FROM THE EDITOR

This issue is devoted mainly to selected papers presented at the 27th Conference of Centre for International Research on Economic Tendency Surveys (CIRET), held in Warsaw, Poland, 15—17 September 2004. The papers were reviewed and submitted for publication in our journal.

There is also a review of important monography published by UN Statistics Division this year on *Household Sample Surveys in Developing and Transition Countries*, in **Book review**, and an **announcement** on the *European Conference on Quality and Methodology in Official Statistics in* 2006.

Business tendency surveys it is a special system of surveys on enterprises and consumers aiming at collection of opinions of particular groups of respondents on present and future situation in the nearest months. These are qualitative surveys e.g. providing general information on tendencies of development in the economy and feelings of consumers so they do not aim at gathering of the definite quantitative information (of which the source is the official short term statistics). This type of surveys is widely spread in Europe and outside. For many years European Union and OECD has been introducing *Joint Harmonised Program of the Business* and *Consumer Tendency Surveys* in the member states with the use of the harmonised set of questions.

Their main subjects were:

- Business tendency surveys, including investment survey: analysis of surveys results, new methods and other chosen subjects;
- Consumer surveys: analysis of results, new methods;
- Business tendency indicators — analysis and new indicators;
- Special surveys.

Out of nearly ninety papers presented at the Conference, the Editorial Board of the *Statistics in Transition* selected thirteen for publication in our journal, taking into account statistical aspects of the papers.

There are following thirteen papers devoted to some problems of business surveys and consumer surveys:

1. ***The usefulness of business surveys data for short-term forecasting — Raw data vs. seasonally adjusted and smoothed one*** (by E. Adamowicz, S. Dudek and K. Walczyk**,** from Poland). Using data for Poland (the business tendency surveys in manufacturing), the authors try to show that the course of transition of the economy from centrally-planned regime to market one is characterised by changes of a seasonal pattern of industrial activity. This makes any conclusions drawn from an analysis of raw BTS indicators are possibly

misleading and calls for application of seasonal adjustment methods which are able to model variable seasonality.

2. ***The Role of Monetary Instruments in the Stabilization Policy in Poland under Transition*** (by R. Barczyk and M. Kruszka from Poland). The aim of the paper is theoretical and empirical analysis of the influence of monetary policy instruments on business fluctuations in the Polish economy during transformation. The first part of paper formulates theoretical hypotheses concerning the influence of the selected instruments of monetary policy on the changes of business activity in Poland under transition. The second part presents morphological features of business fluctuations and also shows empirical relations between the selected instruments used by the central bank and short and medium term changes of business activity in Poland.

3. ***Consumer Surveys and Reality*** (by Maurizio Bovi from Italy). The author investigates the usefulness of Italian consumer surveys as estimation and forecasting tool. To this end, standard consumption equations are estimated and then compared, in terms of in-sample and out-of-sample predictive ability, with corresponding models which differ from them only for the presence of the confidence indicator. The paper focuses on the relationships between subjective and objective information at a disaggregated level. The idea behind is to check if one sub-index is more or less informative than another, and if some outlay is more or less "sentiment sensitive" than another.

4. ***Quality Control and Aggregation Methods in the ISAE Investment Survey*** (by T. Cesaroni, M. Malgarini and G. Rocchetti from Italy). The Institute for Studies and Economic Analysis (ISAE) carries twice a year a survey on investment expenditures of Italian manufacturing firms. The survey, which is part of the Joint Harmonised Program of the European Commission, is based on the same sample of the monthly survey on the manufacturing sector, stratified upon firms' industry, region and size. New methods for quality control and processing of the results are proposed.

5. ***Regional Differentiation of Business Cycles in Poland, 1999—2004*** (by J. Fundowicz, K. Łapiński, M. Peterlik and B. Wyżnikiewicz from Poland). The paper aims at analysis of business cycles in 16 Polish voivodships. This analysis was based on methodology applied for surveys of international business cycles synchronization. Calculations were based on industrial sales and unemployment time series for 16 regions and entire Poland.

6. ***The Use of Banking BTS Data in Diagnosing and Forecasting Situation of Industry and Construction in Poland*** (by J. Garczarczyk and R. Skikiewicz from Poland). The aim of the paper is to construct econometric models describing relationships between the economic situation in banking sector (qualitative data) and macroeconomic indicators showing economic situation

in other sectors of the Polish economy (quantitative data from the Central Statistical Office).

7. ***Measuring the Usefulness of Consumers' Inflation Expectations in Finland*** (Pertti Kangassalo and Kari Takala from Finland). The authors consider the accuracy of consumers' assessments and expectations about the past inflation and about future inflation. The data from the Finnish Consumer Survey are used as indicators of inflation assessments, i.e. valuations concerning the past inflation, and of inflation expectations. These data are available monthly since October 1995. In the Consumer Survey consumers state the inflation rate as a direct percentage change from the previous year's corresponding month

8. ***An Analysis of Inflation Expectations of the Turkish Private Manufacturing Industry*** (by Ercan Karadaş, Fethi Öğünç from Turkey). The main purpose of this paper is to make a detailed analysis of the quantified expected inflation series obtained from the Business Tendency Survey conducted by the Central Bank of the Republic of Turkey. Different representations of simple expectation formation mechanisms that appeared in the literature are discussed in the context of Turkey.

9. ***Occupational Segregation in the Russian Labour Market*** (by A. Klimova from Australia). This paper examines gender inequalities in the labour market in Russia by examining gender occupational segregation. First, a statistical overview of the trends in the Russian labour market outcomes by gender during the Soviet period, perestroika and the period of transition is presented. This is followed by a review of current economic theories of occupational segregation and discrimination in the labour market, including explanations based on tastes, occupational exclusion and statistical discrimination.

10. ***Alternative Sampling Designs — Some Applications of Qualitative Data in Survey Sampling***(by B. Kowalczyk from Poland). The objective of this paper is to present a possibility of application of some alternative sampling designs, which are based on ranked sets in market and consumer surveys. The author introduces and analyzes an example of sales estimation in pharmacies. data, order statistics, ranked sets.

11. ***The Assessment of the Investment Activity of Enterprises Based on the Business Tendency Surveys*** (by H. Sękowska, M. Święcka, K. Walkowska, I. Zagoździńska from Poland). Investment survey includes, among others, size and directions of changes in investment activity as well as motives and sources of its financing. The results of the surveys confirm the notion of still difficult situation of Polish enterprises in the sphere of investment activity, however, larger units as well as those with foreign capital, are in mach better situation.

12. ***Aggregation of Exchange Rate Data and Long Memory Measures*** (by E.M. Syczewska from Poland). The author presents results concerning testing of possible long-memory behaviour of a time series. This task is important from

applied econometrician point of view, e.g., in financial econometrics, as long-memory effects influence not only behaviour of a series, but also statistical properties of various statistics, used to test, e.g., financial market efficiency, to formulate forecasts etc. The author compares results of three methods, used to detect or reject the presence of a long-term dependence in the series

13. ***Business Surveys and Official Statistics in Russia. Which One is Better?*** (by Serguey Tsukhlo from Russian Federation). The author starts with history of business surveys (BS) in Russia and compares information obtained from the surveys conducted by the independent private scientific institutes with information the source of which is the official quantitative statistics. According to his findings, the unreliability of reporting, as a rule, does not depend on a particular sector, size or property form of an enterprise. Survey results have proved to be quite competitive vis-à-vis the official statistical data. Surveys can receive high scores from respondents, providing their organizer is aware of their needs and takes those into account while preparing the results to be sent back to the enterprises.

There is also one **Book Review** devoted to *Household Sample Surveys in Developing and Transition Countries***,** Studies in Methods, Series F, No. 96, Statistics Division, United Nations, New York, 2005, 655 pages (prepared by Jan Kordos). This publication presents several important aspects of conducting household surveys in developing and transition countries, including sample design, survey implementation, non-sampling errors, survey costs, and analysis of survey data. The main objective of this handbook is to assist national survey statisticians to design household surveys in an efficient and reliable manner, and to allow users to make greater use of survey generated data.

An Announcement is connected with the *European Conference on Quality and Methodology in Official Statistics* (Q2006**)** which will be held in the UK in 2006**.** The Conference will be prepared jointly by the UK Office for National Statistics and Eurostat. The Q2006 conference will be the third in a series of scientific gatherings covering important methodological and quality-related topics of relevance to the European Statistical System.

Jan Kordos

The Editor

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 199—219

THE USEFULNESS OF BUSINESS SURVEYS DATA FOR SHORT-TERM FORECASTING

Raw Data vs Seasonally Adjusted and Smoothed One[1]

Elżbieta Adamowicz, Sławomir Dudek, Konrad Walczyk[2]

ABSTRACT

The main aim of the research is to show how important are seasonal adjustment and smoothing the data from business tendency surveys in analyses and short-term forecasting of economic activity (a quantitative index).

Using data for Poland (the business tendency surveys in manufacturing conducted by RIED, WSE), the authors try to show that the course of transition of the economy from centrally-planned regime to market one is characterised by changes of a seasonal pattern of industrial activity. This makes any conclusions drawn from an analysis of raw BTS indicators be possibly misleading and calls for application of seasonal adjustment methods which are able to model variable seasonality.

In order to conduct this research the authors decompose time series of qualitative indicators into seasonal and irregular factors and a trend+cycle component, using the TRAMO-SEATS procedure. Consequently, the authors study patterns of seasonal factors and its changes throughout the transition process and, finally, consistency and correlation of a quantitative manufacturing index with qualitative indicators resulting from the business tendency surveys, seasonally adjusted and smoothed. Using this data some short-term forecasts are made.

Key words: business tendency surveys, seasonal adjustment, seasonality, short-term foecasting, economic transformation

---

[1] The previous version of this paper was presented at the 27th CIRET Conference, Warsaw, in September 2004. This version presents results based on actualised time series.

[2] The Research Institute of Economic Development, The Warsaw School of Economics.

1. Introduction and the aim of the analysis

The Research Institute of Economic Development, Warsaw, has run business tendency surveys (BTS) in manufacturing sector of the Polish economy since 1986. The survey is taken monthly; the data gathered so far comprise 204 observations. Until recently results of the survey were presented as raw times series. Any analyses and forecasts made on the basis of these results also used raw data. As the time series are long enough it is high time they should have been seasonally adjusted and smoothed. This also gives an opportunity to test how decomposing the time series influence their forecasting properties.

An analysis of usefulness of BTS data is very important as one can observe growing interests in results of these surveys in Poland and increasing demand on forecasts based on these results. The problem of incomplete information is widely known in economic research. The abundance of real economic phenomena always preceds possibilities of their desrciption with the use of imperfect but steady improved research techniques. Data collected in business surveys is an additional source of information for all economic agents. Amongst of the most important features of this data are:

- speed of its receipt; BTS data gives their description at the time economic processes take place in;
- duality of description of economic performance. Variables evaluated in BTS provide us with information both on the presence and business expectations, giving a picture of current economic changes and an opinion of the nearest future;
- microeconomic foundations of data obtained. Opinions gathered are made by economic agents, ie firms and households. Using the information to describe macroeconomic phenomena provides the analysis with solid microeconomic basis;
- observations of variables of different characteristics, that response variously at every stage of the trend cycle: coincidently, with some lags or leads. This lets us describe economic fluctuations with more accuracy and forecast potential changes faster;
- observations of events that appear on both sides of the market, ie demand and supply ones. This enables to make use of BTS results by representatives of various economic schools irrespective of their scientific orientation;
- possibility of international comparisons. BTS are run on a harmonized basis, facilitating international comparative analyses.

The abovementioned properties of BTS data make it a valuable supplement to quantitative data and be useful for all economic actors, including politicians responsible for macroeconomic policy.

The aim of the research

In order to study the usefulness of BTS data in macroeconomic policy the authors try to verify conclusions made earlier upon the raw time series, using seasonally adjusted and smoothed data. Particularly, we aimed to answer the following questions:

- does BTS data collected by the Institute properly reflect the dynamics of real economic performance in Poland?
- does eliminating seasonal and irregular fluctuations and outliers influence forecasting properties of the data?
- is it useful in formulating short-term forecasts of economic activity?

In 2002 the similar study was done with the use of raw data.[1] In this paper we present results of the analysis made using seasonally adjusted and smoothed data. The time span of the research covers years 1993—2003, comprising almost the whole period of transition of the Polish economy.

2. Data used and the method of the analysis

The year-over-year index of total sold production in manufacturing, publicised by the Central Statistical Office (CSO), is assumed to be an endogenous variable. The selection of this definition of index is based on our previous studies and other authors analyses of the usefulness of the RIED industrial survey for foresting (see: Adamowicz, Dudek and Walczyk (2004); Tomczyk (2001); Rocki, Tabeau (1995)). It should be stated that whenever these authors compare qualitative and quantitative variables the problem how to define the index of a quantitative variable is a crucial point. This problem was analysed by many researchers (see also: OECD (2003), p. 57—62; Bennett (1984)). In this paper we rely on our empirical analysis which showed that the year-over-year index can be treated as a reference variable for balances from the RIED industrial survey. Thus, the index is calculated assuming sold production be equal to 100 in corresponding period of a previous year (ipyoy).

Monthly business survey data (RIED data) of January 1993 — April 2005 is used as explanatory variables in the analysis. It comprises:

- sold production (q1),
- total orders (q2),
- export orders (q3),
- stocks (q4),
- prices (q5),

[1] Results of this analysis were presented at the 26th CIRET Conference in Taipei (see: Adamowicz E., Dudek S, Walczyk K.: The use of business survey data in analyses and short-term forecasting. The case of Poland, Taipei 2002).

- employment (q6),
- financial standing (q7),
- opinion about general situation of the Polish economy (q8) and
- the RIED business indicator (gbi).

All variables (besides the RIED business indicator) are analysed in two formulas: as a change in relation to previous periods, called ‘state’ (s), and as an expected change in the next 3 or 4 months — ‘expectations’ (p). Time series were adjusted of seasonal and irregular fluctuations, adapting TRAMO-SEATS models. Consequently, all pairs of variables are expressed in forms of raw series, seasonally adjusted series and trend+cycle component. In total there are more than 50 variables to analyse.

In the paper we analyse consistency and correlation of a quantitative manufacturing index with qualitative indicators from the business tendency surveys after seasonal adjustment and smoothing. In the analysis we apply two methods:

- cross-correlation analysis and
- regression analysis.

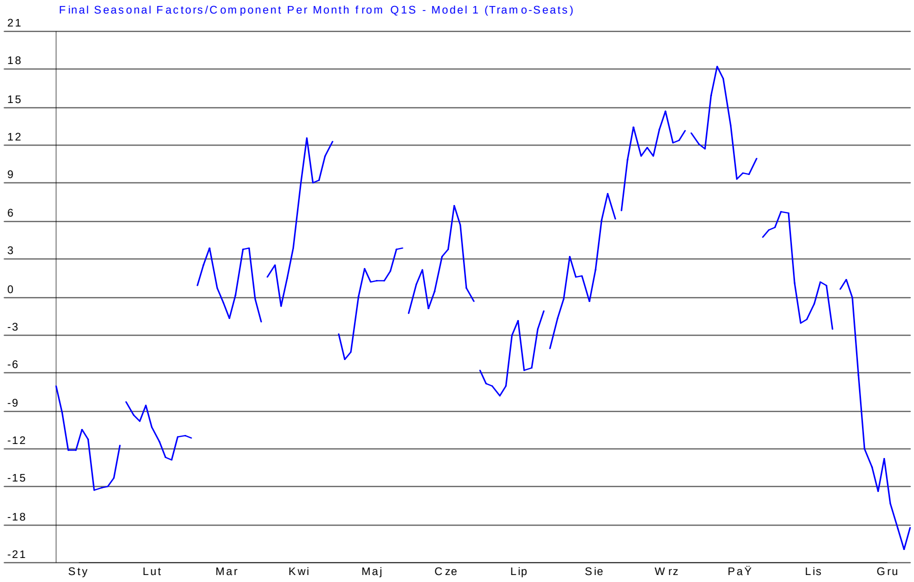
The purpose of the correlation analysis is to quantify consistency between qualitative and quantitative variables by the Pearson`s coefficient of correlation. On the one hand we try to find such independent variables (RIED qualitative indicators) which are most correlated with the index of sold production (CSO manufacturing index), on the other hand we analyse a time profile of explanatory variables, ie whether they lead the tendency of manufacturing production indices or react with lags. When comparing a fit of quantitative variables with qualitative ones we took into account leads and lags from one to twelve months.

The aim of developing regression models is to predict sold production in manufacturing, taking advantage of publication outstrip as well as from predictive power of business trend surveys; in other words, we would like to test forecasting abilities of BTS data.

3. Patterns of seasonality in the transition period of Poland

There are two most popular methods of seasonal adjustment and smoothing used in time series analyses: ARIMA X-12 and TRAMO-SEATS. We applied both of them and obtained similar results. Taking into account the fact that TRAMO-SEATS is model-based and more oriented on stochastic properties of time series we decided to use this method of decomposition of RIED data. We extracted seasonal factors for each of the analysed time series.

Figure 1. Final seasonal component per month from q1s.

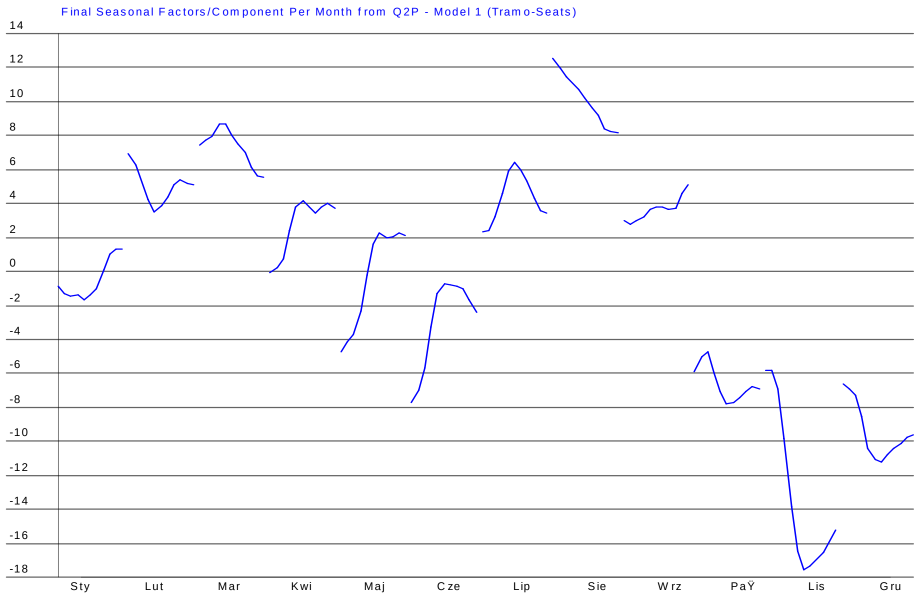


Source: own calculations.

Amplitudes of seasonal fluctuations for each of the variables are different. Sold production (q1s and q1p) and total orders (q2s and q2p) are characterised by high seasonality; the amplitudes amount to 38 and 26 points, respectively. Variability of production is thus demand-led and demand factors are clearly influenced by the weather cycle. Prices (q5s and q5p), opinion about general economic situation — state (q8s), financial standing — state (q7s) and stocks state (q4s) are of medium seasonality (19, 19, 17 and 16 points, respectively). Seasonal fluctuations of all other indicators are low. Employment, state and expectations (q6s and q6p), shows very low seasonality (6 points). As time went by the amplitude of seasonality was diminishing, for stocks state (q4s), prices — state and expectations (q5s and q5p), or increasing, for sold production — expectations (q1p), export orders — expectations (q3p). These changes are probably caused by improvement of firms` stock management or reduction of inflation pressure which made easier price setting.

On the whole, seasonal fluctuations are regular. In case of some indicators a pattern of seasonality is very stable (total orders state, employment expectations, financial standing — state and expectations, opinion about general economic situation — state and expectations). Seasonality of other variables changed throughout the transition process and these changes comprised both amplitudes of fluctuations and displacement of their extrema between months. A pattern of seasonality was biased dramatically as far as sold production (q1s and q1p) and demand indicators (q2p, q3s, q3p) are considered; the details of these changes are given below.

Figure 2. Final seasonal component per month from q2p.

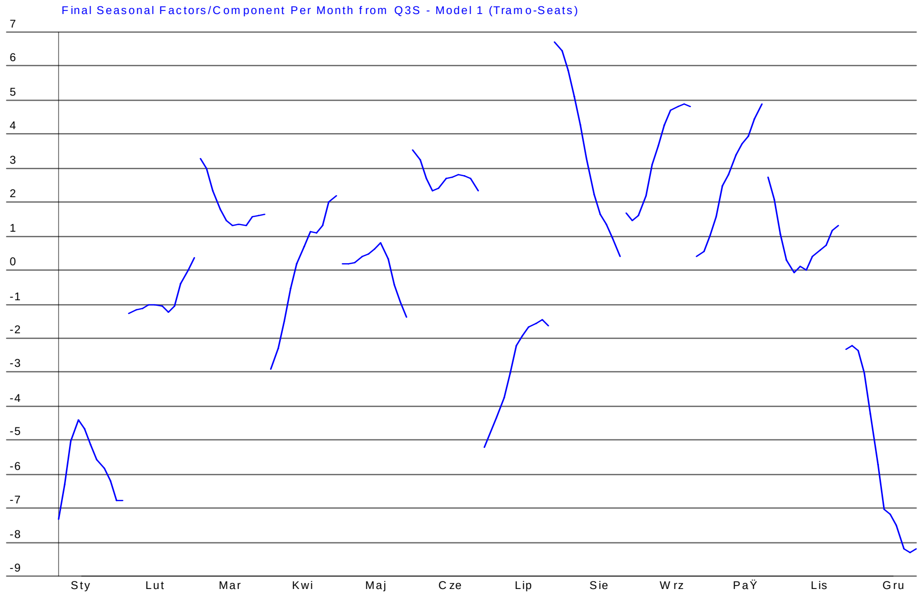


Source: own calculations.

As for manufacturing production (q1s) the beginnings of 1996 and 1998 clearly divide the analysed period into three subperiods (see Figure 1).

Before 1996 the amplitude of seasonal fluctuations of manufacturing production (q1s) was not high, ranging from —12 to 13 points. Three peaks were sharp enough to be noted: the first and the second one came in March and June and they did not exceed 5 points; the third one in October amounted to 13 points. The spring upswing appeared as the economy was recovering after relieved demand in the winter season. It was then followed by a moderate drop of labour productivity in May and a repulse in June which was due to slightly increased purchases induced by preparations to summer holidays. Summer season was bringing about temperate decline of manufacturing production (up to 7 points), caused by lower labour productivity, and was then augmented in October, forestalling the before-Christmas sales. The following downfall in December was the strongest one and “freezing” the economy until the beginning of constructional season next calendar year.

Figure 3. Final seasonal component per month from q3s**.**



Source: own calculations.

The subperiod of 1996 and 1997 was transitory. The size of autumn/winter fluctuations was bigger, ranging from —16 to 18, and the late spring (May/June) business activity changes were overshadowed by the highest rate of GDP growth (7%) since the fall of the communist regime.

Starting from 1998 the spring peak of production fluctuations gets almost the same value as the autumn one (12—13 points), recompensing the November-December downfall which reaches 20 points. The summer descent is, more or less, half of the winter one and is, as a rule, preceded by the short upward swing in June when the rate of the economy is low (eg in years 1999—2001).

Seasonal fluctuations of production expectations index (q1p) are less irregular, but in general, the pattern of changes of these fluctuations over the analysed period of time is corresponding to q1s. Again, year 1996 began a two-year period of "adjustment" after which the amplitude of seasonality doubled from 15 (-7,5 to 7,5) to 30 (-20 to 10).

The seasonality pattern of total orders expectations (q2p) is more regular than the one of q1p but, surprisingly, shows two peaks, in March (8 points) and August (13 points), and two troughs, in June (-7,5 points) and November/December (-17,5 points) (see Figure 2). This means that firms do not interpret market signals properly. Once again, the pattern of seasonal fluctuations changed by the beginning of 1996. Before that the amplitude of fluctuations was smaller (20 points) and June downfalls were even deeper than November ones.

After 1996 the August peak lowered (to 8 points) as well as the November trough (to –16 points) and the mid-year trough moved up (to –2 points).

As far as export orders variable is concerned the beginning of 1996 divides the analysed period into two subperiods (see Figure 3).

Before 1996 the seasonality pattern of export orders displays four peaks in March, June, August (the highest one) and October (the latter dieing out) and the corresponding troughs in April, July, September and January (the deepest one). The magnitude of upward and downward swings was relatively high, except for October one. The amplitude of seasonal fluctuations ranged from –7,5 to 7 but was falling off through the period. After 1996 the pattern of seasonal fluctuations changed; the October upswing disappeared and the size of April and July drops decreased. There is only one big descent in December (reaching –8 points) and the amplitude of fluctuations is a bit smaller (a drop by one point) than before 1996. The pattern of seasonality corresponds to the one of total orders but the amplitude of fluctuations is twice smaller than that of total orders (seasonal factors abroad are weaker).

The results of our analysis suggest that patterns of seasonality of indicators changed during the transition period. Theses changes were caused by two factors: the change of firms` economic behaviour (from centrally planned regime to market one) and the structural change of the Polish economy.

4. Forecasting properties of seasonally adjusted and smoothed BTS data.

4.1. Cross-correlation analysis

In the analysis we assume that correlation between variables is "strong" when a correlation coefficient is greater then 0,6 ($\left|r_{xy}\right| > 0{,}6$). It is certainly a simplified assumption which is not based on any statistical test.

Taking into account economic sense of qualitative variables we look for positive correlation with the industrial production index (ipyoy). Stocks are the only exception. Cross-correlation coefficients are calculated using a standard estimator (see Adamowicz, Dudek and Walczyk (2004)).

The results of calculations are summarised in two tables. Table 1 presents maximum absolute correlation coefficients with corresponding lags or leads for BTS variables and components of decomposition for the whole sample (January 1993 to April 2005). Cross-correlation coefficients for the shortened, second half of the analysed sample (January 1998 to April 2005) are presented in Table 2.

Cross-correlation analysis of the whole sample of raw time series (Table 1) shows that 4 of 17 BTS balances are strongly correlated with the reference variable (according to the ‘rule of thumb’, defined above). The highest correlation coefficient (0,6868) we note for export orders state (q3s) with no lag and no lead

(L=0). More than a half of analysed variables seems to be week indicators of the yearly manufacturing index. The results show that two variables are leading (q4s — stocks state, q4p — stocks expectations), however they are poorly correlated with reference variables, and five are coincident (q2s — total orders state, q3s/q3p — export orders state and expectations, q5s/q5p — price state and expectations). Quite poor correlation for general business indicator (gbi) and production states and expectations (q1s, q1p) requires to be underlined; in addition, the highest correlation coefficients have been obtained for coincident balances.

Table 1. The highest coefficients of cross-correlation with the ipyoy index (whole sample)

Reference variable: ipyoy, sample: 01:1993-04:2005						
BTS series	Raw series		SA series		trend + cycle	
	L	R	L	R	L	R
GBI	1	0,4268	1	0,5368	2	0,5002
Q1S	1	0,3704	-1	0,5614	-1	0,5397
Q1P	3	0,2663	4	0,406	5	0,4666
Q2S	0	0,5744	0	**0,7476**	0	**0,7544**
Q2P	3	0,382	2	0,5091	3	0,5191
Q3S	0	**0,6868**	0	**0,7605**	0	**0,8125**
Q3P	0	**0,6169**	1	**0,7181**	0	**0,8259**
Q4S	-3	-0,2985	-3	-0,4397	-3	-0,4673
Q4P	-1	-0,3647	0	-0,4828	0	-0,5356
Q5S	0	**0,6027**	0	**0,663**	0	**0,7226**
Q5P	0	0,5377	0	0,5978	0	**0,6145**
Q6S	3	0,5155	3	0,5784	4	0,5747
Q6P	3	0,5524	3	0,593	3	**0,6389**
Q7S	1	0,4307	1	0,5136	2	0,4301
Q7P	3	0,3786	3	0,3951	6	0,4068
Q8S	3	**0,658**	3	**0,6893**	3	**0,7413**
Q8P	2	0,5919	2	**0,6022**	3	**0,644**

L – lag/ lead
R – correlation coefficient
Source: own calculations

The results significantly improve after seasonal adjustment of time series; only for three variables we receive poor correlation ($\left| r_{xy} \right| < 0{,}5$). The highest correlation coefficient (0,7605) we note again for export orders state (q3s) with no lag and lead (L=0). Very high correlations we receive also for q2s — total orders state (0,7476) with L=0, q3p — export order expectations with an one-month lag (0,7181). For seasonally adjusted series we receive two leading indicators and five coincident ones (see Table 1).

In general, smoothing time series (decomposing seasonal and irregular fluctuations, outliers and a trend+cycle component) improves results even more (but there are cases in which correlation is lower than for seasonally adjusted series). The highest correlation coefficient (0,8259) is noted for export orders expectations (q3p) with no lag and lead (L=0). Very high correlation are also obtained for q3s — export orders state (0,8125) with L=0 and q2s — total order states with L=0 (0,7544). For trend+cycle components we notice two leading indicators and six coincident ones (see details in Table 1).

On the whole we receive better results for smoothed series (seasonally adjusted and detrended). The best indicator of the manufacturing production index is export orders state and expectations (q3s and q3p) and total orders state (q2s). We receive a few of leading indicators but it should be underlined that as far as coincident indicators are concerned correlations coefficients with one- or two-months lead are close to the highest one. Thus, coincident indicators can be tested in further analyses for forecasting purposes. It is worth to remark that we receive not very acceptable results for production balances and general business indicator what seems to be natural indicators for the quantitative production index. The reason of this is that we observe weaker correlation for these variables in the first half of the sample; in the next step of the correlation analysis the shorter sample is used and consistency results are much better. We receive the highest lead for balances on finished goods stock states and expectations (3—4 months) but correlations for these variables are not very strong. It is remarkable also that balances concerning opinion on general economic situations are lagged against the reference variable (1—2 months).

In the next step of the cross correlation analysis we proceed with the second half of the sample as the graphical analysis shows that consistency between the reference variable and BTS ones is much better in the second half of our sample. This conclusion justifies some general proposition concerning business surveys conducted in countries under transition; it says that an accuracy of expectations formulated by firms increases proportionally to the length of the transition process. This makes us conduct analysis for the sample of 1998 to 2005; the results are presented in Table 2.

Table 2. The highest coefficients of cross-correlation with the ipyoy index (shortened sample)

Reference variable: ipyoy, sample: 01:1998-04:2005						
BTS series	Raw series		SA series		trend + cycle	
	L	R	L	R	L	R
GBI	1	0,5709	1	**0,7916**	0	**0,8199**
Q1S	1	0,4986	0	**0,7919**	-1	**0,8398**
Q1P	0	0,2559	0	0,4662	1	0,5734
Q2S	0	**0,6065**	0	**0,8223**	0	**0,8607**
Q2P	2	0,4405	2	**0,6619**	2	**0,7758**
Q3S	-1	**0,6727**	0	**0,748**	0	**0,807**
Q3P	0	0,5546	1	**0,6865**	1	**0,8532**
Q4S	-3	-0,4029	-3	-0,508	-3	-0,5972
Q4P	0	-0,3121	0	-0,4254	0	-0,5114
Q5S	1	**0,6269**	1	**0,6833**	1	**0,7392**
Q5P	0	0,4478	0	0,5068	-12	-0,5018
Q6S	3	0,5107	3	0,5524	4	0,5388
Q6P	2	0,5113	3	0,5285	3	0,5678
Q7S	1	**0,6487**	1	**0,7742**	1	**0,7893**
Q7P	2	0,5973	2	**0,6612**	3	**0,6956**
Q8S	2	**0,6927**	2	**0,6985**	2	**0,7544**
Q8P	2	**0,6364**	2	**0,629**	2	**0,681**

L – lag/ lead
R – correlation coefficient
Source: own calculations

On average, correlations coefficients for last six years are higher than for the whole sample. Improvements are evident especially for the production balances and the general business indicator.

For raw time series we receive six strongly correlated qualitative indicators with the reference variable (in comparison with four for the whole sample). The highest correlation coefficient (0,6927) is noted for general opinion about economic situation in Poland (q8s), with a two-months lag, and for export orders state (q3s) with an one-month lead (0,6727). The calculations show that two variables are leading and five are coincident. In opposition to above results (for the whole sample) the general business indicator (gbi) and balance on production state (q1s) are quite strongly correlated with the reference variable.

The results significantly improve for seasonally adjusted time series; only for six variables we receive poor correlation. The highest correlation coefficient (0,8223) is noticed for total orders state (q2s) with no lag and lead (L=0). The very high correlations are obtained also for GBI — the general business indicator (0,7916) with an one-month lag, q1s — production state (0,7919) with L=0, q3s — export order state (0,748) with L=0 and for balance on financial situation state (0,7742) with an one-month lag. Taking into account seasonally adjusted series we receive one leading indicator and six coincident ones (see Table 2).

In general the trend+cycle component improves results more (but there are two cases in which correlation is lower than for seasonally adjusted series), only for one variable we receive poor correlation (prices expectations). The highest correlation coefficient (0,8607) is noted again for total orders state (q2s) with no lag and lead (L=0). Very high correlation is also obtained for q3p — export orders expectations (0,8532) with an one-month lag, gbi — the general business indicator (0,8199) with L=0, q1s — production state (0,8398) with an one-month lead, q3s — export orders state with L=0 (0,807) and for balance on financial situation state (0,7893) with an one-month lag. For trend+cycle components we received three leading indicator and four coincident ones (see Table 2).

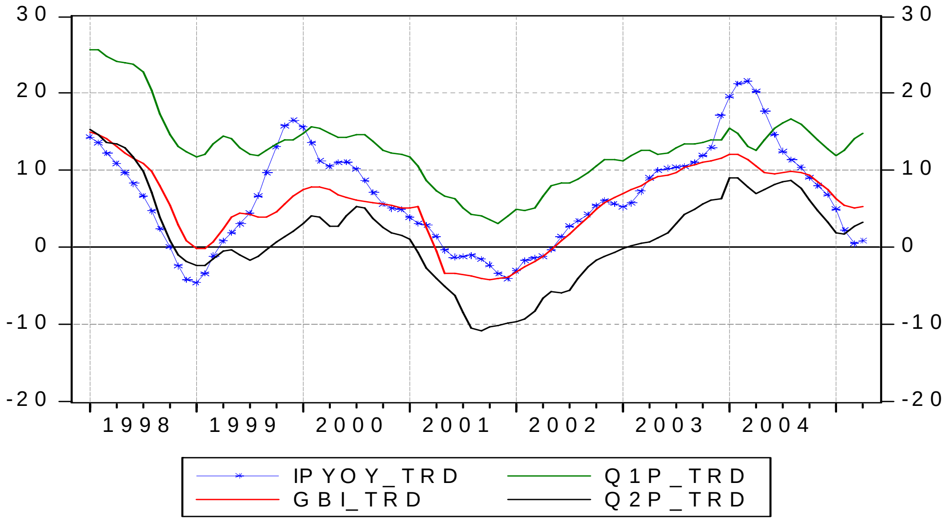
While analysing the second half of our sample we receive, in general, stronger correlation for smoothed series (for seasonally adjusted as well as for trend+cycle component). The best indicator for the manufacturing production index is total orders state (q2s), production state (q1s), the general business indicator (gbi) and, as above, export orders state and expectations (q3s and q3p). These results are similar to other research results. Unfortunately, we receive few leading indicators, even for smoothed time series, and a lot of lagged indicators. As mentioned above correlations coefficients with one- or two-months lead are close to the highest one for coincident variables (for L=0). On the other side the analysed qualitative indicators asymmetrically describe (lead or lag) peaks and troughs of the manufacturing production index what is natural feature of business fluctuations (especially in transition countries). This fact, in connection with quite short time series (several peaks and troughs), influence cross-correlations results, especially a lead/lag profile. To some extent it can be shown on Figure 4, which illustrates combined trend+cycle components for the reference variable and production expectations (q1p), total orders expectations (q2p) and the general business indicator (gbi).

Figure 5 shows the last trough in the manufacturing index is signalled with one- to four-months lead. The qualitative variables are coincident with a trough in the beginning of 1999. On the other side all qualitative indicators are lagged against a peak in the beginning of 2000. This analysis shows asymmetry of a lead/lag profile which causes these variables be lagged in the cross-correlation analyse, despite the fact they represent expectations. It is obvious that asymmetry in leading peaks and troughs in connection with short time series can influence radically a cross-correlation analysis. Due to this fact we do not eliminate from further analyses variables which formally should be useful for forecasting.

The other reason that we do not detect leads in some expectations qualitative variables is the definition of the manufacturing index. It may happen that the production index should be defined in diffferent ways for different qualitative variables. Here we assume the year-over-year index be used for all variables.

As for the whole sample balances concerning opinion on general economic situation are lagged against to the reference variable (1—2 months).

Figure 4. The general business indicator, production expectations balance, total order expectation balance against the manufacturing index (trend+cycle components).



Source: own calculations.

From all the above one can draw the following conclusions:

- the cross-correlation results are on average (for both samples) higher for seasonally adjusted and smoothed time series than for raw ones; this means that seasonal and irregular fluctuations and outliers affect negatively the relation between the manufacturing index and industrial survey variables;
- the cross-correlation results are on average higher in the second half of the analysed sample and a possible way to interpret this is that in the course of the transition process the accuracy of firms` opinions is increasing; to some extent it can be caused by combining public and private sector surveys in 1997;
- seasonal adjustment and smoothing influence a lead/lag profile of cross-correlation analysis; for smoothed series the detection of peaks and troughs is much easier and evident;
- there are few leading indicators and quite a lot of coincident and lagged indicators but it needs to be stated that correlation coefficients are also high for leads and coincident variables can be potential forecasting indicators.

4.1. Regression analysis

In order to test forecasting abilities of BTS data we specify the ADL model (autoregressive distributed lag model) of the following formula:

$$y_t = \alpha_0 + \sum_{q=1}^{Q} \alpha_q \, y_{t-q} + \sum_{k=1}^{K} \sum_{s=0}^{S} \beta_{k,s} \, x_{k,t-s} + \xi_t$$

or

$$\alpha(L) y_t = \alpha_0 + \sum_{k=1}^{K} \beta_k (L) x_{k,t} + \xi_t$$

where:

y_t - reference variable (ipyoy),

$x_{k,t}$ - BTS variables as balances,

$\alpha(L)$, $\beta_k(L)$ - lag polynomial,

α_0, α_q, $\beta_{k,s}$ - parameters,

ξ_t - residuals.

In practice we decide to use simplified model specification with one qualitative explanatory variable:

$$y_t = \alpha_0 + \alpha_1 x_{k,t-s} + \xi_t$$

where notation is as above. Leads up to 3 months (s=0,1,2,3) are tested. We test models for different decomposition components (raw series, seasonally adjusted series and a trend+cycle component) as in the correlation analysis. In the Appendix we present detailed results fo ther set of models. Tables present properties of the model, residuals and forecasting properties. As criteria for comparison forecasting properties of all models we use *ex-post* forecasting error measures (RMSE, MAE, MAPE).

Regressions were estimated with the sample January 1998 — September 2004; we start from 1998 taking into account results from the cross-correlation analysis. For the period October 2004 — April 2005 we make forecasts and calculate forecasting error measures. In this stage of the analysis, to assess an impact of the decomposition and to select good forecasting BTS indicator we focus on forecasting properties.

For raw time series all models (grouped by variables and by leads) are characterised by low adjusted determination coefficients, ranging from 0,007 to 0,457 (see Table 1 in Appendix). The standard error of regression amounts to 5,6-7,6. Residuals in all cases are autocorrelated. In general, all models have quite poor statistical properties. Forecasting properties are however better: RMSE ranges from 3,87 for prices expectations (q5p(0)) to 10,63 for employment expectations (q6p(-3)). Plausible forecasting properties are noticed for production state (q1s(-2)) and for stocks expectations (q4p(0)).

For seasonally adjusted series, in general, models reveal worse forecasting properties: RMSE ranges from 4,6 (prices state (q5s) lagged two months) to 12,1 (employment expectations (q6p)) (see Table 2 in Appendix). The best forecasting properties are achieved for prices state and expectations (q5s and q5p) and stocks expectations (q4p). Again, residuals are autocorrelated. Adjusted determination coefficients are much higher than for raw data; they range from 0,14 to 0,71.

Data removed of irregular component and outliers improve forecasting properties of BTS series. RMSE ranges from 1,76 for prices state to 10,9 for employment expectations. Plausible forecasting properties measured by RMSE are again obtained for prices state and expectations (q5s and q5p) and for stocks expectations (q4p). Residuals are highly autocorrelated for all models. Adjusted determination coefficients are close to those for seasonally adjusted data; they range from 0,12 to 0,77.

Surprisingly, the *ex-post* forecasting errors we calculated for the last six months of the sample are generally higher than those obtained for the corresponding period of the sample analysed in the previous version of this paper (the second half of year 2003). These differences are quite small as far as models with trend+cycle component variables are concerned. Thus, we conclude that the last year was relatively full of irregular fluctuations (and outliers) connected to EU accession disturbances.

Summarizing the results of the regression analysis we can draw conclusions which are to some extent in line with the findings of the cross-correlation analysis:

- the forecasting properties are on average better for seasonally adjusted and smoothed time series; RMSE is much lower for trend+cycle component models than for raw time series;
- in all models residuals are autocorrelated. This problem is, to some extent, connected with asymmetry of a lead/lag profile and the length of the time series; these may cause autocorrelation of residuals and can be solved by assuming AR(1) or MA(1) process for residuals;
- the lower *ex-post* forecasting errors are noted for prices state and expectations (q5s and q5p), with leads up to one month, and for stocks expectations (q4p).

5. Summary and conclusions

This paper summarises the outcome of several attempts to study empirically and eventually adopt and apply a method of seasonal adjustment and smoothing BTS data to improve its descriptive and forecasting properties. The analysis we conducted let us conclude as follows:

1. Patterns of seasonality changed during the analysed period. The changes were probably driven by the process of economic transition or economic reorientation of firms.
2. In order to model changes of economic activity some procedures appropriate to analyse the variability of seasonality should be applied. They will hopefully help answering the question whether changes in patterns of seasonal fluctuations were caused by structural adjustment of the economy, learning effects in firms, resampling or other reasons.
3. Seasonal adjustment and smoothing of time series influence statistic and forecasting properties of all analysed variables by both raising correlation between them and changing their lead/lag profiles.
4. Indications of the cross-correlation analysis changed in the analysed period. Thus, we decided to make calculations for two periods: the whole period of 1993-2005 and its second half. Generally speaking, correlation coefficients for years 1998-2005 were much higher (the highest ones were for total orders state, export orders expectations and productions states). Forecasting properties of models also improved (excluding the last year); the best ones for prices state and expectations, one month leaded.
5. The improvement of statistical properties of RIED data increases the usefulness of prognoses based on them in macroeconomic policy.

It needs to be underlined that time series under the analysis are short and the period was specific in the Polish economy due to the fact that it covered only one business cycle featured by the assymetry of leads and lags in each of the its phases. Thus, the analysis should be followed by further studies using much longer sample.

REFERENCES

ADAMOWICZ E., DUDEK S., WALCZYK K. (2004), The Use of Survey Data in Macroeconomic Analysis, in: Composite Indicators of Business Activity for Macroeconomic Analysisis, ed. Matkowski Z., *Prace i Materiały Instytutu Rozwoju Gospodarczego*, Issue 74, SGH, p. 357—376.

BENNETT A. (1984), Output expectations of manufacturing industry, *Applied Economics*, 16, p. 869—879.

BLAKE A., KAPETANIOS G., WEALE M.R. (2000), Nowcasting EU Industrial Production and Manufacturing Output, NIESR.

CUNNINGHAM A. (1997), Quantifying survey data, *Bank of England Quarterly Bulletin*, p. 292—299.

DEMETRA 2.0 — Pedagogical Manual (2002) EUROSTAT.

DEMETRA 2.0 — User manual (2002) EUROSTAT.

DRAMAIS A. (2003), Forecasting Industrial Production from Business Survey data for the Euro Area, ECOMOD conference on Policy Modeling — Instanbul — 3—5 July 2003.

ETTER R., GRAFF M. (2002), Estimating and Forecasting Production and Orders in manufacturing Industry from Business Survey Data, 26th CIRET Conference, Taipei.

EUROPEAN COMMISSION DG ECFIN (1997), The Joint Harmonised EU Programme of Business and Consumer Surveys, Europena Economy — *Raport and Studies*, No 6, Belgium.

EVIEWS 4 Command and Programming Reference (2000), Quantitative Micro Software.

EVIEWS 4 User’s Guide (2000), Quantitative Micro Software.

GERLI M., PETRUCCI A. (1995), The Econometric Anticipation of the Industrial Production Index. Some Results Based on Business Survey Data, 22nd CIRET Conference, Singapore.

GÓMEZ V., MARAVALL A. (1996) Programs TRAMO (Time series Regression with Arima noise, Missing observations and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the user, Working Paper, nr 9628, *Servicio de Estudios*, Banco de España.

KANGASNIEMI J., TAKALA K., TSUPARI P. (2000), Seasonality and Smoothing of Business Survey Data, 25th CIRET Conference, Paris 2000.

OECD (2003), Business Tendency Surveys — A Handbook, OECD.

OPPENLANDER K.H. (2002), Business Cycle Survey Data: Definition, Importance, and Applications, 26th CIRET Conference, Taipei.

ROCKI M., TABEAU A. (1995), Kwantyfikacja Jakościowych Danych Ankietowych dla Produkcji Przemysłu w Polsce, *Roczniki Kolegium Analiz Ekonomicznych*, Issue 2, SGH, Warsaw.

TOMCZYK E. (2001), Racjonalność Oczekiwań Respondentów Testu Koniunktury in: Analiza Tendencji Rozwojowych w Polskiej Gospodarce na Podstawie Testu Koniunktury, *Prace i Materiały Instytutu Rozwoju Gospodarczego*, Issue no 70, Szkoła Główna Handlowa, Warsaw.

Appendix

Table 1. OLS estimation results for raw series (leads for lowest RMSE).

Reference variable: ipyoy, estimation sample: 01:1998-09:2004, forecasting sample: 10:2004-04:2005									
Explanatory variables		Statistics							
Var	L	R^2 adj	SE	DW	LMF(2)	ARCH LMF(2)	RMSE	MAE	MAPE
GBI	**-2**	**0,181**	**6,87**	**0,7175**	**0**	**0**	**4,187**	**3,7144**	**278,0459**
Q1S	**-2**	**0,1879**	**6,841**	**0,7273**	**0**	**0,0003**	**4,0632**	**3,676**	**264,9215**
Q1P	-3	0,007	7,5644	0,5306	0	0	5,1085	3,9548	395,5217
Q2S	-1	0,3166	6,2754	0,9911	0	0,0697	4,4943	4,0483	330,8358
Q2P	-3	0,0853	7,26	0,5826	0	0	4,9769	4,3099	349,717
Q3S	-1	0,4572	5,5926	1,1247	0	0,1795	5,2282	4,3876	385,2922
Q3P	-2	0,2759	6,4596	0,959	0	0,0062	6,6096	5,6196	604,4468
Q4S	0	0,1008	7,1985	0,602	0	0,0011	4,6879	3,5051	408,4516
Q4P	**0**	**0,0683**	**7,3271**	**0,6075**	**0**	**0,0003**	**4,0961**	**3,4133**	**371,1762**
Q5S	**-2**	**0,1952**	**6,8101**	**0,6439**	**0**	**0**	**4,1199**	**3,1798**	**304,7614**
Q5P	**0**	**0,1759**	**6,8913**	**0,6217**	**0**	**0**	**3,8665**	**3,3861**	**191,547**
Q6S	0	0,215	6,7259	0,6989	0	0,0012	9,022	7,6549	795,303
Q6P	-3	0,202	6,781	0,6835	0	0,0015	10,6349	9,8513	947,446
Q7S	0	0,4062	5,8494	0,9612	0	0,3129	5,9607	5,2511	556,2184
Q7P	-3	0,1614	6,9514	0,671	0	0	5,8249	5,2177	433,6388
Q8S	0	0,4002	5,8789	0,8882	0	0,0009	9,3671	8,8067	833,553
Q8P	-1	0,2772	6,4538	0,792	0	0	9,6408	8,6813	839,1577

Source: own calculations
L – lead, R^2 adj – adjusted determination coefficient, SE – standard error of regression, DW – Durbin-Watson statistics, LMF(2) – autocorrelation LM statistics probability, ARCH LMF(2) - ARCH LM test statistic probability, RMSE - root mean squared average error, MAE - mean absolute error, MAPE - mean absolute percentage error.

Table 2. OLS estimation results for seasonally adjusted series (leads for lowest RMSE).

Reference variable: ipyoy, estimation sample: 01:1998-09:2004, forecasting sample: 10:2004-04:2005									
Explanatory Variables		Statistics							
Var	L	R^2 adj	SE	DW	LMF(2)	ARCH LMF(2)	RMSE	MAE	MAPE
GBI	0	0,6246	4,5701	0,9844	0	0	5,5151	4,5507	176,4311
Q1S	0	0,6376	4,4902	1,2907	0	0,1927	5,2017	4,478	151,9408
Q1P	-2	0,1392	6,92	0,5305	0	0	5,5678	4,6622	172,1903
Q2S	0	0,7095	4,02	1,1097	0	0,0876	6,5437	5,5828	201,2221
Q2P	-2	0,3044	6,2207	0,6707	0	0	5,8656	5,2163	187,2733
Q3S	-1	0,5688	4,8977	1,0988	0	0,0092	5,5495	4,9743	197,8323
Q3P	-2	0,4407	5,5781	0,9785	0	0,0083	6,6353	5,9781	231,2289
Q4S	0	0,1862	6,7286	0,504	0	0	6,0276	4,9526	176,799
Q4P	**0**	**0,1536**	**6,8617**	**0,5594**	**0**	**0**	**4,6804**	**3,658**	**166,3629**
Q5S	**-2**	**0,2149**	**6,6088**	**0,4917**	**0**	**0**	**4,5975**	**3,7902**	**133,8122**
Q5P	**-1**	**0,148**	**6,8843**	**0,4821**	**0**	**0**	**4,6876**	**4,1105**	**112,6293**
Q6S	0	0,2589	6,421	0,5328	0	0,0001	10,1656	8,9573	338,6026
Q6P	0	0,3299	6,1053	0,5742	0	0,0004	12,0565	10,7863	404,6637
Q7S	0	0,6213	4,5901	1,0964	0	0,0437	8,8708	8,0181	309,2153
Q7P	-3	0,2315	6,5383	0,5185	0	0	6,9906	5,8528	219,4395
Q8S	-3	0,2433	6,4881	0,5745	0	0	10,3918	9,0735	345,587
Q8P	-3	0,2204	6,5855	0,5524	0	0	10,1291	8,6288	335,5696

Source: own calculations
L – lead, R^2 adj – adjusted determination coefficient, SE – standard error of regression, DW – Durbin-Watson statistics, LMF(2) – autocorrelation LM statistics probability, ARCH LMF(2) - ARCH LM test statistic probability, RMSE - root mean squared average error, MAE - mean absolute error, MAPE - mean absolute percentage error.

Table 3. OLS estimation results for trend+cycle components (leads for lowest RMSE).

Reference variable: ipyoy, estimation sample: 01:1998-09:2004, forecasting sample: 10:2004-04:2005									
Explanatory Variables		Statistics							
Var	L	R^2 adj	SE	DW	LMF(2)	ARCH LMF(2)	RMSE	MAE	MAPE
GBI	0	0,6849	3,7504	0,1097	0	0	4,0103	3,6821	342,4356
Q1S	0	0,6883	3,7302	0,227	0	0	3,3042	2,9179	289,4421
Q1P	-2	0,2022	5,9679	0,0591	0	0	3,8287	3,2198	333,025
Q2S	0	0,7704	3,2015	0,1174	0	0	5,0395	4,9397	394,2565
Q2P	-1	0,4948	4,7491	0,0818	0	0	4,653	4,2549	396,3618
Q3S	0	0,6838	3,7573	0,0969	0	0	5,4449	5,3151	428,1371
Q3P	0	0,7573	3,2914	0,1719	0	0	5,4254	5,0675	460,699
Q4S	0	0,2353	5,8429	0,051	0	0	3,6406	2,9922	313,9437
Q4P	**0**	**0,2439**	**5,8097**	**0,1215**	**0**	**0**	**2,6695**	**2,6349**	**198,6054**
Q5S	**-1**	**0,395**	**5,1969**	**0,0631**	**0**	**0**	**1,7607**	**1,5589**	**156,7786**
Q5P	**-1**	**0,1247**	**6,2509**	**0,0461**	**0**	**0**	**2,2985**	**2,1709**	**162,2896**
Q6S	0	0,2551	5,7668	0,0687	0	0	8,6329	8,4	673,6561
Q6P	0	0,3787	5,2666	0,0757	0	0	10,9083	10,6164	852,5019
Q7S	0	0,6683	3,8482	0,1193	0	0	7,1328	6,8387	564,0199
Q7P	-2	0,292	5,6221	0,0831	0	0	5,2065	4,958	426,3919
Q8S	-3	0,328	5,4773	0,1089	0	0	9,3671	9,0763	735,592
Q8P	-3	0,2959	5,6065	0,1068	0	0	8,623	8,231	698,4424

Source: own calculations

L – lead, R^2 adj – adjusted determination coefficient, SE – standard error of regression, DW – Durbin-Watson statistics, LMF(2) – autocorrelation LM statistics probability, ARCH LMF(2) - ARCH LM test statistic probability, RMSE - root mean squared average error, MAE - mean absolute error, MAPE - mean absolute percentage error.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 221—238

THE ROLE OF MONETARY INSTRUMENTS IN THE STABILIZATION POLICY IN POLAND UNDER TRANSITION

Ryszard Barczyk[1] **Michał Kruszka**[2]

ABSTRACT

The aim of the paper is theoretical and empirical analysis of the influence of monetary policy instruments on business fluctuations in the Polish economy during transformation. The first part of paper formulates theoretical hypotheses concerning the influence of the selected instruments of monetary policy on the changes of business activity in Poland under transition.

The second part presents morphological features of business fluctuations and also shows empirical relations between the selected instruments used by the central bank and short and medium term changes of business activity in Poland.

The results of empirical analysis made it possible to state that monetary policy in Poland under transformation has had no active anti-cyclical character.

Key words: stabilization policy, monetary policy, transition.

1. Introduction

Observations of the economic processes occurring towards the end of the 20th century indicate that most of the highly developed market economies experienced cyclical fluctuations. Their consequence was, among others, a weaker dynamics of economic growth. Similar processes could also be observed in the Polish economy under transformation. This may mean that the structure of objectives of stabilization policy should be changed and that the state bodies through implementation of anti-cyclical policy should guarantee sustainable and steady economic growth.

The aim of these considerations is to analyse the range and consequences of applying the instruments of monetary policy in order to influence short- and

[1] The Poznań University of Economics, r.barczyk@ae.poznan.pl
[2] The Poznań University of Economics, m.kruszka@ae.poznan.pl

medium-term changes in the economic activity in Poland during systemic transformation. The analysis of anti-cyclical function of the instruments of monetary policy is caused by two factors. The first one results from the properties of instruments belonging to this group because they are highly flexible, which is a precondition for their application in the short and medium run. The second factor is accessibility of empirical data characterizing the decisions made about the use of monetary policy instruments in the Polish economy.

The paper consists of two parts. The first one contains hypotheses concerning the anti-cyclical function of the instruments of monetary policy. The statements are formulated through the prism of theoretical achievement belonging to Keynesian and monetarist economics. The second part is an empirical verification of the theoretical hypotheses. It presents morphological features of business fluctuations in the Polish economy of the transformation period, as well as links between certain, measurable instruments of monetary policy and changes of the economic activity in Poland.

2. Hypothetical influence of the instruments of monetary policy on changes of economic activity in Poland under transformation

Discretionary instruments and automatic stabilizers of business conditions are the tools applied by the state in the process of influencing short- and medium-run cyclical fluctuations. In the group of discretionary instruments the most important role is played by measures of monetary and fiscal policy. The role, range of activity and the mechanism of influence of these instruments are differently interpreted in particular approaches of the modern economics. The Keynesian economists treat monetary policy as much less effective in anti-cyclical activities than fiscal policy. They claim that even if monetary policy is applied its possible impact on the fluctuations of real magnitudes does not occur through the growth or decline in money supply on the market but, above all, through changes in interest rates. Monetarists, as opposed to Keynesian school, question the sense of applying fiscal policy measures in the process of stabilization, both as regards their influence on business fluctuations and on inflation. They advocate the application of monetary policy instruments in the process of stabilization because, in their opinion, there exists a direct link between money supply and level of economic activity.

How the central bank influences the changes in economic activity depends on the character of the monetary policy implemented and the latter depends on a phase of a business cycle[1].

Assuming the Keynesian interpretation of the role of monetary policy, it can be said that in the phase of low economic activity expansive policy (policy of

[1] More on this subject: Bednarczyk (1990), pp.32—47.

cheap money) should be carried out. Then the central bank should have the following aims:

- in the refinancing policy there should occur a decline in the costs of refinancing, i.e. a decline in rediscount rate and lombard rate; the central bank should eliminate quotas and additional qualitative requirements as regards bills of exchange and other securities;
- in the open market policy operations of the central bank should include temporary access of cash offered to commercial banks, increasing their liquidity, i.e. the central bank should buy commercial banks' securities;
- in the policy of obligatory reserve the central bank should decrease the rate of reserve which should lead to a growing supply of credit money in commercial banks;
- in the deposit policy the central bank should cause decrease in deposit accounts of the entities of the public sector, which should lead to the growth in deposits of these institutions in commercial banks;
- in the policy of direct influence the central bank should abolish by way of administrative decisions all bans or directives concerning, e.g. quantity, volume and interest rate on the credits granted, emission activity of commercial banks etc.

In the market economy these activities trigger adjustment processes in the money-credit sector which consist in the growth of commercial banks' cash reserves and wider possibilities of granting credits. In this way the supply of credit money increases, which should cause a decrease in interest rates. Through the mechanism a multiplier-accelerator those factors cause further changes in the real sector, which leads to the growth of demand for consumer and investment goods and consequently to a better utilization of production capacities, growth of employment and volume of production[1].

Similar activities of the central bank can and should be undertaken in the Polish economy of the transformation period as under this system the labour factor is underused, and production capacities are not sufficiently used either. According to the Keynesian economics, a gradual growth of demand due to multiplier mechanisms can accelerate economic recovery.

Because reducing the rate of growth of prices is the main target of the central bank in the Polish economy under transformation, the above-mentioned postulates concerning anti-cyclical monetary policy were not fully accomplished. Divergence between anti-inflationary and anti-cyclical activities of the central bank (particularly in the initial period of transformation) led to a longer phase with low dynamics of growth and to its increased amplitude and intensity.

[1] More on the subject of transmission of monetary impulses to the real sphere see, among others: Drabowski, (1987), p.49; Kruszka (2002), pp.18—29.

According to Keynesian economics postulates, restrictive policy should be implemented in the phase of high dynamics of growth in the market economies. It assumes the so-called cooling of the economy and the essence of these activities can be reduced to the following:

- in the policy of refinancing the costs of refinancing should increase i.e. there should occur the growth of rediscount rate and lombard rate; the central bank should introduce quotas in the sphere of refinancing and additional qualitative requirements as regards the refinanced securities;
- in the open market policy the central bank should reduce cash reserves of commercial banks and their liquidity by selling securities;
- in the policy of obligatory reserve the central bank should cause the growth of the rate of reserve which will decrease cash resources (credit money) of commercial banks;
- in the policy of deposits the central bank should reduce deposits of public institutions on the accounts in commercial banks, which should limit their cash reserves and decline in the supply of their credit money;
- in the policy of direct influence the central bank by way of administrative decisions should introduce bans and directives as regards the number, amount and interest rate of credits and the sphere of emission of securities.

Restrictive monetary policy implemented in the market economy under high level of economic activity, through gradual growth of basic interest rates of the central bank, can lead to increased interest rate on consumption and investment credits and therefore to decreased dynamics of the growth of demand for credit money. These factors can weaken the effects of multiplier mechanisms and hinder the rate of improvement of business conditions.

In the economy under transitiona, when business conditions are favourable, similar activities should be undertaken in the sphere of monetary policy because both in the opinions of Keynesian economists and monetarists they can restrict inflationary processes. A negative consequence of the instruments applied in this way can be persisting disequilibrium on the labour market and limited utilization of production capacities. The anti-cyclical policy conducted in this way, convergent with anti-inflationary policy, can shorten the phase of a high dynamics of growth and decrease its amplitude and intensity.

The activities of the central bank presented above, dependent on the level of economic activity, through changes in the liquidity of commercial banks and changes in the supply of credit money, should lead to changes in the costs of access to credit money, i.e. to the growth or decline in the price of credits and to changes in the demand for money. In the related literature dominates the opinion that effectiveness of the central bank's influence on the real economic processes through changes in credit interest rates is getting smaller and smaller[1]. Implementation of a restrictive or expansive monetary policy and the resulting

[1] See: Markowski (1992), p. 66.

changes in interest rates do not have to determine the demand for credit at all and exert a positive or negative influence on investment activity. This is caused by the fact that[1]:

- growth of interest rate on credits granted by commercial banks under high economic activity very frequently coexists with the growth of interest rate on bank deposits. This may cause a speculative flow of capital e.g. from abroad and contribute to the growth of cash reserves in commercial banks and their increased credit-granting capacity which is contradictory to the assumptions of the implemented restrictive monetary policy of the central bank;
- high interest rate on credits may, contrary to intentions of the central bank, cause sustained or growing inflation. More expensive credits may lead to increased costs of economic activity which can be compensated by a rise in prices;
- if monetary and fiscal policies are not synchronized, a high interest rate used by commercial banks can be levelled by the existing tax system and investment reliefs;
- various sectors of the economy and particular enterprises react differently to changes in interest rates; multinational corporations are the least affected as they can finance their activities from their own means (self-financing). Most sensitive to the changes in interest rates are investment decisions made by small and medium-size firms belonging to the constructing or the trade sector;
- quickly introduced technological progress and high profitability are responsible for the fact that investment activity yields sufficiently good results to cover high costs of credit;
- changes in the functioning of banking systems and transformation (internationalization) of credit markets lead, among others, to the fact that a higher interest rate in one country results in a bigger amount of credits incurred in other countries where the interest rate is lower;
- changes in interest rate very frequently cause changes in exchange rates, which is an additional factor influencing changes in economic activity. Increased interest rate can contribute to revaluation as well as to decline in exports and growth of imports along with all the negative effects for the economic processes in a given country;
- changes in the price of money very frequently lag behind those in the level of economic activity.

Therefore, changes in interest rates influence the development of optimistic or pessimistic climate in decision-making processes. They are more of a psychological factor determining the behaviour of economic entities and provide information on the intentions of a central bank as to the monetary policy implemented.

[1] Ibidem, p. 66 and further on.

Lower effectiveness of the influence of interest rate on short- and medium-run processes in the real sphere as well as frequently incorrect information resulting from these mechanisms caused that followers of the monetarist economics, using the so-called Gibson's paradox and Fisher effect, underline the importance of money supply in the implemented monetary policy[1]. If under low economic activity expansive monetary policy is carried out, then after initial decline in interest rate due to the so-called Keynesian effect of liquidity, in the following periods this interest rate rises as a result of increased demand for bank credit. This increase can be interpreted as the effect of restrictive policy whereas the real reason for it was expansive policy. Fisher's effect in turn, means that the policy of easy money intensifies inflationary expectations and leads to the growth of nominal interest rate. Disappearance of the effect of "money illusion" aggravates current inflationary processes, increases inflationary expectations and raises the nominal interest rate. Therefore, changes in interest rate can lead to erroneous conclusions formulated by monetary authorities, so monetarists emphasize the role of money supply in monetary policy.

Another significant question in the monetary policy is how the supply of money is shaped. Out of the two existing possibilities: indirect influence (fiscal policy, interest rates, exchange rates) and direct influence oriented towards changes in monetary base, monetarists prefer the latter. As the monetary base includes high powered money, i.e. cash and coins in circulation as well as monetary means of commercial banks on accounts in the central bank (reserve money) its change by one unit can cause multiple changes in the supply of money on the market. When the economic activity changes, the highest flexibility, i.e. capability to adjust to the growth or decline of the real values, should be exhibited by the categories which measure the supply of money of the highest liquidity (M0, M1). Reactions of categories M2 and M3 should be more weakly correlated with the growth or decline in business conditions.

An important factor which determines the supply of money and at the same time hinders its regulation is changes in money circulation and whether they are of permanent or temporary character[2]. Any errors made here can significantly influence the real economic situation.

According to Keynesian economists the influence of monetary policy on business fluctuations can have a slight effect as, under expansive policy implemented when economic activity is low, the so-called "liquidity trap" appears[3]. This means that with a low interest rate the growth of money supply will not cause a further decline of interest rate. This phenomenon is connected with slower circulation of money. Another factor diminishing the effectiveness of monetary policy can be the mechanism of making investment decisions by

[1] Ibidem, p. 70.

[2] Ibidem, p. 72.

[3] See: Krupa (1976), pp. 154—157.

enterprises. In this process the most important role is played by expectations of economic entities about the future business conditions and the current interest rate is of much lesser importance. Enterprises' pessimism about the future can reduce propensity to invest more effectively than the price of credit Optimistic forecasts as to the economic situation can contribute to the growth of investment demand despite the fact that interest rate increases. Time lags are also a significant restriction of the monetary policy. The period which passes from the monetary impulse sent by the central bank to changes in real magnitudes is relatively long and positive effects are differentiated in time[1]. In this case the moment of implementation and adequate intensity of the instruments are of particular importance.

Monetarists postulate that in the process of stabilizing business conditions, first of all the instruments of monetary policy should be applied because there exists a direct connection between money supply and level of economic activity[2]. Assuming Fisher equation, they are of the opinion that velocity of money circulation is constant and does not depend on changes in its amount. According to Friedman, this velocity depends on, among others, the structure of the money market and producers' and consumers' habits. The growth in money supply under constant velocity of circulation and constant level of production (depending on structural factors, e.g. production capacities) must lead to the growth of average transaction price, not to the growth of production and employment. Therefore, expansive monetary policy (increase in money supply) is not able to overcome the phase of low dynamics of growth because it only causes increase in prices, inflationary expectations and nominal interest rate. Moreover, it does not increase employment the level of which depends on structural factors. The size of employment is determined by natural rate of unemployment which defines maximum employment without causing inflation under given labour demand and supply and under zero rate of growth in labour productivity.

According to monetarists, interpretation of dependencies between inflation rate and unemployment rate differs significantly from the views of Phillips. They claim that substitution between those phenomena occurs only within a short period of time. In a long run this dependence is shown as a straight line vertical in relation to X-axis which means that a given size of employment can be reached under any level of prices[3]. Thus, stabilization policy should not aim at reducing the rate of unemployment below the natural level because this will increase inflation when there is an increased level of uncertainty in the economy.

Monetarists claim that conducting an effective anti-cyclical policy by means of monetary policy instruments is very difficult. Even if one assumes the

---

[1] Very frequently only after 12 months pass from the recognition that certain activities of the monetary policy should be taken there appear effects in the real sphere. More on the subject see: Markowski (1992), p.77.

[2] More information in: Friedman, Schwartz (1963), pp. 60—62.

[3] More information, among others, in: Winiecki (1986), p. 22.

possibility of change in the velocity of money circulation in a short- and medium-run, it is absolutely impossible to be predicted accurately. Wrong monetary policy leads to economic instability, therefore it should be reduced to controlling the money supply in order to ensure a constant rate of its increase, equal to the dynamics of potential growth in GDP and the assumed rate of inflation.

A significant difference between advocates of Keynesian economics and monetarists concerns also the principles of procedure: should monetary policy be based on fixed principles e.g. fixed rate of growth in money supply, or should it be changed and adjusted to the existing and anticipated economic situation[1]. This argument concerns the question whether an active monetary policy should be implemented or whether a passive policy should be carried out, the one with rigid procedures. Keynesian economists postulate the active monetary policy which should not be restricted by any fixed rules, however each time it should be adjusted to the existing and anticipated economic situation in order to effectively counteract business fluctuations. Monetarists claim that it is necessary to base stabilization policy on fixed principles unchangeable within long periods of time. Therefore, while implementing monetary policy the central bank should concentrate on stabilization of inflationary expectations which will require the principle of constant increase in money supply. Economic stability and a possibility to forecast exactly the increase in money supply will occur when this supply will change in a long period of time at the rate convergent with the dynamics of growth in real output.

3. Identification and analysis of the morphology of business fluctuations in the Polish economy in the years 1990—2003

In the empirical analyses of business fluctuations in economic system under transition, the main problem (apart from definition of oscillations) is selection of statistical data used in the research. This paper assumes that the measures of economic activity will be GDP and sold production of industry (SPI)[2]. Both indicators analysed quarterly are expressed in fixed prices of 1995.

The related literature contains different definition of cycles which sometimes supplement one another but frequently they are mutually exclusive[3]. In this paper Lucas approach was applied. According to this approach the essence of business oscillations is a process of repetitive fluctuations around long-run macroeconomic trends of the time series[4]. This indicates complexity of the business phenomena and therefore a necessity to study a few or more variables explaining the structure

---

1 Markowski (1992), pp. 73—76.

2 More information on estimation of the analysed indicators in Poland see: Barczyk, Kruszka (2003), pp.42—43.

3 More on this subject see: Zarnovitz (1985), pp.523—580.

4 See: Lucas (1977), p. 7.

of the cycle. Secondly, of key importance for the identification of fluctuations in economic activity seems to be determination of long-run trend.

Having a series which expresses the trend of reference index, it is possible to determine deviations of the empirical values from it. Having accepted such approach, the authors decided that oscillations identified with business fluctuations will be determined by the indicators informing about a relative deviation of the GDP or industrial production from a long-run trend estimated for their levels[1]. This means that in our paper business fluctuations are defined as a deviation cycle.

Identification of the cyclical components by means of deviation method creates doubts because the selection of function describing the trend is arbitrary, which may deform the results obtained. This problem, however, concerns only a deterministic trend. In reality a vast majority of macroeconomic magnitudes are the processes of random walk with drift[2]. This means that it is necessary to apply stochastic methods facilitating identification of the growth tendency which is subject to changes in time. To achieve this objective Hodrick — Prescott's (HP) filter was used[3].

According to the above-mentioned procedures of identification of cyclical oscillations, deviations of the real GDP and sold production of industry from their trends were estimated. Both series are presented in Fig.1.

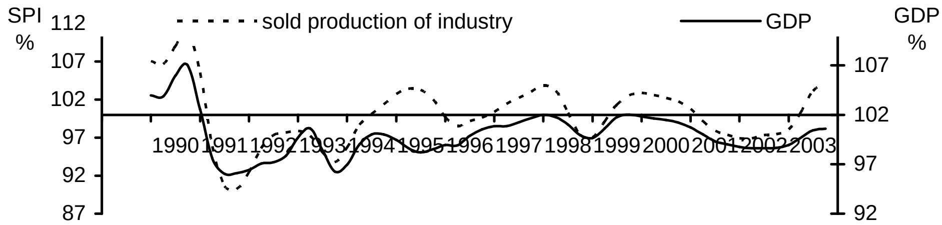
The fluctuations presented cannot be interpreted univocally when identifying the turning points. This is so because successive turning points occur one after another and this is commonly regarded as a phenomenon characteristic of contemporary business cycles subject to considerable deformation as compared with the classical cycles. Therefore, while finding such turning points one should make *a priori* assumptions which in the empirical analyses are dictated by practical aspects.

[1] A seasonal adjustment was carried out before estimation of the growth tendency and deviations from the trend.

[2] See: Nelson, Plosser (1982), pp. 139—162.

[3] Hodrick, Prescott (1997), pp. 1—26.

Figure 1. Fluctuations of GDP and sold production of industry in Poland in the years 1990—2003



Source: Authors' own calculations.

A trough can be identified only when it is preceded by a peak (and *vice versa*). When in the series there appear double peaks or troughs (of relatively equal character), then the moment of the first peak is usually chosen as a turning point. If, however, between possible turning points appear falls (rises) in the values of appropriate indicators, and both sites cannot be regarded as relatively equivalent, then the moment where maximum (minimum) values of a given measure of economic activity appears is regarded as the turning point[1]. Moreover, it should be assumed that particular phases have to last at least for 3 quarters, otherwise this type of dynamic changes may result from purely accidental phenomena, unrelated to business conditions.

If we know when and where the turning point appears, we can also determine amplitudes. This analysis assumes that the amplitude of a phase is absolute difference between the values of extreme deviations occurring within a given period, whereas the amplitude of a cycle is a difference between the amplitudes of growth and decline phases. Using the above-mentioned criteria, turning points were determined in the analysed measures. Next the length of the phases and the cycle as well as the corresponding amplitudes were determined. Table 1 contains relevant information on this subject.

In both cases the first turning point was identified in forth quarter of 1990, so it is located close to the beginning of the analysed series which may mean that its identification is not completely univocal. If, however, the differentiated extremes are accepted, it is possible to state that in the years 1990-2003 in the Polish economy there appeared five complete phases of the cyclical process.

The first one is a phase of decline in business conditions (from forth quarter of 1990 to 1991:2 — according to the GDP, or to 1991:3 — according to dynamics of industrial production). This period witnessed a trough of the Polish transformation recession. Subsequent years saw improvement in the level of economic activity, however, it should be remembered that initially there was a slowdown of decline and only later growth in the real value of measures of

[1] Barczyk, Kowalczyk (1993), p.22.

economic activity. The peak which completed the phase of growth fell on the turn of 1997—1998.

Subsequent quarters are a phase of low dynamics with phenomena indicating a slowdown of economic growth. The trough which seems to have finished that phase was identified in fourth of 1998. It was followed by a 3 or 4 — quarter long phase of improvement in business conditions and at the beginning of the year 2000 another decline phase started to last until the second half of the year 2002.

Table 1. Specification of certain morphological features of business fluctuations in Poland's economy in the years 1990—2003

Morphological features	Phases of business cycles				
	decline phase	growth phase	decline phase	growth phase	decline phase
			Gross Domestic Product		
Duration	1990:4—1991:2	1991:3—1997:4	1998:1—1998:4	1999:1—1999:3	1999:4—2002:3
Length (in quarters) - phase	3	26	4	3	11
- *cycle*	29		7		
Amplitude (in per cent points) - phase	10,95	5,97	2,35	2,37	3,41
- cycle	4,98		0,02		
Coefficient of volatility (in per cent)	2,08				
			Sold production of industry		
Duration	1990:4 – 1991:3	1991:4 – 1997:4	1998:1 – 1998:4	1999:1 – 1999:4	2000:1 – 2002:2
Length (in quarters) - phase	4	25	4	4	10
- *cycle*	29		8		
Amplitude (in per cent points) - phase	20,83	13,64	6,80	5,84	5,58
- cycle	7,19		0,96		
Coefficient of volatility (in per cent)	4,20				

Source: Authors' own calculations

Assuming the existence of five phases, it is possible to group them into two complete cycles. The first one lasted for slightly over 7 years and began with a phase of low dynamics of growth where special attention should be paid to its very big amplitude exceeding 20 per cent points (as regards industrial production) or almost 11 — if GDP is taken as a measure of activity. The observed rate of decline was very fast, unusual in well developed economic systems, which certainly should be related to the implementation of economic reforms. The following phase of growth was much longer but with a smaller amplitude. On these grounds a conclusion can be formulated about a positive amplitude of the cycle and its asymmetry both as regards duration and amplitude.

The second cycle has different morphology as compared to the first one. Altogether it only lasted for two years and the length of its component phases was almost equal. Moreover, fluctuations amplitudes in particular sub-periods were very similar to each other. A subsequent cycle starting from the phase of decline has not finished yet, although one can see that it is going to be much longer than the previous one because the period of unfavourable business conditions exceeded two years.

These morphological features show that in the process of growth the Polish economic system exhibits significant cyclical components. Their repetitive character, however, can by no means be identified with stability of structure either as regards length, frequency or amplitude of fluctuations. This indicates deformation of oscillations characteristic of contemporary business cycles, which in the case of Poland results both from a specific character of transformation and the implemented policy of stabilization.

4. Business fluctuations and monetary policy instruments in Poland under transformation

Identification of changes in the economic activity measured by means of fluctuations in the real sold production of industry and GDP facilitates their comparison with the phenomena observed in the monetary sphere. To do so it is necessary to analyse the magnitudes which seem to be the most significant from the viewpoint of the monetary policy implemented.

According to the theoretical concepts presented in the first part of the paper, the authors take into consideration both money supply and interest rates. Central banks, determining the money supply, group particular monetary assets by adding their values. In this way aggregates are created which are usually denoted by letter *M* with a numerical figure. The higher the figure at *M*, the less liquid are the components in the constructed aggregate [1]. Today the supply of money is described by means of three aggregates: M1, M2 and M3. Somewhere around

[1] Apart from classical aggregates constructed by adding the values of their components, the weighted indicators of DIVISIA type are estimated. The National Bank of Poland (NBP) stopped constructing them in March 2002. See: Cieśla (1999).

there is monetary base (sometimes defined as M0), i.e. the most liquid measure which is a sum of cash and obligatory reserves.

In Poland until February 2002 the widest measure was M2 treated as a sum of cash in circulation, current deposits and all kinds of long-term deposits, deposit certificates and negotiable securities. However, in March 2002 The National Bank of Poland implemented ECB standards[1]. Thus in Poland banking statistics started to use aggregates M0, M1, M2 and M3, defined according to the schedule binding in the EU. This was a significant problem when selecting the variables describing the money supply because the change in classification made it impossible to gather longer series aggregated on the same basis. Due to this the set of monetary variables was limited, the period of analysis was shortened to the years 1993—2003 and only M0, M1 and M2 were used.

In the case of interest rates the analysis included interest on NBP credits, i.e. rediscount rate and lombard rate as well as interest rate on deposits on the Warsaw inter-bank market (WIBOR) for the period of three months. The level of WIBOR is identified with the market cost of capital. Apart from the above-mentioned magnitudes the analysis included interest on deposits in PLN *a vista* and 12-month deposits.

The measures of the amount of money were converted to fixed prices (1995=100), whereas interest rates were analysed on a yearly basis. Monetary indicators (with the exclusion of rates on NBP credits) were subject to statistical decomposition which facilitated the identyfication of the cyclical elements, i.e. they were subject to seasonal adjustment and deviations from the trend estimated by means of HP filter were determined.

The series prepared in the above-mentioned way were used first of all to study correlation with fluctuations in business activity. To achieve this Pearson's correlation coefficients were calculated and then time shifts up to 8 quarters were introduced[2]. It was assumed that longer transpositions would lose the value of significance of the studied links. The results obtained are presented in Tables 2 and 3.

The correlation coefficients inform that there exist visible links between oscillations in business activity and fluctuations in money supply, however, their force and character depend on the category of money and a referential variable. In the case of SPI the strongest correlation appears between fluctuations in this aggregate and the money supply from the central bank (M0) as well as M1, which indicates a strong position of settlements made in the narrow money. As regards GDP a much stronger link is shown with broad money measure M2.

[1] Biuletyn Informacyjny NBP (2002) No.3, p.5.

[2] In further research the shift showing the highest absolute value of the correlation coefficient was taken into account.

Table 2. Matrix of correlation between fluctuations of GDP, SPI and oscillations of monetary variables (without time lags) in Poland in the years 1993—2003

Explained variables	Explaining variables							
	M0	M1	M2	RR	LR	ADI	12MI	WIBOR3 M
GDP	**0,22**	**0,44**	**0,50**	0,10	0,08	0,00	**0,35**	**0,30**
SPI	-0,05	0,17	-0,16	0,12	0,10	**0,44**	**0,34**	**0,39**

Explanations: RR – rediscount rate; LR – lombard rate; ADI – interest rate on a vista deposits; 12MI – interest on 12-month deposits; WIBOR3M – three-month interest rate on deposits on the Warsaw inter-bank market. The values of correlation coefficients significant at level 5% are printed in bold.

Source: authors' own calculations

Table 3. Matrix of correlation between fluctuations of GDP, SPI and oscillations of monetary variables (time lags included) in Poland in the years 1993-2003

Explained variables	Explaining variables							
	M0	M1	M2	RR	LR	ADI	12MI	WIBOR3M
GDP	**0,63**	**0,67**	**0,71**	**-0,40**	**-0,37**	0,11	**0,53**	**0,51**
	(+5)	(+3)	(+4)	(-4)	(-4)	(+2)	(+2)	(+2)
SPI	**0,50**	**0,50**	**0,45**	-0,23	-0,20	**0,58**	**0,41**	**0,47**
	(+6)	(+4)	(+5)	(-4)	(-4)	(+1)	(+1)	(+1)

Explanations: numbers in brackets inform about time shift of the explaining variables. Mark (+) shows the lag calculated in quarters compared with fluctuations in GDP or SPI, whereas (-) indicates the lead.

Source: authors' own calculations

The time shifts applied suggest that oscillations appear first in the real sphere and then the turns occur in fluctuations of particular *M*. Therefore, in Poland the changes in economic activity correspond after some time to the increasing amount of money in circulation. Similar patterns were observed for all the measures of money, although transpositions changed from 2 to 6 quarters.

Analysing the links between cyclical fluctuations and interest rates it can be noticed that the NBP rates are more strongly correlated with fluctuations in the GDP than with oscillations in the sold production of industry. Moreover, this is a relation of reverse character and the NBP rates lead the changes in GDP. This can suggest that general business conditions are sensitive to the steps taken in the sphere of interest rate policy. However, this result has to be interpreted cautiously. There is a possibility of apparent spurious correlation resulting from nonstationarity of the series of interest rates on the NBP credits[1]. Statistical decomposition of these rates, facilitating identification of the cyclical components, was not justified because their administrative determination hinders the application of the previously used instruments.

This type of objections did not concern other interest rates. In that case it was assumed that these were pro-cyclical variables in relation to oscillations from the real sphere (fluctuations in the GDP and a vista deposits were an exception). Moreover, it should be noticed that the extent of pro-cyclical character increased when interest rates lagged behind.

The next step in this research was estimation of regression equations. Fluctuations in GDP or sold production of industry were assumed as dependent variables. First the set of independent variables included oscillations in money supply then — only changes in interest rates and finally equations considering all the analysed magnitudes. In the estimated equations time series were used — shifted according to the results of correlation measurements. Equations (1) — (6) present the polynomials obtained in this way[2].

(1) $GDP_t = \mathbf{73{,}748} - 0{,}002\ M0_{t+5} + \mathbf{0{,}137\ M1_{t+3}} + \mathbf{0{,}128\ M2_{t+4}};\quad R^2 = 0{,}800$

(2) $GDP_t = \mathbf{100{,}924} - 0{,}471\ RR_{t-4} + 0{,}385\ LR_{t-4} - 0{,}255\ ADI_{t+2} + 0{,}276\ 12MI_{+2} + + 0{,}099\ WIBOR3M_{t+2};\quad R^2 = 0{,}400$

(3) $GDP_t = \mathbf{78{,}687} - 0{,}021\ M0_{t+5} + \mathbf{0{,}093\ M1_{t+3}} + \mathbf{0{,}138\ M2_{t+4}} - 0{,}382\ RR_{t-4} + 0{,}359\ LR_{t-4} - 0{,}351\ ADI_{t+2} + \mathbf{0{,}152\ 12MI_{t+2}} + \mathbf{0{,}167\ WIBOR3M_{t+2}};\quad R^2 = 0{,}872$

(4) $SPI_t = \mathbf{77{,}328} - 0{,}019\ M0_{t+6} + \mathbf{0{,}242\ M1_{t+4}} - 0{,}032\ M2_{t+5};\quad R^2 = 0{,}333$

(5) $SPI_t = \mathbf{97{,}320} - 3{,}032\ RR_{t-4} - 2{,}627\ LR_{t-4} + \mathbf{1{,}147\ ADI_{t+1}} + 0{,}145\ 12MI_{t+1} + - 0{,}170\ WIBOR3M_{t+1};\quad R^2 = 0{,}339$

(6) $SPI_t = \mathbf{70{,}452} - 0{,}042\ M0_{t+6} + \mathbf{0{,}362\ M1_{t+4}} + 0{,}020\ M2_{t+5} + 0{,}766\ RR_{t-4} - 0{,}706\ LR_{t-4} + \mathbf{1{,}681\ ADI_{t+1}} + 0{,}005\ 12MI_{t+1} - 0{,}341\ WIBOR3M_{t+1};\quad R^2 = 0{,}628.$

All of the obtained equations were statistically significant. However, it should be remembered that the starting source material covered 44 quarters and the introduction of time lags still reduced the set of observations. Therefore, the results obtained must be interpreted cautiously and treated as a starting point for further studies.

The highest coefficients of determination were obtained in the case of modeling of the GDP oscillations because it turned out that about 85 per cent of the changes in the economic activity in Poland can be explained by means of fluctuations in the monetary magnitudes included in the analysis (see: equation

[1] See: Syczewska (1999), pp.37—42.

[2] Variables significant at 5 per cent significance level are printed in bold.

(3)). The goodness-of-fit of the models obtained for fluctuations in sold production of industry should be evaluated as being much worse.

Analysing the variables which describe the amount of money it can be seen that information from the correlation calculus is confirmed. Widely defined money (M1, M2) exert a stronger influence on the GDP, whereas the narrow M1 is the only significant determinant of the oscillations of SPI (equation (4)). In the case where the amount of money could be regarded as an instrument linked with the changes in economic activity, its proportional influence on referential variables was proved, which is convergent with the hypotheses formulated on the grounds of monetarism.

Equations (2) and (5) show that interest rates generally cannot be regarded as autonomous instruments of shaping economic activity 1. However, a joint study on the influence of the amount of money and interest rates (relations (3) and (6)) presents certain links between interest rate on a vista deposits and changes in GDP and SPI.

It should be underlined that monetary policy of the central bank which would apply only the instruments controlled by this bank has not an anti-cyclical character in Poland. This is proved by the fact that oscillations in the GDP and SPI are not significantly determined by fluctuations in M0 and changes in NBP interest rates on credits. On the other hand, however it would seem that monetary authorities can indirectly influence the economic activity by shaping the supply of broad money. For example the policy of obligatory reserves and open market operations are the instruments which make it possible, although partially, to influence the creation of money by a commercial banking system. An obstacle to this type of interpretation, however, is the lagging character of oscillations of monetary aggregates as compared with changes in business conditions.

Very similar conclusions can be drawn as regards the links between market interest rates and fluctuations in the GDP and sold production of industry. The central bank can indirectly determine liquidity of the banking sector which certainly influences the price of money on the inter-bank market (WIBOR). However, once again the lagging (not the leading) character of changes in this indicator is visible, moreover its oscillations are pro-cyclical. This means that the monetary authorities do not pursue an active counter-cyclical policy.

Conclusions

The conducted research shows that Poland's economy under transformation is an economic system where significant cyclical fluctuations can be observed. Their morphology is similar to the characteristics of contemporary business fluctuations exhibiting a considerable deformation in comparison with classical

[1] See: Orłowski (2000), pp.39—40.

definitions of business cycles. The morphological features of oscillations in business conditions prove that indicators of sold production of industry are much more sensitive to business fluctuations than those of GDP.

Taking into account relations between fluctuations in the real indicators and the magnitudes regarded as instruments of monetary policy, there is a possibility to verify their links by Keynesian and monetarist approaches. The results of statistical analyses seem to show fairly univocally that the Polish monetary policy is not pro-Keynesian. This is proved by the fact that there is no significant influence of the central bank interest rates on changes of the economic activity and lagging character of the market rates shaped indirectly by the decisions of the monetary authorities.

On the other hand, many facts prove that activities of the Polish central bank were carried out according to the monetarist theory. It is confirmed by a significant positive correlation between fluctuations in money supply and indicators in the real sphere of the economy, with a simultaneous time lag of the former. This can be interpreted as a clearly visible implementation of the fixed rules of a passive monetary policy linking the rate of changes in money supply with the dynamics of the GDP. Moreover, a pro-cyclical character of the price of capital corresponds to the conclusions derived from the neoclassical school[1].

The presented analysis shows, therefore, that the NBP's priority was not to counteract excessive changes in the economic activity. Its policy was clearly oriented towards curbing inflation. However, it should be underlined that this is one of the aspects of a correctly implemented stabilization policy of widely approached business conditions. This policy should be supplemented with instruments from the fiscal sphere which, however, goes beyond the research range assumed, although there is no doubt that further research should be conducted in this field.

REFERENCES

BARCZYK R., M. KRUSZKA (2003), Morphology of Business Fluctuations in the Polish Economy under Transition, [in:] Business Cycle Diagnosing in Poland, eds. K. Piech, S. Pangsy-Kania, DW Elipsa, Warszawa, pp. 34—58.

BARCZYK R., Z. KOWALCZYK (1993), Measurement of Business Cycle, PWN, Warszawa-Poznań.

BARRO R. J. (1997), Makroeconomics, PWE, Warszawa.

BEDNARCZYK J. L. (1990), Monetary Policy against Inflation. The Case of Capitalist Economy, PWN, Warszawa.

[1] More on the subject: Barro (1997), p.271; Kruszka (2003), pp. 179—180.

CIEŚLA N. (1999), Construction of DIVISIA Monetary Aggregates in Poland, Materials and Studies No 89, NBP, Warszawa..

DRABOWSKI E. (1987), Money Influence on Economy in Keynesian and Monetarist Theory, PWN, Warszawa.

FRIEDMAN M., A. J. SCHWARTZ (1963), Money and Business Cycles, Review of Economics and Statistics, vol. 45, pp. 32—64.

INFORMATION Bulletin NBP (2002), No 3, Warszawa.

KRUPA A..(1976), Wachstumspolitik. Wachstumskriesen und die Möglichkeit stabiler Wirtschaftsentwicklung in der Bundesrepublik Deutschland, R. Oldenbourg Verlag München-Wien.

KRUSZKA M. (2003), Fluctuations of Economic Activity and Changes of Prices and Interest Rates in Poland [in:] Business Cycle Diagnosing in Poland, eds. K. Piech, S. Pangsy-Kania, DW Elipsa, Warszawa 2003, pp.175—192.

KRUSZKA M. (2002), Business Cycle and Changes on the Money Market, AE Poznań.

LUCAS R. E. (1977), Uderstanding Business Cycle, reprint [in:] Business Cycle Theory, ed. F. E. Kydland, Aldershoot-Brookfield.

MARKOWSKI K. (1992), The Role of Government in Market Economy, PWE, Warszawa.

NELSON C. R., C. I. PLOSSER (1882), Trends and Random Walks in Macroeconomic Time Series; Some Evidence and Implications, Journal of Monetary Economics, vol. 10, pp. 139—162.

ORŁOWSKI W. M. (2000), Has the economic policy in Poland in the 1990s had an anti-cyclical character?, ZBSE GUS and PAN Research Bulletin, vol. 9, No 3—4, pp. 29—42.

SYCZEWSKA E. M. (1999), Long–run Relationships Analysis: Estimation and Veryfication, Wydawnictwo SGH, Warszawa.

WINIECKI J. (1986), The Sources of Inflation in Market Economy and Planed Economy, PWN, Warszawa.

ZARNOVITZ V. (1985), Recent Work on Business Cycles in Historical Perspective: A Review of Theories and Evidence, Journal of Economic Literature, vol. 23, s. 523—580.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 239—258

CONSUMER SURVEYS AND REALITY

Maurizio Bovi[1]

ABSTRACT

This paper investigates the usefulness of Italian consumer surveys as estimation and forecasting tool. To this end, standard consumption equations are estimated and then compared, in terms of in-sample and out-of-sample predictive ability, with corresponding models which differ from them only for the presence of the confidence indicator. Unlike mainstream literature, the present work focuses on the relationships between subjective and objective information at a disaggregated level. The idea behind is to check if one sub-index is more or less informative than another, and if some outlay is more or less "sentiment sensitive" than another. Working with disaggregated data, it is shown that the qualitative information obtained from household surveys improves both the goodness-of-fit of consumption equations and their forecasting performances. For instance, as somewhat expected, a serene mood lead to an increase in the consumption of recreation services, while medical care expenses are not driven by psychological motivations.

JEL classification: C22, C52, D12, E21, E27.

Key words: consumer sentiment; surveys; consumption; time series.

1. Introduction

Surveys of people in households add a psychological dimension that can be tentatively quantified for use in the study of economic systems. In fact, the effect of consumers attitudes on economic activity is a subject of great and growing interest among policymakers and economists. For instance, the worsening of households confidence has often been reported as one of the main causes of the early '90s world-wide recession. The peculiar nature of this subjective

---

[1] Institute for Studies and Economic Analyses (ISAE), Department of Macroeconomics, Piazza dell'Indipendenza, 4 — 00185 Roma — Italy. E-mail: m.bovi@isae.it. I would like to thank Roberto and Valentina for their patient assistance and the participants of the 27th CIRET Conference of September 2004 in Warsaw for helpful comments. The opinion expressed herein are solely those of the author and not necessarily represent the views of the ISAE.

information suggests to accurately analyse its advantages and drawbacks, on the one hand, and a careful use of it, on the other hand. Theoretically, the psychological dimension has never had a relevant role in economic literature. Excluding the attempts by Katona (1951, 1960) and the Keynesian allusion to “animal spirits”, households’ and firms’ behaviours have always been studied without considering their emotional attitude to act. Needless to say, it is very hard to gather reliable information about this latter. The mainstream solution is based on directly interviewing agents. The results of these surveys are then arranged into “balances”, which are the differences between the percentages of positive and negative answers. The Italian consumers’ sentiment index monthly worked out by ISAE[1] is based on a representative sample of 2,000 respondents, and concerns the consumers’ expectations on the overall economic situation and on their own financial situations. It is computed as the arithmetic average of the (seasonally adjusted) net balances of positive minus negative responses to nine questions on: (1) anticipated personal financial conditions over the coming year; (2) anticipated economic conditions over the coming year; (3) anticipated job availability conditions; (4) whether now is a good or bad time to save; (5) personal financial conditions over the past year; (6) economic conditions over the past year; (7) personal opinion regarding the household’s budget; (8) personal saving possibilities over the coming year; (9) whether now is a good or bad time to buy major household items.

Provided that consumers’ sentiment indices (CS) have additional information content about the economic situation beyond that already contained in quantitative variables, they can be very helpful because of the timeliness of their availability. The ISAE index is diffused in the first days of the last decade of the reference month, whereas the promptest quantitative measure (the GDP preliminary flash estimate) is released by the Italian National Institute of Statistics (ISTAT) with a delay of forty days. Despite the popularity of the qualitative indices, there is little consensus about their ability to collect information on consumer spending that is not already captured by economic fundamentals. In response to the widespread belief that consumers’ opinions and expectations influence the direction of the economy, a growing number of studies have set out to analyze the relationship between consumer attitudes and economic variables. So far, the literature on the predictive power of CS have showed mixed conclusions (Leeper, 1992; Carroll, Fuhrer, and Wilcox, 1994; Matsusaka and Sbordone 1995; Berg and Bergstrom, 1996; Batchelor and Dua, 1998; Bram and Ludvigson, 1998; Golinelli and Parigi, 2003).

In order to reflect and represent total economic activity, surveys data are used to calculate a composite indicator. Thus, in the literature this latter overall measure is often compared with the performance of a reference variable which is also all inclusive. The GDP is the obvious choice, but the CS has often been

[1] For a description of the ISAE Consumers’ confidence indicator, see Martelli (1998).

related to total private consumption as well. However, it can be useful to invert the process, that is to collate subjective with objective information at a less aggregated level. In that is the main novelty of this paper, which is the first formal empirical investigation of household attitudes associating four different sub-indices with six categories of consumer spending. The idea behind is to check if one emotional measure is more/less informative than another, and if some outlay is more/less “sentiment sensitive” than another. In the former case it may happen that even if the overall index turns out to be uninformative, it could be not true for some of its components. In the latter case one can expect that a serene mood could lead to an increase in the consumption for, *e.g.*, recreation, while expenses incurred for medical care should be driven by psychological motivations only, if any, to a lesser extent.

From the methodological point of view I use a two-step procedure to determine the in-sample and out-of-sample informative power of CS. First, I consider a standard baseline equation for consumption growth that does not include attitudinal survey measures. Then, I add consumer sentiment indicators to the baseline equation and test which CS, if any, improve the goodness-of-fit (adjusted R-squared) and/or the root mean square error (RMSE) of the benchmark model. As in most of the analyses in the field (for an exemption, see Golinelli and Parigi, 2003) the statistical framework is univariate. This may raise simultaneity bias issues, but in the present empirical design the Johansen procedure suggests the existence of, at most, only one cointegration vector. This means that the single-equation approach can be validly pursued because the weak exogeneity requisite holds (Engle *et al.*, 1983). Within the single equation framework, I follow the general-to-specific approach in its Wickens-Breusch (1988) form which i) increases the efficiency of the estimates; ii) allows testing the statistical significativity of the error correction term in a single step; iii) is particularly suited in cases, such as the present, dealing with both theoretical and empirical information. Also, since the residuals are homoschedastic innovations and the parameters are verified to be constant, the baseline models are empirically admissible and coherent (Hendry, 1995). Therefore they are both reliable and strong contenders for the CS-augmented models.

Results show that the qualitative information obtained from household surveys improves both the explained variance and the RMSE of the baseline consumption equations. These improvements are all the more evident when working with disaggregated data, *i.e.* linking a particular kind of consumption to a particular sub index. For instance, consumption for services seems to be sensitive to households’ personal conditions and expectations in the sense that (alternatively) adding these sub-indices in the relative baseline model one may improves both the in sample and the out of sample performances of the baseline equation. This result does not hold for other indices, in particular for the most aggregate one which, actually, shows the lowest predictive ability. In other words, while there is not a clear winner there is an identifiable looser, the overall index.

As refer to the health care outlay, psychological measures result useless in forecasting exercises. It can be seen as an indirect clue that in order to better understand the informative content of CS a careful treatment of macroeconomic data is needed. Furthermore, as pointed out by the literature (for Italy, Parigi and Schlitzer, 1997), subjective indicators help to shed some light about extraordinary economic episodes. Sub-sample analyses show that throughout the two-recessions period 1992—2003 some CS sub-indices increase their forecasting ability while, once again, it does not hold for the overall indicator. Finally, the out-of-sample experiments performed over the last four years (2000—2003) show better results when using some sub indices. Thus, even during the recent puzzling behaviour of the Italian climate measures, languishing in a very low level as opposed to a relatively sustained growth in the total consumption, a disaggregated analysis is still able to shed some light about the predictive ability of qualitative measures.

The paper is organised as follow. The next section describes the surveys data, Sections 3 and 4 outline the empirical framework, Section 5 collects the findings. Concluding remarks close the paper.

2. Surveys Data

The monthly qualitative data are drawn by ISAE and cover the sample period 1982.01 to 2003.12. Since macroeconomic data are quarterly, I converted the monthly frequency of ISAE data by a simple average. As mentioned, the overall index (1980=100) is formed by nine questions. Following the ISAE procedure, I exhaustively aggregate these latter into two pairs of sub-indices. The first couple deals with present and future situations; the other pair distinguishes between personal and economic conditions in the country as a whole. More in detail, the questions are collected in the following way:

- Personal:
  1) anticipated personal financial conditions over the coming year;
  4) whether now is a good or bad time to save;
  5) personal financial conditions over the past year;
  7) personal opinion regarding the household's budget;
  8) personal saving possibilities over the coming year;
  9) whether now is a good or bad time to buy major household items;
- General
  2) anticipated economic conditions over the coming year;
  3) anticipated job availability conditions;
  6) economic conditions over the past year.
- Present:

4) whether now is a good or bad time to save;
7) personal opinion regarding the household's budget;
9) whether now is a good or bad time to buy major household items;
5) personal financial conditions over the past year;
6) economic conditions over the past year;

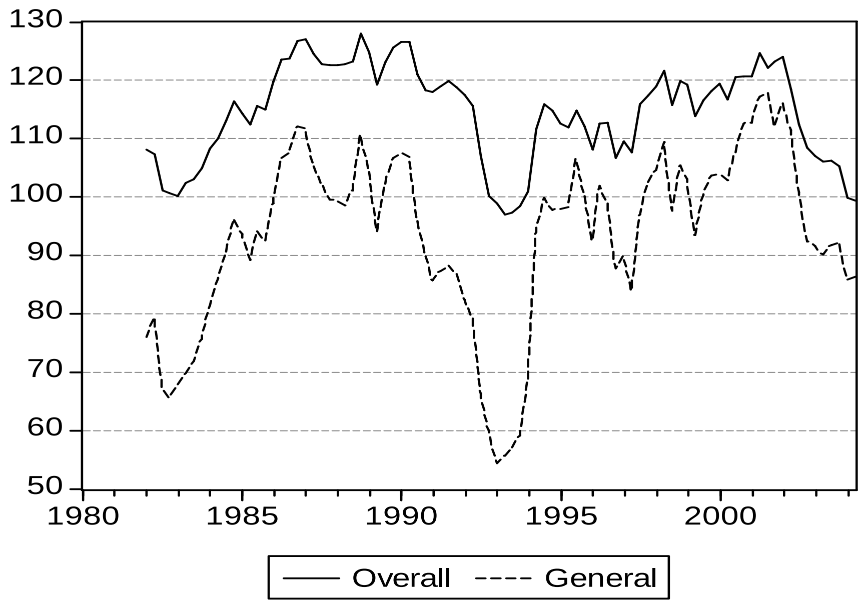
- Future:
  1) anticipated personal financial conditions over the coming year;
  2) anticipated economic conditions over the coming year;
  3) anticipated job availability conditions;
  8) personal saving possibilities over the coming year.

Clearly, the indices Present and Personal on the one side, and General and Future on the other side, are highly correlated because they share some items. The following figures 1-4 report the (seasonally adjusted) time series of the five indices (the overall and the four sub-indices).

Figure 1. Italian Consumer Sentiment Indices (1980=100, seasonally adjusted). Overall vs Personal

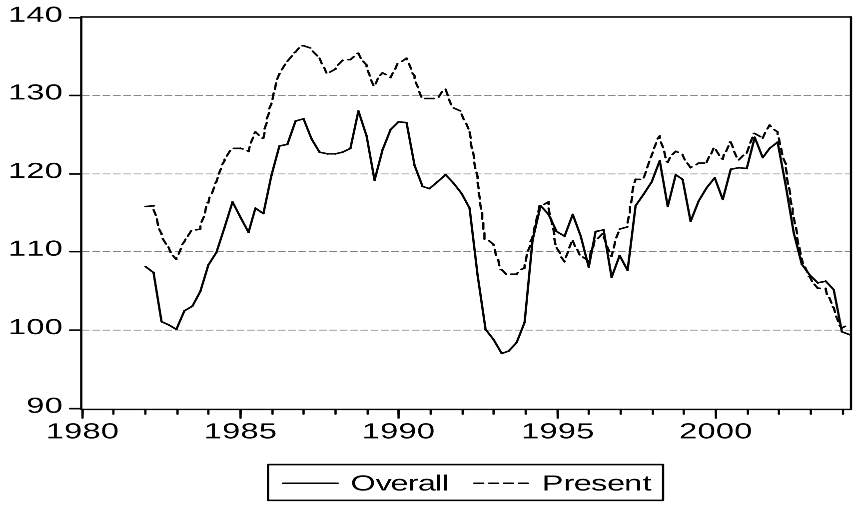
Source: ISAE

Figure 2. Italian Consumer Sentiment Indices (1980=100, seasonally adjusted). Overall vs General



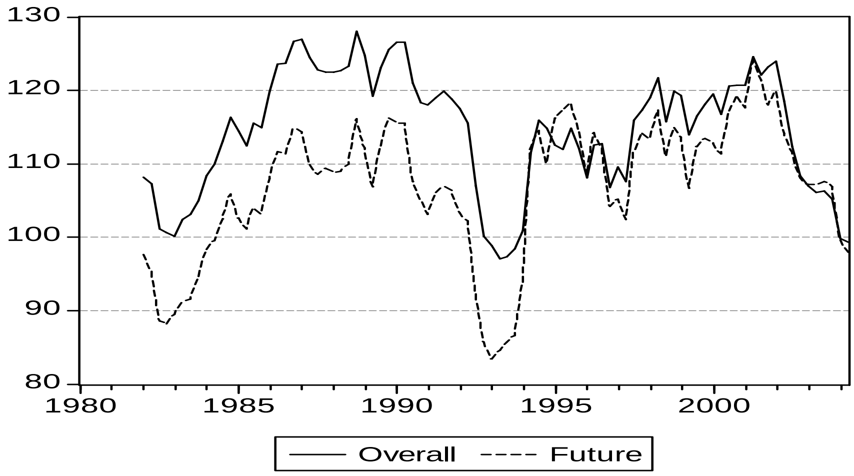
Source: ISAE

Figure 3. Italian Consumer Sentiment Indices (1980=100, seasonally adjusted). Overall vs Present



Source: ISAE

Figure 4. Italian Consumer Sentiment Indices (1980=100, seasonally adjusted). Overall vs Future



Source: ISAE

A visual inspection of the figures 1-4 suggests the presence of some stylised fact. First, the indices Personal and Present have been nearly always higher than the Overall one. Consequently, the opposite holds for the other two complementary measures. Second, starting from the mid-90s there seems to be a tendency to the reduction of the gap between indices. This could be partly due to the change in the methodological approach to the households — now interviews are made by phone, while they were direct before the 1995. Third, and more interestingly, General and Future show a larger reaction to the early 90s recession and then a positive trend up to the present slowdown in the economic activity. On the contrary, Personal and Present remained more constant in the last decade, while show a stronger reduction in the recent years. As a consequence, these indices languish now in an historical minimum, whereas the lowest values for General and Future were reached in 1993. A tentative explanation may be found in the European commitments signed at Maastricht in 1992. In fact, in that year there was the last devaluation of the Lira, while henceforth there was a tightening of the fiscal stance. May be that people felt the devaluation as a deterioration of the system more than of their personal condition (in the overall index there is not a question referring to the inflation), while the protracted fiscal consolidation (and the following impossibility to implement counter-cyclical policies) impinged especially on the individual position. For instance, the strong reduction in the interest rates caused by the fiscal stance have had a dramatic impact on the disposable income of Italian families. Also, the public sector reduced its role as employer of last resort. Moreover, one can think that the perceptions about system-wide economic conditions are subjected to the same information sources (*e.g.* the mass-media). To the extent this is true, people's sensations could tend to follow an "epidemic" pattern ending up with single-clustered expectations. This could help to explain the higher volatility of the general index as compared to that of the personal one. Altogether this suggests that each sub index can tell a different, and potentially useful, story about the relationships between macroeconomic variables and people's feelings. This calls for the deeper and more formal analyses which I perform in the next sections.

3. Some Theory, Data Description and Preliminary Econometrics

The economic literature pointed out several determinants of consumption. The path-breaking paper by Davidson *et al*'s (1978) specifies two causes of long-run consumption: income and inflation. However, alternative consumption theories and empirical evidence from advanced industrial countries suggest that consumption is likely to be influenced by additional variables, such as wealth, demographic factors, liquidity constraints, and uncertainty (Muellbauer and Lattimore, 1995). The timing of the variables has been investigated as well. Nelson (1987) showed that consumption growth is correlated with lagged growth in disposable income. Working on the famous Hall's (1978) random walk hypothesis, Campbell and Mankiw (1989, 1990) have argued that consumption growth is a random walk once the response of consumption growth to the contemporaneous change in income is taken into account.

In an attempt to mediate between data availability, degrees of freedom and theoretical suggestions, I select a set of "standard" regressors, namely: the real wage; the real interest rate (loans to households); the annual rate of inflation. I then add the (seasonally adjusted) number of new registration of cars (Car) because it is used by ISTAT to estimate durable goods quarterly data. These determinants are used to model several kinds of consumption[1]. A widespread used way to disentangle the total consumption of households deals with the durability of the goods, which gives consumption for durable goods, non durable goods and services. Then, according to the goal of this paper I analyse two other kinds of consumption. The first wishes to insulate the consumption for "happiness" and to this end I collect services items such as recreation, hotels, and "abroad" (a proxy for holidays). The second tries to define the expenses for "illness" and it is formed by disbursements for health care. The logic behind is that a serene mood could lead to an increase in the consumption for recreation, while outlays incurred for medical care should be driven by psychological motivations only, if any, to a

[1] As will be explained in the text, I use one only general unrestricted model for each kind of consumption, thus the variable "Car" enters all the consumption models (but Illness). The logic of the inclusion is that some non durable and services expenditures are related to the ownership and the use of a car (insurance, gasoline, maintenance, etc.).

lesser extent. In other words the idea is to check if and how much the information content of subjective indicators depends on the type of the macroeconomic variables under analysis. To sum up, I have six baseline models according to six kinds of consumption (total, durable goods, non durable goods, services, happiness, and illness), which are regressed on four right-hand-side variables (Z= real wages, real interest rate, inflation rate, new registration of cars).

Since the aim of the work and the use of a very general unrestricted Autoregressive Distributed Lag (ADL) model as a starting point, my approach is "one-model-fits-all". For instance, it is obvious that the expenditure for Illness should depend on demographic factors as well. However for the comparisons of section 5, it is sufficient that the models be an empirically admissible and coherent representation of the data. It is verified by the analysis of the residuals.

Before detailing the baseline models (see section 4), it is important to recall that the first necessary step to validly estimate a model is the univariate analysis of the stochastic properties of the series involved. The attention devoted to this topic is well deserved for several reasons. First, in contrast to stationary or trend stationary time series, models with a stochastic trend are persistent so they have time dependent variances that go to infinity with time. In other words, shocks have permanent effects on the values of the process. Second, when a series is used in regressions with other variables the interpretation of the regression results can depend on whether the variables involved are trend (TS) or difference stationary (DS). This phenomenon is related to the "nonsense" and "spurious" regression literature surveyed by Phillips (1986). It is also well known that unit root tests are based on asymptotic critical values. One expects in finite samples that the use of asymptotic critical values will result in over-rejection. I address this problem by performing two unit root tests. The first (NP) was worked out by Ng and Perron (2001). It yields both substantial power gains and a lower size distortions over the standard unit root tests, maintaining the null of unit root. NP offer four test statistics based on the GLS detrended data y_t^d. Altogether these statistics are modified versions of Phillips-Perron Z_α and Z_t statistics (1988), the Bhargava (1986) R_1 statistic, and the Elliot *et al.* Point Optimal statistic (1996):

$$MZ_\alpha = \left(T^{-1}(y_t^d)^2 - f_0\right)/2\kappa \tag{1}$$

$$MSB = \left(\kappa/f_0\right)^{1/2} \tag{2}$$

$$MZ_t = MZ_\alpha \;\text{X}\; MSB \tag{3}$$

$$MPT = \bar{c}^2 \;\; \kappa - \bar{c}T^{-1} \;\; (y_t^d)^2)\,/f_0 \quad \text{(if deterministic term = constant)} \tag{4}$$

$$MPT = \bar{c}^2 \;\; \kappa + (1-\bar{c})T^{-1} \;\; (y_t^d)^2)\,/f_0 \quad \text{(if deterministic terms = constant, trend)} \tag{5}$$

where = $\sum_{t=2}^{T} (y_{t-1}^{d})^2)/T^2$ and f_0 is an estimate of the residual spectral density at the zero frequency[1]. The choice of the autoregressive truncation lag, p, is critical for correct calculation of f_0. Here p is chosen by using the modified AIC suggested by Ng and Perron (2001).

The second is the KPSS test (Kwiatkowski *et al.* (1992)), which can be thought as complementing the NP one because it tests the null hypothesis that the variable follows a TS stochastic process. Suppose the NP test fails to reject the unit root null because of low power. The KPSS test which has (trend) stationarity as the null should indicate the data have no unit roots. On the other hand, if the KPSS test rejects the trend stationarity null, then we have stronger evidence for unit root presence. That is, consistent results from NP and KPSS tests yield more persuasive evidence on data persistence, while conflicting results indicate uncertainty associated with the interpretation of the individual test outcomes[2]. The KPSS test is based upon the residuals from the OLS regression of y_t on the deterministic term x_t:

$$y_t = x_t'\delta + u_t \tag{6}$$

The LM statistic is be defined as:

$$\text{LM} = \sum\nolimits_t^n S(t)^2 / (T^2 f_0) \tag{7}$$

where f_0 is an estimator[3] of the residual spectrum at frequency zero and where S(t) is a cumulative residual function:

$$S(t) = \sum_{r=1}^{t} \hat{u}_r \tag{8}$$

based on the residuals $\hat{u}$ = yt – x' $\hat{\delta}$(0). I maintain the same lag length selection criterion already used in the NP test.

The following table 1 collects the empirical results**.**

---

1 The frequency zero spectrum method used is the AR-GLS detrended.

2 Conflicting results might also indicate the presence of fractionally integrated processes. I do not address this issue here.

3 The frequency zero spectrum method used is the Kernel-Bartlett sum-of-covariances.

Table 1. Unit root tests (quarterly data 1980.1—2003.4)

	MZα	MZ_t	MSB	MPT	KPSS
Total consumption	-6.75826	-1.70761	**0.25267**	**13.6021**	0.179203
Δ(Total consumption)	**-21.1496**	**-3.24773**	0.15356	4.33431	**0.064255**
Durables	-8.05009	-1.97502	**0.24534**	**11.4133**	0.118274
Δ(Durables)	-3.62172	-1.34559	**0.37153**	**25.1592**	**0.064629**
Non Durables	-5.98427	-1.63820	**0.27375**	**15.1369**	0.174084
Δ(Non Durables)	-2.71788	-1.02638	**0.37764**	**29.2161**	**0.109472**
Services	-2.79698	-0.93964	**0.33595**	**25.8587**	0.268486
Δ(Services)	**-17.9541**	**-2.98072**	**0.16602**	**5.17083**	**0.050400**
Happiness	1.07461	0.98842	**0.91980**	**60.9777**	1.270553
Δ(Happiness)	-0.58889	-0.40363	**0.68541**	**26.1136**	**0.086438**
Illness	-3.80728	-1.28162	**0.33662**	**22.6197**	0.192348
Δ(Illness)	-0.46344	-0.47952	**1.03469**	**195.224**	**0.136969**
Real wage	0.76764	0.56129	**0.73119**	**38.8100**	1.045664
Δ(Real wage)	-3.55903	-1.08201	**0.30402**	**6.89841**	**0.103472**
Real interest rate	-2.19994	-0.88833	**0.40380**	**33.7675**	0.280794
Δ(Real interest rate)	**-42.4318**	**-4.60587**	0.10855	2.14864	**0.027553**
Inflation	-13.5926	-2.59510	**0.19092**	**6.77348**	1.116020
Δ(Inflation)	**-36.1758**	**-4.24189**	0.11726	2.58059	**0.074645**
Car	-9.09355	-2.12656	**0.23385**	**10.0447**	0.329684
Δ(Car)	-6.17382	-1.75653	**0.28451**	**14.7597**	**0.059292**
Overall sentiment	-5.32131	-1.45439	**0.27331**	**16.5535**	**0.098587**
Personal	-4.47922	-1.24193	**0.27726**	**18.4745**	**0.141919**
General	-6.88406	-1.80451	**0.26213**	**13.2966**	**0.084934**
Future	-8.91567	-1.98958	**0.22316**	**10.6821**	**0.070759**
Present	-6.01753	-1.58465	**0.26334**	**15.0102**	**0.105081**

Lag length criterion: Modified AIC; constant and trend included. All variables, but the interest rate, are in log. Car = # of new registration of cars. Happiness = Consumption for recreation, hotels, restaurants and "abroad". Illness = Consumption for health care. Bold values imply stationariety (5%), that is imply the rejection of the null of unit root or the non rejection of the null of stationariety (in the KPSS case). $\Delta X_t \equiv X_t - X_{t-1}$.

Source: ISAE (indices); Bank of Italy (interest rate); ISTAT.

As usual when dealing with short samples, results about the statistical nature of the time series are mixed. However, the overall picture emerging from Table 1 leads to think that the macroeconomic variables are I(1), while indices are stationary. This latter finding is somewhat expected because it is hard to think about an everlasting "irrational exuberance (or apprehension)" in all agents. Since the quantitative (economic) time series are integrated of the same order, the next step is to test whether them (alternatively each type of consumption and its above mentioned quantitative[1] determinants) are cointegrated. In order to save space I do

[1] Given that qualitative indices result stationary they can not enter the ECM (see Section 4).

not report the results of Johansen procedure which seems to suggest that, at most, there is only one cointegration vector. This means that the single-equation approach can be validly pursued because the weak exogeneity requisite holds (Engle *et al*., 1983).

4. The baseline models

Within the single-equation framework, I follow the general-to-specific approach in its error-correction (ECM) version to modelling consumption (Davidson *et al*., 1978). The major feature of the ECM is that it distinguishes between short term and long term effects. The specification of long-run components in the consumption function draws upon economic theory, while dynamics are a clear case of "measurement without theory", at which Hendry (1995) refers to the "let the data speak" approach. This is another reason, beyond their stochastic properties, why I do not enter the CS indices in the error correction term - they are measurement without theory. This statistical background seems to be quite suitable in the present context with mixed information coming from both empirical and theoretical applications. The starting "general" system is an ADL(4,4) reparametrised *a là* Wickens-Breusch (1988) in order to estimate (in one single step and more powerfully in small samples) the possible presence of the ECM. Formally (i=0,…,4; j=1,…,4):

$$\Delta C_t = \alpha + \Sigma\gamma\Delta \mathbf{Z_{t-i}} + \Sigma\eta\Delta C_{t-j} + \boldsymbol{\beta}\mathbf{Z_{t-1}} + \delta C_{t-1} + \varepsilon_t \qquad (9)$$

where $\Delta X_t \equiv X_t - X_{t-1}$, C_t is the consumption variable, and $\mathbf{Z_t}$ is the right-hand-side variables vector as defined in section 3. To account for monetary illusion in the short run, I enter the first differences of the interest rate and of the (log)wage in nominal term. Unreported[1] diagnostic tests suggest that four lags are sufficient for a reliable general unrestricted model to start with. After having obtained the results for the whole set of the variables (the "general" model), I sequentially delete the insignificant variables to select a parsimonious (specific) model[2] for increasing the precision of the estimates (Hendry, 1995). The specification search ends according to two information criteria based on the log-likelihood value. The first one, proposed by Akaike (1973) penalizes the log-likelihood by 2n/T for n parameters and a sample size of T; the second (Schwarz, 1978), by ṇlog(T)/T.

As requested in order to see if the model captures the essential characteristic of the data (denoted congruence), in each step of the testing down procedure I check if i) the parameters are constant (via a recursive analysis); ii) the residuals

[1] The whole set of empirical results/tests are available upon request from author.

[2] Variables not significant at the conventional levels (5%) are considered irrelevant and excluded from the regression. At the end of the specification search I performed an F-test on the joint significativity of the deleted variables. Evidence suggests to proceed with the "specific" model.

are homoschedastic innovations. All in all it results that the baseline models are empirically admissible and coherent. Thus, in the next section I will compare CS-augmented models with very strong and reliable contenders. In order to save space, I do not display all empirical results. Some diagnostic tests referring to all these models are reported in the following section (see table 2).

5. The Information Content of Consumer Sentiment Indices

To investigate the usefulness of Italian consumer surveys as estimation and forecasting tool over the period 1982.1-2003.4, I estimate new models which differs from the relative baseline only because of the inclusion of alternative contemporaneous and lagged confidence indicators. According to the results of section 3, CS indices are stationary processes, thus they enter the baseline models in log-levels. As for the other right-hand-side variables, four lags proved to be sufficient for a reliable general unrestricted model to start with. Then I repeat the testing down procedure. Finally, I compare each benchmark model with the corresponding sentiment-augmented in their specific form in terms of goodness-of-fit (adjusted R^2) and out-of-sample forecast performance statistics (root-mean-square % error). These latter are computed by estimating the specific model over the period 1980.1—1999.4, and then by performing both static and dynamic forecasts over the next sixteen quarters (up to 2003.4). Table 2 collects the findings of all the experiments.

Table 2. Standard vs Consumers Sentiment Augmented Consumption Growth Models

		Models of Total Consumption							
Sample		*Baseline*	+ Overall	+ Present	+ Future	+ General	+ Personal	+ Pres+Fut.	+ Gen+Pers
1980-03	**Adj-R^2**	*0.68628*	0.68185	**0.68932**	0.68299	------	**0.71101**	0.68448	**0.70838**
1987-03		*0.69367*	0.68006	**0.69604**	0.68324	------	**0.69911**	0.68303	**0.69438**
1992-03		*0.70079*	0.68000	**0.71558**	0.68372	------	**0.74194**	**0.70500**	**0.76010**
2000-03	**RMSPE_s**	*0.41075*	0.43506	0.42196	0.45071	------	0.41118	0.44582	0.43037
2000-03	**RMSPE_d**	*0.86358*	1.24603	**0.85639**	1.17518	------	0.88937	1.00572	0.88808
		Models of Consumption for Durable Goods							
80.1-03.4	**Adj-R^2**	*0.86882*	------	------	**0.87429**	**0.87225**	**0.87303**	**0.88208**	**0.88157**
87.1-03.4		*0.87101*	------	------	**0.87881**	**0.87241**	**0.87841**	**0.88511**	**0.88002**
92.1-03.4		*0.89576*	------	------	**0.91646**	0.89416	**0.90138**	**0.90070**	0.89348
00.1-03.4	**RMSPE_s**	*1.17977*	------	------	1.30916	1.23579	1.40113	**1.17287**	**1.14684**
00.1-03.4	**RMSPE_d**	*2.11762*	------	------	2.72201	**1.77434**	2.47737	**1.97769**	**2.06919**
		Models of Consumption for Non-Durable Goods							
80.1-03.4	**Adj-R^2**	*0.45967*	**0.45993**	**0.46477**	**0.46388**	------	**0.48301**	**0.48087**	**0.51496**
87.1-03.4		*0.44770*	0.43897	0.41167	0.42465	------	0.39069	0.42286	**0.54438**
92.1-03.4		*0.38288*	**0.38853**	0.38195	0.35771	------	0.37636	0.34416	0.23787
00.1-03.4	**RMSPE_s**	*0.61607*	0.68117	0.80861	0.72136	------	0.64605	**0.54664**	0.64906
00.1-03.4	**RMSPE_d**	*1.01191*	1.60196	2.09897	1.64914	------	1.13034	**0.77583**	**0.89510**

		Models of Total Consumption							
Sample		*Baseline*	+ Overall	+ Present	+ Future	+ General	+ Personal	+ Pres+Fut.	+ Gen+Pers
		Models of Consumption for Services							
80.1-03.4	**Adj-R^2**	*0.38368*	0.37394	**0.39074**	**0.38955**	0.37713	**0.39543**	**0.40650**	**0.39972**
87.1-03.4		*0.33333*	0.30451	0.33329	**0.35948**	**0.34316**	**0.35000**	**0.35092**	**0.34690**
92.1-03.4		*0.43385*	**0.47593**	**0.52668**	**0.43832**	**0.47651**	**0.46776**	0.38137	**0.43655**
00.1-03.4	**RMSPE_s**	*0.35411*	0.42201	0.50067	0.45292	0.53487	0.37995	0.51405	0.41917
00.1-03.4	**RMSPE_d**	*1.09121*	**1.01232**	1.20677	**0.94159**	1.23044	**0.80676**	1.33975	**0.90666**
		Models of Consumption for Happiness							
80.1-03.4	**Adj-R^2**	*0.36959*	------	0.36677	**0.38009**	0.35287	**0.38307**	**0.41198**	**0.38851**
87.1-03.4		*0.25953*	------	**0.27455**	**0.28592**	**0.28745**	**0.29088**	**0.40523**	**0.32817**
92.1-03.4		*0.29642*	------	**0.32362**	**0.40172**	**0.36660**	**0.32998**	**0.31313**	**0.44910**
00.1-03.4	**RMSPE_s**	*1.20972*	------	**1.11277**	**1.19431**	1.39082	**1.18250**	1.31710	1.26380
00.1-03.4	**RMSPE_d**	*2.07997*	------	2.33869	2.15515	2.31238	**2.01811**	2.31472	2.36184
		Models of Consumption for Illness							
80.1-03.4	**Adj-R^2**	*0.352144*	------	------	------	------	------	------	------
87.1-03.4		*0.362282*	------	------	------	------	------	------	------
92.1-03.4		*0.387924*	------	------	------	------	------	------	------
00.1-03.4	**RMSPE_s**	*2.086742*	------	------	------	------	------	------	------
00.1-03.4	**RMSPE_d**	*9.229285*	------	------	------	------	------	------	------

Happiness = Consumption for recreation, hotels, restaurants and abroad. Illness = Consumption for health care.

All variables, but the interest rate, are logged. The reported statistics are drawn from the specific models obtained after the testing down procedure. In all cases, the general model is the ADL(4,4) reported in the text as equation [9]. To account for monetary illusion in the short run, I enter the first differences of the interest rate and of the wage in nominal term. The general baseline model includes: real wage, real interest rate, inflation rate, new registration of cars. The general CS-augmented models just add to it the contemporaneous and the first four lags of the log-level of the index/indices reported in the second row ("+ Overall"; "+ Present" etc.). RMSPE = root-mean-square % error. Tag=_s refers to static forecasting exercises; tag=_d refers to dynamic forecasting exercises. They are computed on sixteen-steps (2000.1-2003.4) out-of-sample forecasts (of the log-level of the dependent variable) based on regressions over the period 1980.1-1999.4. Bold values emphasise lower R^2 and higher RMSPE than those realised by the baseline model which, in turn, implies that climate indices have additional information content beyond that contained in macroeconomic variables. The other values support the contrary proposition. Cells filled with (---) mean that none of the five terms (one contemporaneous plus the first four lags) of the relative CS index is survived to the testing down procedure. Clearly, in these cases the best model turns out to be the baseline one.

The picture emerging from the empirical examination indicates some intriguing fact. The upper rows of Table 2 show that the total consumption growth depends especially on present and personal economic circumstances, while questions about general conditions and subjective expectations do not display bold values, that is they do not have extra information content. The out-of-sample incremental ability of qualitative indices is quite low. As already emphasised by the literature (e.g. Parigi and Schlitzer, 1997), during extraordinary periods as the "tormented" 1992—2003 one in Italy, the emotivity of spending decisions increases. For instance, the variance explained by the model augmented with the General and the Personal indices (+Gen.+Pers.) amount to 76%, whereas the corresponding statistic for the baseline model is 70%. In other sub-samples the difference between these statistics is much smaller. Looking at the "Durable Goods" test, things change. The Overall and the Present indices result orthogonal, while models using other indicators show better performances even if computed throughout different samples. Also, there is evidence of a greater forecasting power as indicated by the RMSPE bold values. The experiments regarding the consumption for non-durable goods indicate that it has emotional content especially in the long term, the most part of the bold values being in the full sample estimation. It can be due to the low volatility of this item as compared to the much more schizoid behaviour of household sentiment. Another result is that couples of indices perform better than single indices probably because, following the previous logic, they are more smoothed than single indices. In particular, the

General and the Personal indices appear to be good for in sample analyses, while the pair Present and Future presents better out of sample statistics. Turning the attention to the expenses for services, findings suggest that the answers about the expectations and the personal feelings of participants in the economy provide the most useful insights on the functioning of this aggregate. As pointed out by the sub sample analyses, adding qualitative information may be very useful in particular periods. The ad hoc trials referring to peculiar collections of disbursements, Happiness and Illness, support the intuition that some kind of outlay is more emotional driven than others. With the exemption of the dynamic forecasting exercises, the CS-augmented models outperform the relative baseline in the most part of the empirical designs. On the other hand, as expected, results for Illness show that in no case the subjective indicators survive the testing down procedure1.

By looking vertically downward table 2 we can note that analysing and drawing conclusions on the aggregate information stemming from surveys could be misleading or, more properly, that it could not be the best way to exploit the informative power of subjective feelings collected among economic agents. The column "+ Overall" has the less number of bold values which are aimed to indicate situations where the CS-augmented model outdoes the corresponding baseline one. Similarly, poor performances are realised by the General index which, on the other hand, is the best choice in order to perform dynamic forecasting exercises about non durable goods. Good results are achieved by using the Personal index, in particular when it is applied to services and to Happiness goods or when it is coupled to the General index. The "+Gen.+Pers." pair beats the baseline model seventeen out of twenty five trials, whereas the model augmented by Present and Future turns out to be the best as refer to the out-of-sample dynamic forecast of the consumption for durable and non durable goods. Finally, the out-of-sample experiments performed over the last four years (2000-2003) show better results when using some sub indices. Thus, even during the recent puzzling behaviour of the Italian climate measures, languishing in a very low level as opposed to a relatively sustained growth in the total consumption, a disaggregated analysis is still able to uncover the predictive ability of some indicators.

6. Concluding remarks

This paper presents new evidence on the extra information content of surveys of people in households beyond that already contained in quantitative variables. The main novelty of this work is the focus on the relationships between

[1] The low predictive power of this baseline model could be link to the absence of, *e.g.*, demographic variables which logically should be inserted. However, as explained in the text (see section 3), what is sufficient here is obtaining empirically admissible and coherent models. And it has been verified by the residuals analysis.

subjective and objective information at a less aggregated level than that is usually analysed. In particular, the overall sentiment index is divided into four sub-indices related to the opinion about the i) current, ii) future, iii) general, and iv) personal situation. In turn, the total private consumption is divided in five items, i) consumption for durable goods, ii) consumption for non durable goods, iii) services, iv) consumption for “happiness”, v) consumption for “illness”.

From the methodological point of view, a two-step procedure to determine the in-sample and the out-of-sample informative power of consumer confidence is performed. First, starting from a general unrestricted ADL model with standard economic regressors, parsimonious consumption equations are found, estimated and used as a forecasting tool. Second, their adjusted R-squared and root mean square errors are compared with those of corresponding models which differ from them only because of the presence of the confidence indicator.

The main conclusions of the paper can be summarized as follows. Surveys data seem to provide autonomous information. No single index has imposed itself as being clearly superior to the others. The worst results are realised by the most aggregate measure, while much better performances can be obtained by associating a particular sub index to a particular consumption item. The emotional content of spending decisions is different depending on the nature of the goods/services under scrutiny. It is different over time as well, especially if there happen to be extraordinary episodes. All in all, these findings may help to explain why the literature, analysing more aggregate quantitative and qualitative information, shows mixed results. Finally, a word of caution. It may be that a different weighting system could increase the information content of the overall index. But this is beyond the scope of this paper and it is relegated in the agenda.

REFERENCES

AKAIKE, H. (1973), Information theory and an extension of the maximum likelihood principle. In Petrov, B.N. and Saki F.L. (eds.) Second International Symposium of Information theory. Budapest.

BHARGAVA, A. (1986), On the Theory of Testing for Unit Roots in Observed Time Series, *Review of Economic Studies*, Vol. 53, 369—384.

BATCHELOR, R. and P. DUA (1998), Improving Macro-economic Forecasts: The Role of Consumer Confidence, *International Journal of Forecasting*, vol. 14, n. 1, 71—81.

BERG, L. and R. BERGSTROM (1996), Consumer Confidence and Consumption in Sweden, *Uppsala University Working Paper*, 7.

BRAM, J. and S. LUDVIGSON (1998), Does Consumer Confidence Forecast Household Expenditure? A Sentiment Index Horse Race, *Federal Reserve Bank of New York Economic Policy Review*, 4, June, 59—78.

CAMPBELL, J.Y. and G.N. MANKIW (1989), Consumption, Income, and Interest Rates: Re-interpreting the Time Series Evidence, In *NBER Macroeconomics Annual*, O.J. Blanchard and S. Fischer (eds.), Cambridge: MIT Press.

CAMPBELL, J.Y. and G.N. MANKIW (1990), Permanent Income, Current Income and Consumption, *Journal of Business and Economic Statistics*, 8, July, 265—79.

CARROLL, C.D., J.C. FUHRER and D.W. WILCOX (1994), Does Consumer Sentiment Forecast Household Spending? If So, Why?, *American Economic Review*, 84, December, 1397—1408.

DAVIDSON, J.E.H., HENDRY, D.F. SRBA, F. and YEO, J.S. (1978), Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal*, 88, 661—692.

GOLINELLI, R. and PARIGI G. (2003), What is this thing called confidence? A comparative analysis of consumer confidence indices in eight major countries, *Temi di Discussione*, Banca d'Italia, 484, September.

ELLIOTT, G. ROTHENBERG T.J. and STOCK J.H. (1996), Efficient Tests for an Autoregressive Unit Root, *Econometrica,* Vol. 64, 813—836.

ENGLE, R.F., D.F. HENDRY, and RICHARD, J.F. (1983), Exogeneity, *Econometrica,* 51, 277-304.

HALL, R.E. (1978), The Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy*, 86, December, 971—87.

HENDRY, D.F. (1995), *Dynamic Econometrics*. Oxford: Oxford University Press.

KATONA, G. (1951), *Psychological Analysis of Economic Behaviour*, McGraw Hill, New York.

KATONA, G. (1960), *The Powerful Consumer: Psychological Studies of American Economy*, McGraw Hill, New York.

KWIATKOWSKI, D. PHILLIPS, P.C.B. SCHMIDT, P. and SHIN, Y. (1992), Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root, *Journal of Econometrics*, 54: 159—178.

LEEPER, E. M (1992), Consumer Attitudes: King for a Day, *Federal Reserve Bank of Atlanta Economic Review,* July, 1—15.

MARTELLI, B. (1998), Le Inchieste Congiunturali dell'ISCO: aspetti metodologici, *Rassegna di Lavori dell'ISCO*, Vol. XV, n. 3, Rome.

MATSUSAKA, J. and A. SBORDONE (1995), Consumer Confidence and Economic Fluctuations, *Economic Inquiry*, 33, 2, 296—318.

MUELLBAUER, J and R LATTIMORE (1995), *The Consumption Function: A Theoretical and Empirical Overview*, in Handbook of Applied Econometrics, M.H. Pesaran and M. Wickens (eds.), Blackwells.

NELSON, C.R. (1987), A Reappraisal of Recent Tests of the Permanent Income Hypothesis, *Journal of Political Economy*, 95, June, 641—46.

NG, S. and P. PERRON (2001), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica*, 69(6): 1519—1554.

PARIGI, G, and SCHLITZER, G. (1997), Predicting consumption of Italian households by means of survey indicators. *International Journal of Forecasting*, 13, 197—209.

PHILLIPS, P.C.B. (1986), Understanding Spurious Regression in Econometrics. *Journal of Econometrics*, 33(31): l—40.

PHILLIPS, P.C.B. and PERRON, P. (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, Vol. 75, 335—346.

SCHWARZ, G. (1978), Estimating the dimension of a model, *Annals of Statistics*, 6, 461—464.

WICKENS, M.R. and BREUSCH, T.S. (1988), Dynamic specification, the long run and the estimation of transformed regression models, *Economic Journal*, 98, 189—205.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 259—292

QUALITY CONTROL AND AGGREGATION METHODS IN THE ISAE INVESTMENT SURVEY

Tatiana Cesaroni, Marco Malgarini, Gaia Rocchetti[1]

ABSTRACT

The ISAE Investment survey provides quantitative information about investment plans and structure, and qualitative data on the factors influencing firm' behaviour. In dealing with business surveys a special attention should be devoted to the reliability of the results and to appropriate weighting methods; this is particularly important when quantitative as well as purely qualitative data are concerned. In this respect, we propose deterministic and statistically-based methods for the treatment of missing data and outliers. After a careful deterministic control, a fairly simple econometric model is used for missing data, based on information on sector, geographical location, number of employee and investment plans for the previous years. The robustness of sample estimates in case of extreme observations is then tested with the use of the winsorized mean. In processing the results, estimate of aggregated investment plans is then derived using sample weights; aggregation of qualitative data is finally based on appropriate size weights.

1. Introduction

The Institute for Studies and Economic Analysis (ISAE, previously ISCO) carries twice a year a survey on investment expenditures of Italian manufacturing firms. The survey, which is part of the Joint Harmonised Program of the European Commission, is based on the same sample of the monthly survey on the manufacturing sector, stratified upon firms' industry, region and size. It is a postal survey, differing in that from the monthly one, which uses the Computer Aided Telephonic Interview (CATI) system; together with the relatively higher

[1] Tatiana Cesaroni: ISAE and University of Roma "La Sapienza"; Marco Malgarini: ISAE; Gaia Rocchetti: ISAE and Saint Anna School of Advanced Studies, Pisa; t.cesaroni@isae.it; m.malgarini@isae.it; g.rocchetti@isae.it

complexity of the questionnaire, this implies a much lower response rate with respect to that of the monthly survey[1].

The survey provides qualitative information on the factors eventually influencing revision of investment plans during the year; besides, it also provides an estimate of the structure of investment expenditures and of the rate of growth of manufacturing firms investments in two consecutive years (the previous and the current year in the spring survey — performed between March and April of each year — and the current and following year in the autumn survey — performed between October and November). More precisely, in the case of the ISAE survey, firms are not asked directly for an evaluation of the rate of growth of their capital expenditures: instead, they provide a quantitative estimation of actual investment in three consecutive years (investments of the two previous years and of the current one in the spring survey, investments of the previous, current and next year in the autumn survey).

Because it provides also quantitative information besides the more traditional qualitative ones, the investment survey requires a more careful control on the quality and general reliability of answers and, on the other side, a somewhat different approach to aggregating procedures[2]. Until recently, the controls for the quality of answers and their internal reliability were based only on deterministic methods, aimed at checking for the existence of possible outliers and missing data. As for the processing of quantitative data, the methodologies used by ISAE involved a simple sum of firm level data; for data on investment structure and factors influencing investment plans, it was based on a weighted average of firms' answers, using investment reported by response units as weights.

New methods for quality control and processing of the results are proposed in this paper; in particular, after a presentation of the reference Universe and the theoretical and actual sample in section 2, section 3 deals with the problem of reliability of survey results, describing the deterministic procedures used by ISAE to check for survey quality, and introducing new, fairly standard, statistically based techniques for the identification and treatment of missing data and outliers. The methods used in processing both qualitative and quantitative data are discussed in section 4, while the results obtained for the period 2000—2004 (autumn survey) are included in a statistical appendix. Some consideration about possible future research concludes the work.

[1] Starting from the April 2005 survey, a part of the interviews has been performed with the CATI system.

[2] In the monthly survey, the quality control is performed with deterministic methods; the aggregation of the qualitative results is usually based on a four-step procedure, in which an internal weight is used to aggregate firms results in the sample strata, and then external weights are applied to progressively aggregate the results (see Malgarini et al., 2005).

2. Target Universe, theoretical and actual sample

2.1. The Target Universe and the theoretical sample

The target Universe for both the monthly and the Investment survey on the manufacturing sector includes all the firms with more than 10 employees operating in the Class D of the NaceRev.1.1 Classification. Starting from the year 2000, the statistical unit considered for the survey is the firm: the choice of the firm instead of the local unit is based on the fact that the latter, previously considered as statistical unit, in the past was often reporting to the firm' headquarters in order to provide answers (see on this Malgarini, Margani and Martelli, 2005) The sampling units are extracted from the most recent official archive of active firms, provided by the National Institute of Statistics (ISTAT), denominated ASIA; the method used for extracting the sample is the *optimal allocation to strata* , in which strata characterised by an higher internal variability are over-sampled with respect to those with a lower variability (on this topic see again Malgarini, Margani and Martelli, 2005 and Martelli and Rocchetti, 2004). The sample is stratified along 3 axes, namely, industry, region and size of the firms. More in particular, the geographical stratification is based on the Italian administrative regions, aggregating *Valle d'Aosta* and *Piemonte*, for a total of 19 regions. At the size level, 3 groups are considered, namely the firms with a number of employees between 10 and 49 (Small firms), those from 50 up to 249 employees (Medium-size firms) and the big ones (with more than 250 employees). At the industry level, the sample is stratified on 21 main sectors, based on the old NACE 1970 Classification (the so-called "ISAE Branches"). The major differences between the old classification and the NACE Rev.1.1 are the inclusion of furniture in the wood & wood product sector, instead that in the miscellaneous final sector of the manufacturing classification, and a slightly different classification of some branches of the mechanical industry.

All in all, we have a total of 19 (Regions) x 3 (size groups) x 21 (industries) = 1.197 sample strata, with on average 4.000 monthly interviews. The number of interviews per strata is therefore quite low, and this may bias the resulting estimation. However, it should firstly be considered that some of the sample strata may be indeed void, both because those strata are void even for the universe of manufacturing firms or because no firm is selected in that particular strata; as a consequence, the average number of firms in each relevant sample strata (i.e., the non-void sample strata) is higher then what may be inferred at first sight. However, anticipating some of the consideration in section 2.2 below, the ISAE Investment Survey presents much lower response rate with respect to that of the monthly survey, with an average value of 20% of responses. Hence, bias could be stronger then those eventually observed in the monthly manufacturing survey, since a greater part of originally defined strata will be void. In order to reduce possible bias, we therefore decided to merge the sample strata, trying to minimize the number of empty cells. More in details, while retaining the stratification in 3 size groups and 21 industries, the 19 regions are collapsed into the four standard geographical partitions (North West, North East, Centre, *Mezzogiorno*); the resulting weights are then calculated adding up those relative to the original strata. With this settings, we have 4 (geographical partitions) x 3 (size groups) x 21 (sectors) = 252 sample strata, with on average 800 monthly interviews. Table 1 presents the industry composition of the theoretical sample obtained aggregating further (for presentation purposes) the 21 industries into 8 main sectors[1], and compares it with that of the target Universe, together with the relative sampling fraction. The adoption of the optimal allocation to strata method in selecting the theoretical sample results in over-sampling of the industries of food, beverages and tobacco, metallic and non metallic minerals, transport equipment and chemicals; on the other hand, textiles, clothing and footwear and mechanical sectors are slightly under-sampled, due to their relative lower internal variability. On average, in the manufacturing sector the sampling fraction is equal to 4.3%; it rises to 8.5% in the transport equipment and is as low as 3.4% in the mechanical sector.

Table 1. Target Universe and theoretical sample

Sectors	Target Universe		Theoretical sample		Sampling fraction
	N	%	N	%	
Textile, clothing, footwear	20582	22.1	828	20.5	4.02
Paper, printing, publishing	5033	5.4	214	5.3	4.26
Chemicals, rubber, plastics	7295	7.8	379	9.4	5.19
Metallic & non metallic minerals	9445	10.1	549	13.6	5.81
Mechanical sectors	32427	34.8	1109	27.5	3.42
Transport equipments	2177	2.3	184	4.6	8.45
Other manufacturing n.e.c.	9800	10.5	440	10.9	4.49
Food, beverages, tobacco	6474	6.9	337	8.3	5.20
Total manufacturing	93233	100	404	100	4.33

2.2. The actual sample

In the case of the monthly survey on the manufacturing sector, the actual monthly sample is approximately equal to the theoretical one: in fact, the use of the C.A.T.I. system allows substituting in "real time" for non-responding firms with new firms extracted from the same sample strata. As a result, the number of

[1] Some of the 8 industries considered in the table are an aggregation of the 21 ISAE branches; in particular, the "Textile, clothing and footwear" industry is the aggregation of branches 1, 2 and 5 above; "Paper, printing and publishing" is the branch 4; "Chemicals, rubber & plastic products" aggregates branches 6, 7, 10, 11 and 18; "Metallic and non metallic minerals" aggregates branches 8 and 9; "Mechanical sectors" comprises industries 12, 13, 14, 15 and 19; "Transport equipment" those in the branches 16 and 17; Other Manufacturing n.e.c the branches 3 and 21.

firms interviewed each month is very close to the 100% of the total sample size, and the structure of the theoretical sample is almost identically matched in the actual sample. More precisely, an average proportion of 1% of non-answering firms is substituted each month with new firms extracted form the same sample strata.

In the case of the bi-annual survey on investments of manufacturing firms, however, the more "traditional" postal technique is used to carry the interviews: the questionnaire is sent (usually mid January and mid September respectively for the spring and autumn survey) using the postal service to all the firms answering the monthly survey, and they have to return it, again using the postal service, in time for the scheduled completion of the survey. Together with a relative higher complexity of the questionnaire (which includes also quantitative questions besides the more traditional qualitative ones) and with the relative low frequencies of the survey, that may imply some difficulties in finding the appropriate person to answer the survey within the firm, this results in a much lower response rate with respect to that of the monthly survey on the manufacturing sector. More in detail, in the period April 2000-November 2004 the number of responding firms varied from a maximum of 927 (23% of the sample) to a minimum of 723, only the 18% of the sample. In this respect, table 2 presents the industry and size composition of the actual sample in the period considered, both with respect to the number of answering firms, and of their relative importance in the sample in terms of employees. Looking at the number of answering firms, however, the distribution of firms appears fairly stable, both at the industry and size level. The industries with the higher representativeness in the actual sample are those of the mechanical (representing a share between 26.5 and 31.1% of the sample), textile, clothing and footwear (between 17.2 and 14.1% of the sample), and metallic and non metallic minerals sectors (between 12.1 and 15.5 of the sample); from this point of view, the composition of the actual sample closely reflects that of the theoretical one (see again table 1), even if the textile sector appear to be underrepresented with respect to the theoretical sample.

At the size level, at least the 53% of the sample is represented by small size firms (10—49 employees); around 30% of the respondents are firms with a number of employees between 50 and 249, whilst the large size firms (>250 employees) accounts for little more than 12% of the sample.

The internal variability of the actual sample from one survey to another is higher looking at the industry and size composition in terms of the number of employees of the responding firms. In particular, a major variability is found for the mechanical industries (accounting in terms of employees for a share varying between 19.4 and 35.6% of the total) and for transport and equipments (between 29.6 and 40% of the total). Overall, in all the survey considered these 2 sectors accounts for at least 58% of the total number of employee of the interviewed firms. At the size level, over 80% of total employees are from large size firms, around 14% from medium size firms and 5% for the small ones.

Looking at tables 1 and 2, two major considerations emerge: the response rate of the survey is low, as is quite typical for postal surveys; however, the industry and size composition of the sample in terms of the number of responding firms is quite stable in time, favouring comparability of the results[1]. On the other hand, the sample is much less stable looking at the distribution of answering firms' employees: looking more carefully at the data, this seems to be due to a high substitution rate, from one survey to another, in the medium and small size groups. Indeed, in terms of employees, large size firms accounts for a quite stable 80% of the sample, while the fraction of small firms varies between 6 and 4.5% of the sample, and that of medium size firms between 16.5 and 10.9% of the total employees of answering firms. In this respect, it should be considered that the weighting system used for the qualitative questions (described in section 4.2) is based on a "scale factor", represented by the actual investment expenditures of the responding firms; the quite high variability of the employees' distribution may raise a risk of high variability of those weights from one survey to another[2]. On the other hand, the use of the investment itself as a scale factor seems to be quite appropriate, also in the light of the recent OECD (2003) contribution that recommends the use of variable-specific weights, when available, in the processing of the results: for instance, in weighting the firms' evaluation on the structure of investment expenditures the more appropriate "scale factor" indeed is, according to the authors of this paper, the investment of the firm. In other words, in this case a possible trade-off arises, between the choice of an appropriate scale factor and the risk of inducing high variability in the sample estimates. In this paper, we propose to accept the risk of higher variability in order to implement an appropriate weighting scheme.

[1] If responding firms are distributed in a largely comparable way between sectors and size groups in different surveys, the results may be considered highly comparable from one survey to another.

[2] The authors wish to thank an anonymous referee to a previous version of the paper for having raised this point.

3. Data reliability

To assess the quality of survey results an important aspect to investigate is that of their internal reliability, i.e. the accuracy with which the survey measures the target variable (OECD, 2003, p. 25). In the case of the investment survey, reliability issues may be even more relevant then for the others current survey conducted by ISAE, because of the fact that it gathers also quantitative information beside the more usual qualitative ones. For all the survey, and in particular for that on investments, ISAE usually implements a series of quality controls in order to increase the reliability of survey data; until recently, the controls were mostly based on “judgemental” procedures, i.e. on the autonomous judgement of the analyst dealing with the data. More in detail, the “judgemental” controls were primarily aimed at dealing with problems related to missing information and the presence of possible outliers in the answers provided by the firms. In this context, missing data problems arise when questionnaires comes incomplete in one or more of the answers, because the firms are not able, or are not willing, to respond to some of the questions; for instance, in the case of the quantitative questions on the actual amount of investment, it may happen that a firm is able to provide an answer only to one or two of the years for which they are asked for, but not for all the three years requested in the questionnaire. On the other hand, outliers occur when data provided by the firms are apparently inconsistent, either with respect to the rest of the data provided in the same questionnaire, or comparing two or more questionnaires of the same firm in consecutive surveys[1]. Their detection and treatment is needed in order to reduce the variability of estimates.

In what follows, the usual “judgemental” procedures used by ISAE are described (3.1), together with a proposal of new statistically-based methods (3.2) for dealing with both missing data and outliers.

3.1. The “judgemental” procedures

3.1.1. Missing data

Currently, missing data for the investment survey are identified in the data entry stage; in practice, given the postal nature of the survey, data are manually entered in the data base: at this stage, the analyst checks and reports for missing data (in the sense described above). The following step is to proceed with the telephonic re-interview of the firms that missed some of the answers. If the firms are able to complete the information set, the procedure goes back to the data entry stage, integrating the answers previously provided with the new data. If instead the data are still missing, the questionnaire is however kept in the database for the

---

[1] For istance in the case of investments survey the amount of investments indicated by a firm for a given year can be very different from the investments provided from the same firm for the same year in another consecutive survey.

information it actually provides, leaving the possibility of correcting for missing values with statistical methods.

A major upgrading of the data entry stage introduced recently is worth mentioning here: in fact, for the question on investments, it should be considered that in the ISAE version of the harmonised questionnaire the firm can possible answer "no investment" for any particular year on which the survey is carried. Until recently, if the firm was non answering for one of the years considered (i.e., in the case of missing data according to the definition above), this was considered as a "no investment" answer, even if no indication in this sense was provided by the firm. This procedure at the data entry stage could have possibly resulted in the past in an under estimate of the value of investments in a particular year. For this reason, starting from 2002, and looking as far back as the April 2000 database, all the micro-data have been carefully re-checked, distinguishing between actual missing data and "no investment" answers; this has allowed to treat differently actual missing data (using the statistical procedure discussed below) and the "no investment" answers, that were computed as such (i.e., zero investment for that firm in a particular year) in the processing of the results[1].

3.1.2. Outliers

The presence of outliers is fairly common in surveys based on a random sample as those carried by ISAE[2]. More precisely, from a theoretical point of view, it is possible to distinguish, on the basis of the Chambers (1986) classification, between non representative, and representative outliers: the former may be a consequence of mistakes made at the stage of collecting and editing the answers or entering them into the data base, the latter arising from the fact that results derive on a random sample, based on a selection of units, and not from the entire universe. In this sense they can be interpreted as a realisation of a case with a low probability.

Since non representative outliers are interpreted as errors they do not contribute to provide relevant information on the variability of non sampled units of the target universe, therefore it is necessary to proceed to their detection and correction (Smith, 1987). On the other hand, representative outliers provide information on non sampled units of the universe. They are correctly recorded and cannot be considered as errors. In any case their presence can also introduce a bias in the sample estimates especially if the parameter estimate is obtained weighting each observation with an "expansion factor"[3]. For this reason it is advisable to proceed to their treatment with appropriate statistical methods in

---

1 The authors wish to thank Paola Bellincampi and Massimo di Tommaso for their crucial contribution in dealing with the problems in the data entry stage.

2 See on this Banca d'Italia, 1996.

3 The expansion factor allows to take into account the differences existing between the structure of sample strata and population strata (see section 4 for a description of expansion factor).

order to reduce their impact on the sample estimates variability (see Battipaglia, 2000).

As for missing data, the identification of non representative outliers is performed in the data entry stage. In the first step, the analyst proceeds to verify possible inconsistencies “internal” to the same questionnaire: for instance, as far as quantitative questions on investment expenditures are concerned, a careful check on the consistency of the answers in two (or more) consecutive years is performed, having in mind also the relative dimension of the firm (e.g., a particular check is carried for large investments declared by small firms). In a second step, the inter-temporal consistency of the answers is checked, using consecutive questionnaires when available: for instance, in the case the same firm in two consecutive questionnaires indicates very different investment expenditures for the same year, this information is checked with that on revision plans and with the other qualitative answers of the firm. This procedures are applied both to quantitative and qualitative answers of the survey.

If the data are considered as “inconsistent”, in the sense specified above, they are checked for possible mistakes in the data entry stage or in the collection of questionnaire, and in case they are corrected in the database; then, for surviving outliers, the firms are re-interviewed in order to check for possible errors on their part in the provided answers. If the firm corrects the answer originally provided, outliers are considered of a non representative type, and they are duly corrected in the database. In the case the firm confirms its answer, outliers are considered as representative and the data are included in the processing of the results, if necessary treating them in a following step with appropriate statistical methods (see below).

3.2. Statistical procedures

3.2.1. Missing data

The detailed analysis of survey results shows that the use of re-interviewing techniques is not able to completely eliminate missing data problems from the data-sets. Table 3 shows non-response rate for the quantitative questions on investment expenditures in the period November 2000[1]-April 2004; in the case of sample surveys with unequal inclusion probabilities for different units (as is the case of the ISAE investment survey), response rates are calculated as follows (see OECD, 2003, p. 26):

[1] For the April 2000 survey it has not been possible to distinguish between missing data and “no investment” answers; for this reason the survey has not been considered in the following analysis.

$$M = \frac{\sum_{i=1}^{n''} \frac{1}{f_i}}{\sum_{i=1}^{n} \frac{1}{f_i}} \tag{1}$$

where: n=number of enterprises in the survey; n’’= number of non-responding firms; the f_i’s are the sampling probability of each reporting unit i, belonging to a given stratum, calculated as the ratio between N, the total number of reporting units in the target universe in each stratum, and n, the number of reporting unit in the sample, for the same stratum.

Missing data represent up to 5,6% of the total answers, not a negligible measure given the relatively low response rate characteristic of the investment survey. Moreover, ISAE usually provide data broken down by sectors and size classes; as shown in table 4 (calculated for the spring 2004 survey; similar results do emerge for the other surveys, available upon request), the rate of non response could differ among strata, becoming quite large in particular cells. Also, it should be considered that (anticipating in this some of the considerations of section 4 below) in the phase of results processing the presence of missing data implies the need to adjust the weighting scheme in order to take into account the differences in the number of answering firms in each strata. From these considerations stem the opportunity to minimize missing data presence in the database, at least those on quantitative variables, using statistical techniques to impute them.

Table 3. Missing data (as a weighted percentage of total answers)

	Spring survey				Autumn survey				
Investment at time: (spring/autumn survey)	2001	2002	2003	2004	2000	2001	2002	2003	2004
t-2/t-1	2.28	1.95	3.99	2.03	1.65	2.87	2.33	1.13	2.42
t-1/t	0.74	0.89	2.62	0.51	1.02	0.18	2.36	1.57	1.94
t/t+1	1.88	3.66	3.81	3.87	4.47	3.39	5.59	2.83	4.63

Table 4. Missing data by sectors, size classes and geographical partition (as percentage of total answer)

Spring 2004	t-2/t-1		t-1/t		t/t+1	
	MD % (*)	MD % weighted (**)	MD % (*)	MD % weighted (**)	MD % (*)	MD % weighted (**)
Textile, clothing, footwear	0.81	0.84	0.00	0.00	4.88	5.19
Paper, printing, publishing	4.17	3.57	0.00	0.00	8.33	11.98
Chemicals, rubber, plastics	3.23	2.98	0.00	0.00	3.23	2.12
Metallic & non metallic minerals	1.79	2.05	0.00	0.00	4.46	5.67
Mechanical sectors	1.63	1.43	0.82	0.7	2.04	2.28
Transport equipments	0.00	0.00	0.00	0.00	2.44	5.34
Other manufacturing n.e.c.	2.04	2.92	0.00	0.00	0.00	0.00
Food, beverages, tobacco	5.71	7.16	2.86	3.87	2.86	3.87
Small size firms	2.86	2.24	0.82	0.59	4.29	4.06
Medium Size Firms	1.18	1.2	0.00	0.00	1.18	1.22
Large Size firms	1.18	1.64	0.00	0.00	2.35	1.52
North-west	1.79	0.70	0.45	0.36	2.24	3.85
North-east	1.29	1.83	0.00	0.00	2.27	2.05
Centre	3.59	4.81	0.00	0.00	5.99	6.90
Mezzogiorno	3.05	3.02	2.29	2.84	3.05	2.82

(*) *(n"/n)x100* (**) $M = \sum_{i=1}^{n''} \frac{1}{f_i} \Big/ \sum_{i=1}^{n} \frac{1}{f_i}$

In this respect, in what follows a conditional mean imputation method is proposed, based on a simple regression model, which uses the information internally provided by the survey' questionnaire. More in particular, to estimate the missing information relative to the investment expenditure of a firm, slightly different models are proposed according to whether an estimate is needed for expenditures at time t-2, t-1 or t (in the spring survey) or at time t-1, t and t+1 (in the autumn survey). The models rely upon some preliminary assumption, in particular:

(i) firm response probability does not depend on the absolute value of missing investment itself, but only on the characteristics of the firms in the sample; in other words, we assume that missing data are Missing At Random (MAR): let Y be a matrix of complete data, and R the matrix of indicator variables whose values are zero or one depending on whether the corresponding elements of Y are missing or not. Assuming that data are MAR means that:

$$P(R \mid Y_{obs}, Y_{miss}) = P(R \mid Y_{obs})$$

where Y_{obs} are the fully observed variables and Y_{miss} those presenting missing data;

(ii) all the covariates in the econometric models adopted should be fully observed, i.e. there are not missing data in the variables used as regressors.

In what follows, we select the following covariates for the models estimating missing investment data in period t-1, t and t+1:

a) in order to estimate investment at time t-1 of the generic i-firm, the model uses the number of employees of the firm as scale factor. This variable is supposed to be always completely observed. Moreover, preliminary analysis of correlation matrix between quantitative variables in each data-set shows that investment and number of employees are strongly correlated;

b) for estimation of missing data in t and t+1, the models also include information on investment expenditures, respectively, for years t-2 (in case t-1 missing data are estimated), t-2 and t-1 (estimating missing data in t) and t-1, t (for estimating t+1 investment missing data). The investments of the i-firm may be thought to show some autocorrelation;

c) furthermore, as observed above, table 4 shows that partial non-response distribution substantially differs among strata, thus we also include in the regression dummies variables selecting the sector[1], geographical partition and size class of the firm, in order to properly compute conditional mean in each cell.

Finally, it should be considered that in this phase of analysis our main goal is to impute correctly the investment values at firm level, in order to estimate the population rate of investment variation, more than to preserve general distribution properties (as, for example, overall investment expenditure variability in the sample); thus a deterministic single imputation technique is adopted without adding any random effect to conditional mean. The following analysis of models adequacy to predict missing value is therefore conducted having in mind the aim described above.

Econometric Models

Models described in equations (2)-(4) are proposed respectively to estimate missing data on investment expenditure at time t-2, t-1 and t for the spring survey, while models in equations (5)-(7) estimate missing information for the years t-1, t, t+1 in the autumn survey:

$$I_i^{t-2} = const + \gamma_r + \alpha_d + \delta_s + Num_i + \delta_s Num_i + \varepsilon_{isr} \tag{2}$$

[1] On the basis of some preliminary results, in the model only the 8 sectors presented in table 1 above are considered for the construction of the dummy variables.

$$I_i^{t-1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-2} + \delta_s Num_i + \delta_s I_{isr}^{t-2} + \varepsilon_{isr} \qquad (3)$$

$$I_i^{t} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-2} + I_{isr}^{t-1} + \delta_s Num_i + \delta_s I_{isr}^{t-2} + \delta_s I_{isr}^{t-1} + \varepsilon_{isr} \qquad (4)$$

$$I_i^{t-1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + \delta_s Num_i + \varepsilon_{tisr} \qquad (5)$$

$$I_i^{t} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-1} + \delta_s Num_i + \delta_s I_{isr}^{t-1} + \varepsilon_{isr} \qquad (6)$$

$$I_i^{t+1} = const + \gamma_r + \alpha_d + \delta_s + Num_i + I_{isr}^{t-1} + I_{isr}^{t} + \delta_s Num_i + \delta_s I_{isr}^{t-1} + \delta_s I_{isr}^{t} + \varepsilon_{isr} \qquad (7)$$

In models (2)-(7), I_i^t are investment expenditures of the i-firm in a particular year t, Num_i is the number of employees of i-firm, γ_r α_d and δ_s are dummies variable accounting for the firm' 4 geographical partitions, 3 size classes and 8 ISAE sectors, ε_{isr} is a normally distributed, *i.i.d.* stochastic term.

The model is estimated both using standard OLS and Weighted Ordinary Least Squares (WOLS), obtained applying OLS regression on observations multiplied by the inverse of their sample probability, 1/fi. WOLS methods are often adopted in order to take into account random samples in which sample probability varies among units (Winship and Radbill, 1994). The models' robustness is then tested adopting the standard inference procedure. In particular, we assume that data could be affected by Heteroskedasticity that is somewhat typical of regressions on levels observed in heterogeneous units. Hence, in order to compute the proper confidence intervals, we use the White (1980) Heteroskedasticity consistent estimator for the coefficients standard errors. Moreover, the robust estimator is fairly recommended when weighted regressions are used, as weighting observations clearly introduces Heteroskedasticity (see again on this Winship and Radbill, 1994).

Tables 5—6 show the R^2 and the usual Wald test on the significance of the variables used in models (2)—(4) for spring 2004 survey and (5)—(7) for Autumn 2003 survey[1], estimated both with OLS and WOLS. R^2 are generally pretty high, showing that the models are able to explicate a large proportion of the original variance; according to the Wald test, geographic dummies are not jointly significant in both data-sets, while in the Autumn 2003 survey also the size class effect is not statistically significant. Furthermore, the comparison between OLS and WOLS regressions highlights some differences in the jointly significance of coefficients for dummies, mainly for Models (2) in Spring 2004 and (7) in Autumn 2003 survey.

<u>The complete case analysis</u>

1 Similar results for the other surveys are available with the authors upon request.

As stated above, models (2)-(7) are aimed at correcting for missing data, in order to provide a better estimate of the aggregate rate of growth of investment. For a better assessment of the capacity of the models of reaching this goal, in the following we propose also a "complete case" analysis, eliminating from the database the firm with missing data. Hence, we perform a simulation analysis to verify whether the estimates of investment variation between two periods radically differ with and without missing data imputation, and whether differences arise when using WOLS instead of OLS. More in detail, rate of variation r_c is computed on a "complete" data-set, obtained by deleting firms that present missing data for investment in t-1, t or t+1; in a second step, we select three random sub-sample of firms, i.e. one for firms presenting missing data for I_{t-1}, one for those with missing data in It and the last for the firms with non-response for I_{t+1}, and in these we substitute for generated missing data, alternatively, with the estimate provided using OLS and WOLS. Thus the investment rate of variation is computed again in the three different settings:

1) "complete case" rate of variation (r_{CC}), obtained deleting from the data set the firms presenting missing data in any of the investment level in different t;
2) the rate of variation imputing value with OLS regression models (r_{OLS});
3) the rate of variation imputing value with WOLS regression models (rW_{OLS}).

Each rate is computed for several iterations, each time extracting different random sub-samples of missing data. Table 7 summarizes the differences between the rate computed on the complete data-set (rc) and the mean values of the rates estimate for the Spring 2004 survey with the methods listed in 1-3 above (namely r_{CC}, r_{OLS} and r_{WOLS}) for 10 independent iterations. We can observe that complete case analysis induces stronger bias than the adoption of the imputation method we proposed. Also, the WOLS regressions seem introduce less bias than OLS.

To conclude with, the models we propose could present some misspecification problems on particular data-sets. At the same time, the simulation shows that the estimation of the rate of investment variation, filling in missing data with WOLS regression methods, provides better estimate than those obtained with OLS regression and, above all, those obtained with case deletion. Actually, we could observe a slight bias adopting WOLS regression, though we expect that the bias will disappear as number of iterations radically grows.

From this considerations, stem the choice to adopt the models (2)—(7) described above; also, the WOLS regression method has been applied.

Table 7. Simulation results for r_c-r_{IMP}

Spring 2004		**CC**		**WOLS**		**OLS**	
t/t-1							
	r_c	**$m(r_{CC})$**	**$r_c - m(r_{CC})$**	**$m(r_{WOLS})$**	**rc – $m(r_{WOLS})$**	**$m(r_{OLS})$**	**$r_c - m(r_{OLS})$**
Small size firms	-24.02	-22.70	**-1.32**	-23.30	**-0.72**	-24.44	**0.42**
Medium Size Firms	-18.41	-16.01	**-2.40**	-17.80	**-0.61**	-18.89	**0.48**
Large Size firms	-8.34	-8.49	**0.15**	-8.41	**0.07**	-8.47	**0.13**
Total	-15.09	-14.29	**-0.80**	-14.79	**-0.30**	-15.43	**0.34**
t+1/t							
	r_c	**$m(r_{CC})$**	**$r_c - m(r_{CC})$**	**$m(r_{WOLS})$**	**rc – $m(r_{WOLS})$**	**$m(r_{OLS})$**	**$r_c - m(r_{OLS})$**
Small size firms	-16.46	-17.77	**1.31**	-14.39	**-2.07**	-12.39	**-4.07**
Medium Size Firms	-10.66	-12.81	**2.15**	-10.72	**0.06**	-10.64	**-0.02**
Large Size firms	17.53	16.74	**0.79**	17.84	**-0.31**	17.89	**-0.36**
Total	3.16	2.06	**1.10**	3.88	**-0.72**	4.49	**-1.33**

3.2.2. Outliers

Once the data have been checked in the data entry stage for the presence of possible non representative outliers, they are included in the database in order to proceed with the processing of the results. In this respect, figure 1 presents the sampling distribution of the absolute level of investments in the April 2003 survey[1] for the period t, t-1 and t-2, considering a fixed investment class and leaving open the last investment class. Empirical distributions of firms' answers are asymmetric, with more than 70% of observations concentrated in the first classes and a fairly number of outliers in the last open class. The presence of outliers is confirmed also looking at the investment-per-worker distribution of answers: in this respect, figure 2 shows the same asymmetry in the firms' data, indicating that this asymmetry doesn't depend only on the size of the firm, i.e. it is not due exclusively to the sample composition and to the presence of large firms in the sample. Rather, it seems to be a proper feature of the answers provided by the firms; in particular, in the case of per capita investments, we can see that more than 33% of the answers lies in the first two investment classes.

[1] The distributions are fairly similar for all the surveys of the period April 2000—April 2004.

Figure 1. Empirical distribution of the absolute level of investments (April 2003)

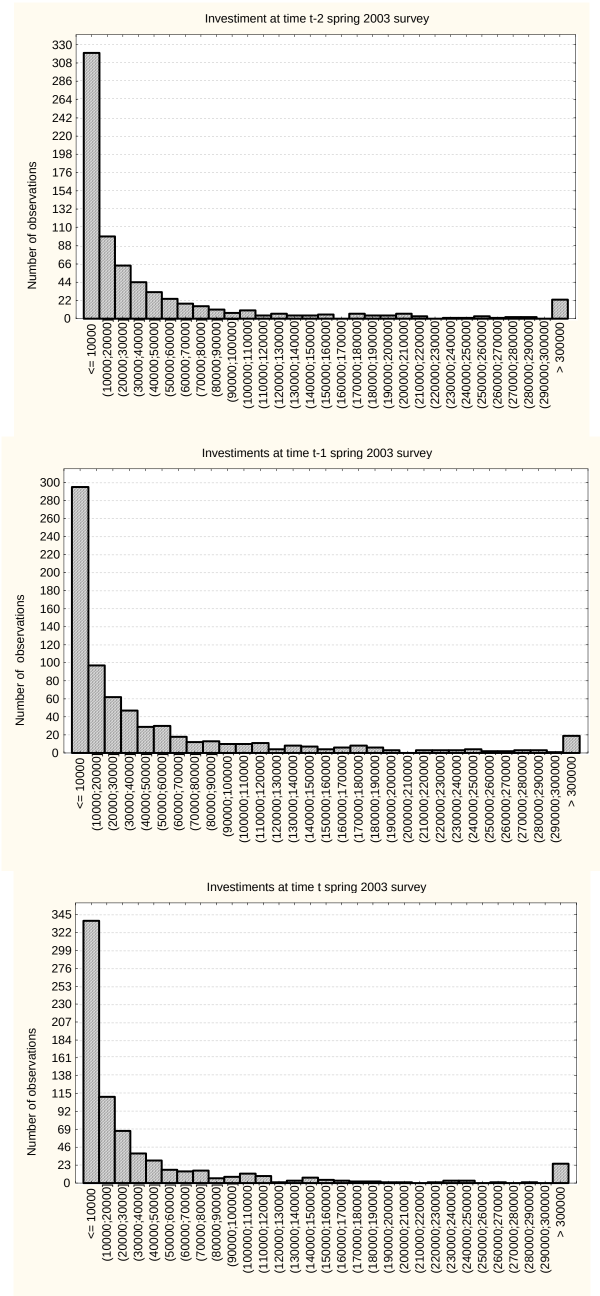
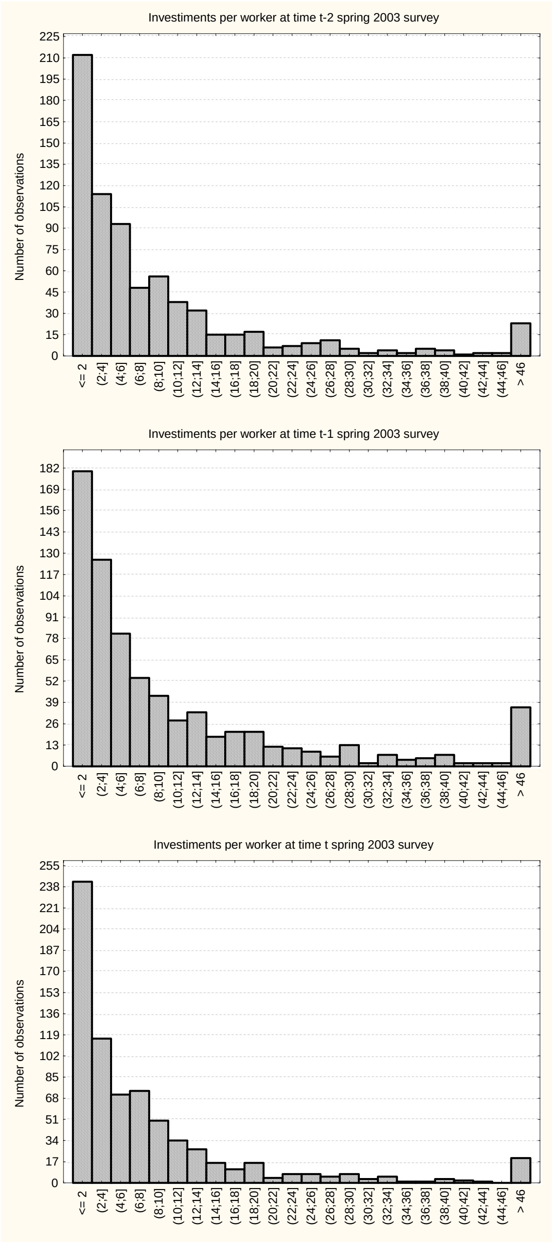
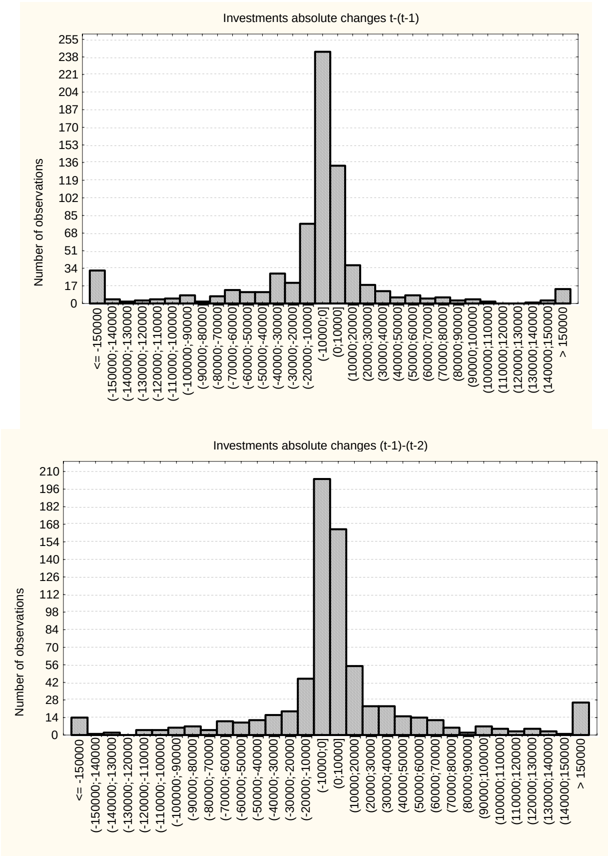


Figure 2. Empirical distribution of per-worker level of investment (April 2003)



Since the main purpose of the investment survey is to determine an estimate of the growth investment rate for the Italian manufacturing firms, we inspect the empirical distribution of absolute changes of investments. Looking again at the results for the April 2003 survey, we can notice (as we expect) that these distributions (figure 3) are characterised by a certain degree of symmetry and by the presence of outliers in both tails[1].

Figure 3. The distribution of absolute changes of investment (April 2003)



The analysis of the distribution of firms' answers has provided some first evidence that extreme observations of representative type are present in the database. In order to limit their influence on the variability of estimates in what follow we use appropriate statistical methods. More in particular we calculate robust estimators of the growth investment rate, and we compare the results with that obtained simply aggregating firm' answers on the basis of the procedure described in section 4 below: a large difference between the results may be

[1] Whereas in the empirical distribution of investment levels representative outliers arise from those firms that effect a huge amount of investments, in the distribution of absolute changes they are rather originated by those firms whose investments are strongly revised with respect to the previous year.

interpreted as evidence of a significative impact of outliers on the sample estimates.

From a theoretical point of view, different approaches have been proposed in order to produce robust estimators for data coming from sample surveys[1]; on the basis of the existing literature the choice of the appropriate estimator depends on rather heterogeneous factors, like the aim of the research, the kind of information we want to obtain from data, the degree of knowledge of the phenomena involved in the study (Gismondi, 2002). However, it is possible to select some desirable properties that a robust estimator should have: such an estimator should indeed be robust, easy to implement and to interpret, and show low variance and a low bias (Hulliger, 1999). The most commonly used robust estimators of sample distributions are trimmed and winsorized means. In what follows, we propose some considerations on the use of trimmed and winsorized means and we motivate the application of the winsorized mean for the evaluation of the impact of outliers on ISAE survey data.

In particular, trimmed means are calculated eliminating the k smallest and the k highest observations, after having determined thresholds on both sides of the distribution: i.e., the trimmed mean eliminate all the k observation exceeding (both from the left and from the right) a predetermined threshold. The threshold may be fixed “a priori” (e.g., observation exceeding a certain value) or determined eliminating a fixed fraction of the original observation (e.g. the 10% of the original distribution). Denoting y_k as the trimmed mean we have:

$$y_k = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} y_i \tag{8}$$

Also in the case of the winsorized mean, firstly it is necessary to determine the fraction of observation to be considered as extreme values (e.g. those exceeding a particular value, or a fixed proportion of the distribution); in the second step, the so-defined extreme values are replaced with the values preceding the first observation excluded from the distribution: in other words, in this case extreme observations are not eliminated from the distribution, but are substituted with the last available observation that lies just before the predetermined threshold. Defining y_k, as the winsorized mean we have:

$$y_k = \frac{1}{n}\left\{(k+1)y_{(k+1)} + \sum_{i=k+2}^{n-k-1} y_{(i)} + (k+1)y_{(n-k)}\right\} \tag{9}$$

With respect to the trimmed means, winsorization has the advantage of preserving all the observations, giving a different weight to extreme values of the sample data distribution. This feature turns out to be particularly useful when

[1] See Chambers (1997) for the model based and Hidiroglou and Srinath (1981) for the designed based approach.

researchers have to deal with a limited number of observations, as it is the case for the ISAE investment survey. In order to produce robust estimates of the investment variation rate with the winsorized mean is possible to use two different approaches (see again Battipaglia, 2000), i.e. one-sided and two-sided winsorization. The first one is applied to the distribution of relatives changes of investments, I_i^t / I_i^{t-1}. The second one is applied to the distribution of absolutes changes of investments $I_i^t - I_i^{t-1}$ that is symmetric and shows outliers in both sides of the distribution. Since one sided winsorization introduces by construction a bias in the estimates, in this paper we decided to use only the two sided winsorization, calculated as follows[1]:

$$\Delta I^{wins}{}_t = \frac{\sum_{i=1}^{n} f_i^s \left[I_i^t - I_i^{t-1} \right]^{wins}}{\sum_{i=1}^{n} f_i^s I_i^{t-1}} \tag{10}$$

where ΔI_t^{wins} is the winsorized growth investment rate, f_i^s is the expansion factor (see section 4) and $\left[I_i^t - I_i^{t-1} \right]^{wins}$ is the winsorized distribution of absolute changes of investments for a given cut off.

Table 8 provides the estimates for the growth investment rate calculated applying (10) to the results of the ISAE investment surveys for the period November 2000-November 2004, using different cut off rates, and compares them with the non-robust estimator obtained on the basis of expression (11) described below in section 4. In particular the winsorisation has been applied, respectively, to an amount equal to 2, 5 and 10% of total observations.

The results in table 8 show that the use of winsorized means on ISAE data introduces relevant changes in the estimate of the growth rate of investment providing evidence that representative outliers may have a relevant impact on the final estimates. On the other hand, results do not appear to be particularly sensitive to the choice of the cut off rate. This is due to the fact that the observations playing a particular influence on the parameter estimates are mainly concentrated in the last 2% of the distribution. In this sense, results in table 8 seem to suggest that the evaluation of survey results should take into consideration the role that extreme observation play on the estimate of the growth rate of investments; for this reason, since the autumn 2004 survey ISAE publish survey results in term of the robust estimators described above.

[1] In the following, we show the application of winsorized mean on the original data, pre-multiplied by the sample probability (f); see below, section 4.

4. Processing the results of the ISAE investment survey

Until spring 2002 the processing of the Investment survey results was based, for quantitative data on investment plans, on the simple sum of firm' level data on investment expenditures. The growth rate of investments was then calculated on the basis of the point estimates of yearly capital expenditures. For the structure of investments, and for the qualitative data on revision plans, aggregated results were calculated as a weighted average of firms' answers, using firm investments as weights.

The simple sum of firms' answers previously used by ISAE to provide an estimate of total manufacturing investment expenditures did not take into consideration the structure of the sample; this implies that this method provides only a sample estimate, without giving proper information about the target universe. For this reason, ISAE proposes to introduce a new procedure in order to obtain an estimate of the investment expenditures of the target universe; this estimate should be based on the structure of the actual sample. In particular, accordingly to the methods used for a similar survey by the Bank of Italy (Banca d'Italia, 1996), in the new aggregating procedure the sample information are "expanded" to obtain the target universe estimation, using the sample probabilities as the expansion factor. We have:

$$I^t = \sum_{i=1}^{n} f_i \; I_i^t \qquad (11)$$

where: I^t = investment expenditures of the manufacturing sector
I^t_i = investment expenditures of the response unit i
$f_i = N_h/n_h$, i.e the inverse sample probabilities, calculated as the ratio between the number of firms in the h-stratum of the target Universe (N_h) and the number of the response units in the h-stratum of the sample (n_h).

The use of the expansion factors f_i allows taking into consideration, in the estimate of the capital expenditures of the manufacturing sector, the structure of the sample. In practice, the use of (11) implies that the response units are considered to be representative of the firms included in the each stratum (defined on the basis of the region, sector and size of the response unit) of the target universe; in other words, using (11) means to attribute to the non sampling units belonging to each strata the same investment expenditures of the response units sampled in the strata.

For the question on the structure of investments, response units are asked to report whether capital expenditures are aimed at i.) replacing worn-out plant or equipment or ii.) extend production capacity, or else if they are designed either to iii.) streamline production or to iv.) other objectives (for example, pollution control, safety, etc.). Firms are asked to indicate the percentage of total expenditures destined to each of these categories in the current and the following year. Aggregate results are obtained as a weighted average of the firm level structure of the investment expenditures, using as weights the information on actual investment expenditures for the year considered, provided in the same questionnaire:

$$\mathrm{A}_j^t = \sum_{i=1}^{n} \alpha_{ij}^t \frac{f_i \; I_i^t}{I^t} \tag{12}$$

where: $_j^t$ = aggregate share of investment in each of the j-possible destination (j=1,…4)

$_{ij}^t$ = share of investment in each of the possible destination for the firm i

$f_i \; I_i^t$ = total amount of investment for firm i, multiplied by its inverse sampling probability

I^t = investment expenditures of the manufacturing sector (see 11 above).

The EU harmonised questionnaire includes this question only in the autumn questionnaire; in the ISAE implementation of the survey, the question is asked also in the spring questionnaire[1].

Besides quantitative information on investments plans and structure, the survey provides also qualitative assessment on factors influencing eventual revision of the original investment' plans during the year. More in detail, firms are asked to report if investment expenditures are influenced by i.) the current trend of demand, ii.) the availability of financial resources and profits, 3.) technical factors and iv.) other factors (which may include public policies). For each of these factors, firms indicate if they are considered as very stimulating, stimulating, without influence, or either with a limiting or very limiting influence on investments. In other words, for any k-factor (k= 1, …. 4), multiple-choice percentages should be calculated, having in this case j=5 reply options, and using firms investment expenditures as weights:

$$R_{k,j}^t = \sum_{i=1}^{n} r_{i,k,j}^t \frac{f_i \; I_i^t}{I^t} \tag{13}$$

where: $R^t_{k,j}$, = aggregate percentage of answers of type "j" (for j=1,…5), concerning factor k (for k=1,…4)

$r^t_{i,k,j}$, = 0/1, depending on the answer of type "j" given by enterprise i, concerning factor k

---

[1] The ISAE version of the harmonised questionnaire includes also other qualitative questions about investment plans for the years ahead, and a more detailed analysis of expenditures destination; for a description, see the Italian version of this paper, (Cesaroni, Malgarini, Rocchetti, 2005).

5. Final considerations and future research

The survey on investments of manufacturing firms differs from the other surveys usually performed by ISAE in that it also provides quantitative information about actual investment plans and their structure, besides more usual qualitative data about reasons influencing them. The gathering of quantitative information raises further potential problems concerning the quality of the data, especially for the possibility of higher non response rates and referring more in general to the issue of data reliability.

In this respect, the paper shows that only about 20% of the actual sample (on a theoretical sample of 4.000 Italian manufacturing firms) usually responds to the survey; moreover, the ex-post analysis of the questionnaires also demonstrates that non-response rates for the question on investment plans may amount up to almost 6% of the actual sample. On the other hand, concerning the reliability of the data, the analysis of the distribution of the firms' answers provides evidence on the presence of outliers, even after controlling for non sampling errors with the use of deterministic techniques. High non response rates and reliability problems linked to sampling errors may diminish the overall quality of survey results.

For these reasons, in this paper we propose the use of appropriate statistical methods dealing both with high non response rates and outliers. More in particular, concerning the treatment of non-responding firms, the paper presents a fairly simple econometric model aimed at estimating missing data on the basis of information, gathered in the questionnaire itself, about the firm geographical location, size, sector of activity and the past investment of the firm itself, when available. Albeit presenting some specification problems, the model seems however to provide us with estimates of missing data that an ad-hoc simulation exercise has proved to be able to improve the general inference about the parameter of interest (i.e., the rate of growth of investment of manufacturing firms). Further research in this field is however advisable, in order to have a better assessment of the performance of the models, extending the use of simulation techniques and improving the general fit. Moreover, the use of more modern interviewing methods (e.g. CATI system, or techniques based on the use of the Web) may help to generally improve survey response rates, as it has been the case for the monthly survey on the manufacturing sector.

As far as the treatment of outliers is concerned, the paper shows that the distribution of answers to the general question on investment plans is actually characterised by a certain degree of asymmetry and by a rather "fat" right-hand tail (or, by long symmetric tails if looking at the distribution of the rate of change of investments), even after controlling for non representative outliers with deterministic methods. This result is interpreted as a first evidence of the existence of extreme values in the answers provided by the firms, suggesting that sampling outliers may well be present in the data set. The application of the winsorized mean has confirmed this hypothesis, showing that the inference on the parameter of interest may be significantly affected by the application of this method, and suggesting the opportunity of publishing robust estimator results. Further research is however needed, in order to provide a better assessment of the effects of the application of these methods on the properties of the resulting estimators.

To conclude with, section 4 presents the methods used by ISAE to aggregate firm-level data and provides estimates about the growth rate of manufacturing investments, their structure, and the factors influencing investment' decision of the firms. Finally, the Appendix provides the complete data-set obtained applying the techniques described in the paper to the April 2000-November 2004 surveys.

REFERENCES

BARNETT, V. LEWIS, T. (1994) Outliers in statistical data, third edition.

BATTIPAGLIA, P. (2000), "Robust Estimates of Investment from the Bank of Italy's Business Survey", London School of Economics Statistics Research Report, LSERR56, Aprile 2000.

BANCA D'ITALIA (1996), "Metodi e risultati dell'indagine sugli investimenti delle imprese industriali", Supplementi al Bollettino Statistico, Note Metodologiche e informazioni statistiche, Nuova Serie, Year VI, n. 59, 21 October .

BANCA D'ITALIA (various years), Indagine sugli investimenti delle imprese industriali, Sintesi dei risultati, Roma.

CESARONI, T. MALGARINI, M., ROCCHETTI, G. (2004), "L'inchiesta ISAE sugli investimenti delle imprese manifatturiere ed estrattive: aspetti metodologici e risultati", Documento di lavoro ISAE n. 50/2005

CHAMBERS, R.L. (1986), "Outlier Robust Final Population Estimation", Journal of American Statistical Association, Vol. 81, n. 396.

EUROPEAN COMMISSION (1997), DGII's Qualitative By-Yearly Business Survey on Investment, II/082/97. En, paper presented in the "Seminar on short term investment surveys", Luxembourg, January 27—28.

EUROPEAN COMMISSION (2002), The Joint Harmonized EU Programme of Business and Consumer Surveys, User Guide 2002.

GISMONDI, R. (1992) "Un confronto tra metodi di identificazione di osservazioni outliers in indagini longitudinali finalizzate alla stima di una variazione: proposte teoriche e verifiche empiriche".

HIDIROGLOU, M. SRINATH, K. (1981) "Some estimators of a population total from simple random sample containing large units".

KOKIC, P, P. SMITH (2000), Outlier Robust estimation in sample surveys using two-sided winsorisation,

MALGARINI, M., MARGANI, P. e B.M. MARTELLI (2005), "New Design of the ISAE Manufacturing Survey", Journal of Business Cycle Measurement and Analysis, Vol. 2, n. 1, OECD, Paris.

MARTELLI, B. M. e G. ROCCHETTI (2004), "The ISAE Market Services Survey: Methodological Upgrading, Survey Reliability, First Empirical Results", 27th CIRET Conference, Varsavia, Settembre 2004.

OCSE (2003), Business Tendency Surveys: A Handbook, Paris 2003.

SMITH, T.M.F. (1987), "Influential observation in survey sampling", Journal of Applied Statistics, Vol. 14, pp.143—152.

WHITE, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticy", Econometrica vol. 48, pp. 817—38.

WINSHIP, C. and L. RADBILL, "Sampling Weights and Regression Analysis", Sociological Methods and Research, Vol. 23 No. 2, pp 230—257.

Appendix

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 293—307

REGIONAL DIFFERENTIATION OF BUSINESS CYCLES IN POLAND, 1999—2004

Jacek Fundowicz, Krzysztof Łapiński, Marcin Peterlik, Bohdan Wyżnikiewicz[1]

ABSTRACT

The paper aims at analysis of business cycles in 16 Polish voivodships. This analysis was based on methodology applied for surveys of international business cycles synchronization. Calculations were based on industrial sales and unemployment time series for 16 regions and entire Poland. In order to eliminate seasonality the X-11-ARIMA procedure was applied. Next, in order to identify trend in time series, Hodrick — Prescott (HP) filter was applied.

Four groups of regions with similar course of business cycles fluctuations were identified. Two of them are characterized by similar fluctuations of industry production levels and two other — with unemployment rate convergence. The regions in the identified groups are also located in close geographical neighborhood. In three of four groups of regions, the analysis of some macroeconomic indicators allowed authors to indicate the hypothetical leader (leaders) of each "macro-region".

Key words: Business Cycles, Regional Economic Activity.
JEL Classification: E32, R12.

1. Introduction

The paper aims at analysis of business cycles in 16 Polish regions (voivodships). Groups of regions with similar pattern of business cycles developments and fluctuations were identified. It was assumed that regional business cycles fluctuations in Poland are significantly differentiated.

Interconnections between business cycles on national and regional levels are complex. On the one hand, regions of higher economic potential influence domestic macroeconomic situation and on the other hand, regions with weaker

[1] The Gdansk Institute for Market Economics (Instytut Badań nad Gospodarką Rynkową), Gdańsk, Poland. E-mail addresses: jacek.fundowicz@ibngr.edu.pl, krzysztof.lapinski@ibngr.edu.pl, marcin.peterlik@ibngr.edu.pl, bohdan.wyznikiewicz@ibngr.edu.pl.

economic potential normally react with a delay to domestic macroeconomic changes. Pattern of dispersion of business cycle impulses from stronger regions to weaker regions is another research issues analyzed in the paper.

Administrative reform of 1999 has reduced a number of voivodships from 49 to 16 and therefore the period of observation only covers 1999—2004 (March). One of reasons for such reform was to create regions large enough to be able to conduct effective and independent regional policy. Population of new regions varies from 1,0 to 5,1 million people, GDP from 8,4 to 75,5 billion euro (PPS) in 2001[1] Little is known about the nature of regional business cycles developments in Poland. Few pioneer surveys were undertaken in this area[2].

As far as business cycle dispersion is concerned, regions within a country can not be regarded as countries within a continent like, for example European Union and individual countries within EU–15 or EU-25. Economic connections are stronger between regions within a country than between countries. On the other hand, regions in Poland are historically biased. Such factors as the level of industrialization, regional dependence on agriculture and low propensity to liberal (or non-liberal) behavior are historically or geographically rooted. Therefore, regions in Poland react differently on business cycles developments.

2. Data and methodology

Analysis of links between regional business cycles fluctuations was based on methodology applied for surveys of international business cycles synchronization. This methodology assumes that business cycle fluctuations can be identified by decomposition of time series on a trend and deviation from such a trend. Such methodology is characteristic for the real business cycle approach. It is also applied in eclectic models which unifies features of classic models and ones belonging to so called new classical economics. Oscillations around the trend represent the path of business cycles fluctuations while correlation coefficients between deviations from trends in regions — the convergence level of business cycles fluctuations between regions.

Due to the 1999 reform of administrative division in Poland which reduced the number of regions from 49 to 16, the analysis covers comparable data from January 1999 to March 2004. It has to be remarked, that the survey focus on links (ties) between medium-term regional business cycles fluctuations and not the

---

[1] Eurostat data.

[2] Michał Domański, Wiktoria Parchomenko, Marcin Peterlik, Bohdan Wyżnikiewicz (2001), Realisation of Entrepreneurs Expectations in Research with Business Condition Test, [in:] Business Tendencies on the Market of Financial Services in Poland, Akademia Ekonomiczna, Poznań.

course of business cycles. So far, during economic and system transformation in Poland only one full business cycle covering 1991—2001 has been identified[1].

The main variable describing the course of business cycle fluctuations is gross domestic product. Regional quarterly GDP estimates do not exist for Poland and annual time series are too short to look for statistically significant relationships. For this reasons the survey was conducted on two available monthly series: on regional industrial sales and on unemployment rates.

In the case of industrial sales the attempt was undertaken to use data in constant prices. Producers Price Indices (PPI) were regionally differentiated. The Central Statistical Office (GUS) does not publish regional PPI, but publishes regional Consumer Price Indices. Regional PPI were estimated for regions using a pattern of regional CPI proportions.

Another factor of distortion for the description of business cycles relationships using industrial sales data is the regional differentiation of industrialization in Poland. This issue is important since in transition economies the industry diminishes its contribution to the GDP. Such distortion was eliminated by applying constant-base (January 1999=100) indices of industrial sales for each region. The analysis was conducted with the data compiled in such a way, which reflected real changes of output.

Both dynamics of regional industrial production and changes in unemployment rates depend on seasonality. The strength of seasonal factors is similar in individual regions, however it is not identical. Analysis on seasonally unadjusted time series could cause interference to similarity between regions as related to real connections between them. Correlation coefficients between seasonally unadjusted series could be overestimated. They would not show similarities between business cycles fluctuations, but similarities between seasonality patterns.

In order to eliminate seasonality the X-11-ARIMA procedure was applied. This procedure was developed in the mid sixties by the US Bureau of the Census. Currently the majority of econometric and statistical software has option of adjusting time series with ARIMA related methods[2].

The next stage of analysis was the identification of medium-term trends for industrial sales and unemployment in each region and later deviations from the trends expressed in percentages. It was assumed that such deviations represent the course of business cycles fluctuations.

[1] Bohdan Wyżnikiewicz (ed.) (2004), Macroeconomic Situation and Forecasts for Polish Economy 2003—2006, GIME Quarterly Bulletin No. 41, January 2004.

[2] Jacek Kotłowski (2002), The Methods of Time-Series Smoothing with ARIMA Models, Papers and Proceedings No. 73, Research Institute of Economic Development, Warsaw.

In order to identify trend in time series free from seasonality and price effects, Hodrick — Prescott (HP) filter was applied[1]. The methodology of HP filter relies on time series decomposition to growth element representing a trend and to cyclical element[2]. It can be presented as:

$$y_t = g_t + c_t, \qquad \text{for } t = 1, \ldots, T, \tag{1}$$

where: y_t – empirical time series,
g_t – growth element,
c_t – cyclical element.

Finding series (g_t) and (c_t) is reduced to solving:

$$\min_{g_t} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} [(g_{t+1} - g_t) - (g_t - g_{t-1})]^2 \right\}, \tag{2}$$

where λ is so called "smoothing parameter" which is positive. Expression $(y_t - g_t)$ is a cyclical element ($c_t = y_t - g_t$). When $\lambda \to 0$, trend moves to the initial series $(c_t \to y_t)$, when $\lambda \to \infty$, then trend is linear. Since in literature there is no unanimity on optimal value of parameter λ, the values proposed by Hodrick and Prescott are commonly accepted:

- $\lambda = 100$ for annual time series,
- $\lambda = 1600$ for quarterly time series,
- $\lambda = 14400$ for monthly time series.

Consequently, in this paper the value of 14400. was adopted for parameter λ. The results in this stage of calculations were time series of deviations from the trend presented as:

$$\theta_{it}^{I} = \frac{c_{it}^{I}}{g_{it}^{I}} * 100\% \qquad \text{and} \tag{3}$$

$$\theta_{it}^{U} = \frac{c_{it}^{U}}{g_{it}^{U}} * 100\%, \tag{4}$$

[1] R. J. Hodrick, E. C. Prescott (1980), Postwar US Business Cycles: An Empirical Investigation, Working Paper, No. 451, Carnegie — Mellon University, Pittsburg; reprinted [in:] Journal of Money, Credit and Banking, Vol. 29, No. 1, February 1997, p. 1—16.

[2] In the paper a cyclical element concerns medium-term fluctuation, which means that the term "cyclical" was used in traditional meaning.

where θ^I are series of deviations for industrial sales, and θ^U – series of deviations for unemployment rates. Sub-indices i are for regions and entire Poland (then $i = 1, 2, ..., 17$).

The next step of calculation was cross correlation coefficients between percentage shares of deviations in a growth element. This procedure was applied for both industrial sales and unemployment rates. As a final output two matrices were received:

$$I = \begin{pmatrix} cor\theta^I_{1,1} & & & \\ cor\theta^I_{2,1} & cor\theta^I_{2,2} & & \\ \dots & \dots & \dots & \\ cor\theta^I_{17,1} & cor\theta^I_{17,2} & \dots & cor\theta^I_{17,17} \end{pmatrix}, \tag{5}$$

$$U = \begin{pmatrix} cor\theta^U_{1,1} & & & \\ cor\theta^U_{2,1} & cor\theta^U_{2,2} & & \\ \dots & \dots & \dots & \\ cor\theta^U_{17,1} & cor\theta^U_{17,2} & \dots & cor\theta^U_{17,17} \end{pmatrix}. \tag{6}$$

The results were presented in the Appendix in the Tables A1—A2.

3. Regional business cycles fluctuations

3.1. Results of analysis

Four groups of regions with similar course of business cycles fluctuations were identified using described above research methodology. Calculations were based on industrial sales and unemployment time series for 16 regions and entire Poland. Graphs showing deviations from trends for each region are presented in the Figures A1—A2 in the Appendix. Correlation coefficient between regional deviations from trends are presented in tables A1—A2.

The main feature differentiating the course of two time series is higher volatility of seasonally adjusted industrial sales growth rates than unemployment rates. In the case of unemployment rates, average variation coefficient for 16 regions was 0,15 and changed from 0,08 to 0,22. In the case of industrial sales, growths rates average level of variation coefficient was 0,44 and changed from 0,27 to 0,84.

Such differences can be explained by nature of two variables. Industrial output is a typical economic stream. Its dynamics depends on current conditionalities and may vary from one period to another.

The level of unemployment rate is more stable. First reason is connected with low elasticity of labor market in Poland. Labor market regulations make impossible fast adjustment of labor demand to changes in business conditions. Second reason of lower flexibility on labor market than that of industrial production is connected to the fact, that changes in output result from changes in demand, which are independent from companies. They affect all companies in the same time (if sector differences are not taken into account). Changes in employment level depend from companies decisions and therefore they do not taken in the same time. Due to different employment policy in companies, changes in employment level appear with time lags and are not uniform. Therefore the unemployment variability is smaller.

It should be pointed out that both analyzed variables are sensitive toward business cycle factors, however, as shown above, with different elasticity.

The consequence of the described differences is higher synchronization level of business cycles fluctuations of the unemployment rates than of the industrial sales. It is reflected in correlation coefficients of deviations from trends. Correlation coefficients for unemployment deviations change between 0,43 and 0,95 with average 0,77. Corresponding values for industrial sales are between -0,09 and 0,89 with average 0,35. Table 1 shows main parameters.

Table 1. Correlation coefficients for regional unemployment rates and industrial sales

	Unemployment rates		Industrial sales	
	Correlation coefficients, seasonally adjusted time series	Correlation coefficients deviations from the trend	Correlation coefficients, seasonally adjusted time series	Correlation coefficients deviations from the trend
MIN	0,08	0,43	0,27	-0,09
MAX	0,22	0,95	0,84	0,89
Average	0,15	0,77	0,40	0,35

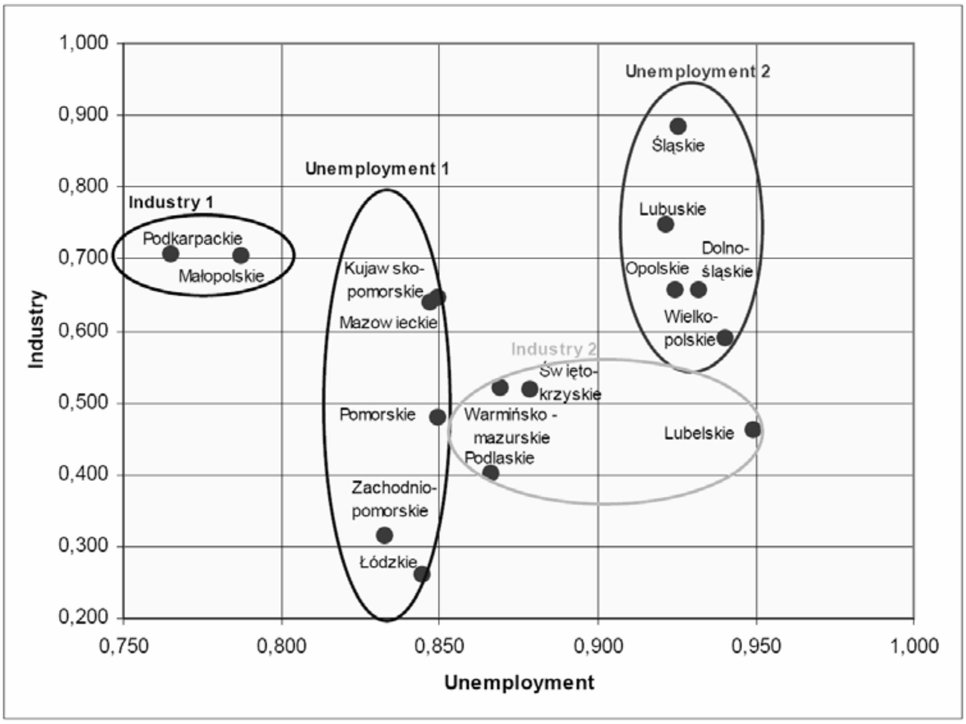
Source: Authors' calculations.

3.2. Groups of regions

Graph 1 illustrates the distribution of correlation coefficients between 16 regions and entire Poland for two variables. There are correlations between unemployment rates in regions and entire Poland on horizontal axis presented, and correlations of industrial sales on vertical axis. The higher the correlation coefficient, the higher level of business cycles convergence. Low correlations mean differences between fluctuation patterns.

The closer position of a given region to the upper right corner of the graph, the biggest similarity of regional business cycles fluctuation to national pattern. Close position of two regions (in both dimensions) shows relative synchronization of business cycle fluctuations.

Figure 1. Correlation coefficients between regional and national unemployment rates and industrial sales, January 1999 – March 2004



Source: Authors' calculations.

Two-dimension presentation of results was widened by identification of four groups of regions with similar pattern of medium term business cycle fluctuation. Inclusion of regions to groups Industry 1 and Industry 2 were due to similarity of business cycles fluctuations of industry production, while groups Unemployment 1 and Unemployment 2 were formed when fluctuation similarity of unemployment rates appeared. Identified groups were marked with ellipses.

It should be pointed out, that regions in the identified groups are located in close geographical neighborhood. In other words, regions in one group create geographical "macro-regions" from the point of wiev of observed similarities. It is shown on the Map 1.

Map 1. Groups of regions with similar course of business cycles fluctuations

Source: Authors' calculations.

Two regions of south-eastern Poland, Małopolskie and Podkarpackie, form the group Industry 1. The group Industry 2 is created by regions from so called "eastern wall of Poland" (Warmińsko-Mazurskie, Podlaskie, Lubelskie) and region Świętokrzyskie. The group Unemployment 1 includes regions of northern Poland (Zachodniopomorskie, Pomorskie, Kujawsko-Pomorskie) and two centrally situated regions (Mazowieckie i Łódzkie). The fourth group, Unemployment 2, is composed by south-western Polish regions: Lubuskie, Wielkopolskie, Dolnośląskie, Opolskie and Śląskie.

Criteria for regions grouping are related to the course of medium-term business cycles fluctuations in the period under consideration. Such grouping requires *ex post* analysis of structural similarities among regions in groups and differences between groups. Table 2 contains basic characteristics of four groups.

Table 2. Basic characteristics of four groups of regions, 2001

	GDP per capita	Share in Polish GDP	Share of industry in Gross Value Added	Share of agriculture in Gross Value Added
Unemployment 1	117,8	41,7	21,9	3,6
Unemployment 2	104,8	35,4	28,7	3,4
Industry 1	81,0	11,2	24,6	2,5
Industry 2	74,5	11,8	21,0	6,4
Poland	100,0	100,0	24,5	3,8

Source: Authors' calculations based on Central Statistical Office data.

The names of four groups of regions may be misleading. The grouping made on the base of business cycles synchronization does not reflect a variety of structural features of the regions. Econometric and statistical procedures presented in the section 2 do not take into consideration structural characteristics of the regions. However, it is plausible to find *ex post* common features of regions identified as similar in the undertaken procedure.

The group Industry 1 is composed only by two regions in the south-eastern corner of Poland. These regions are less developped, low share of agiculture in production can be explained by unfavorable conditions in the prevailng mountainous areas. Share of industry close to the national average is due to the only large aglomeration (Cracow) in this region. Both regions have common historical roots. They belonged to the former Austro-Hungarian Galicia. Historical economic relations are probably strengthened by close geographical location.

Regions of the group Industry 2 are characterized by the lowest economic potential in Poland. Share of agriculture in value added is high. For this reason these regions are known as "Poland B". Consequently, share of industry is the lowest in this group of regions. This fact may explain why medium-term business cycle divergence appears: low share of industry leads to higher sensibility to business cycles impulses. Such finding seems to be supported by elimination of industrialization effect from business cycles.

It is difficult to find common economic characteristics for the regions in the group Unemployment 1. The largest and therefore the most influential on other Polish regions is Mazowieckie. It has strong economic ties with Łódzkie. Regions Zachodniopomorskie and Pomorskie have similar economic structures: maritime economy and tourism. Region Kujawsko-Pomorskie is a "bridge" between two former pairs of regions in the group Unemployment 1.

Regions of the fourth group Unemployment 2 are located in the south-western Poland. Share of indutry is the highest in the economy, share of agriculture below national average. All regions in this group have economic relations with German economy and apparently their business cycles are influenced nor only by ties than by signals from Polish regions but also by ties with German economy.

3.3. Patterns of dispersion of business cycles impulses in the identified groups of regions

At present stage of analysis, the research methodology applied in the paper did not allow for direct identification of channels of regional business cycles impulses in the four groups of regions. In other words, the pattern of similarities in regional business cycles can not be explained straightforward. However, the results of above described analysis and general knowledge makes it possible to formulate hypothesis on the most probable ways of dispersion of business cycles impulses within four distinguished groups of regions.

In the case of the group Industry 2 all four regions are similar: high dependence on agriculture and low economic potential. In addition, none region can be considered as a leading region in the group. Such pattern suggests that regions in this group are all "receivers" of the business cycles impulses from other regions and do not generate such signals themselves. In other words, regions in the group Industry 2 depend mainly on business cycles in Poland.

The smallest group of regions Industry 1 has a leader, which is region Małopolskie. Large agglomeration of Cracow is a reason why the region Małopolskie dominates over the region Podkarpackie as far as business conditions and economic impulses are concerned.

In the group Unemployment 1 with no doubt the leading region is Mazowieckie. This region influences economic situation of Łódzkie region due to the geographical location. Among three northern regions in this group the strongest region is Pomorskie which dominates over the other two regions.

In the group Unemployment 2 composed of five regions, three of them (Wielkopolskie, Dolnośląskie and Śląskie) have large agglomerations and therefore their economic potential is important. None of the three regions can be considered as a leading region which could be able to send business cycles impulses to other regions in this group.

4. Conclusions

The main result of the conducted research is identification of four groups of regions with similar business cycle fluctuations. Two of them are characterized by similar fluctuations of industry production levels and two other — with unemployment rate convergence. The regions in the identified groups are located in close geographical neighborhood. Such pattern of transmission of business cycles impulses seems to be typical for the present stage of transformation.

The analysis allows to draw a hypothesis that similarties in sensivity on business cycles in regions depend the most on economic structure as well as level of development of the regions. These characteristics mainly result from historical roots.

The research methodology applied in the paper did not allow for direct identification of channels of regional business cycles impulses in the four groups of regions. However, in three of four groups of regions, the analysis of some macroeconomic indicators allowed authors to indicate the hypothetical leader (leaders) of "macro-regions". It should be pointed out that ways of impulses dispersion described in the paper have to be treated only as a hypothetical ones.

The main obstacle for the more advanced researches is the lack of regional time series and more suitable economic data which could be used for analysis. Undoubtedly the problem is substantial and requires further analyses and researches.

Acknowledgement

Authors wish to thank Kamila Rydzewska for assistance in calculations and fruitful participation in discussions of the results.

REFERENCES

MICHAŁ DOMAŃSKI, WIKTORIA PARCHOMENKO, MARCIN PETERLIK, BOHDAN WYŻNIKIEWICZ (2001), Realisation of Entrepreneurs Expectations in Research with Business Condition Test, [in:] Business Tendencies on the Market of Financial Services in Poland, Akademia Ekonomiczna, Poznań.

R. J. HODRICK, E. C. PRESCOTT (1980), Postwar US Business Cycles: An Empirical Investigation, Working Paper, No. 451, Carnegie — Mellon University, Pittsburg; reprinted [in:] Journal of Money, Credit and Banking, Vol. 29, No. 1, February 1997, p. 1—16.

JACEK KOTŁOWSKI (2002), The Methods of Time-Series Smoothing with ARIMA Models, Papers and Proceedings No. 73, Research Institute of Economic Development, Warsaw.

BOHDAN WYŻNIKIEWICZ (ed.) (2004), Macroeconomic Situation and Forecasts for Polish Economy 2003—2006, GIME Quarterly Bulletin No. 41, January 2004.

Appendix

Figure A1. Unemployment — Deviations from trend in voivodships and Poland

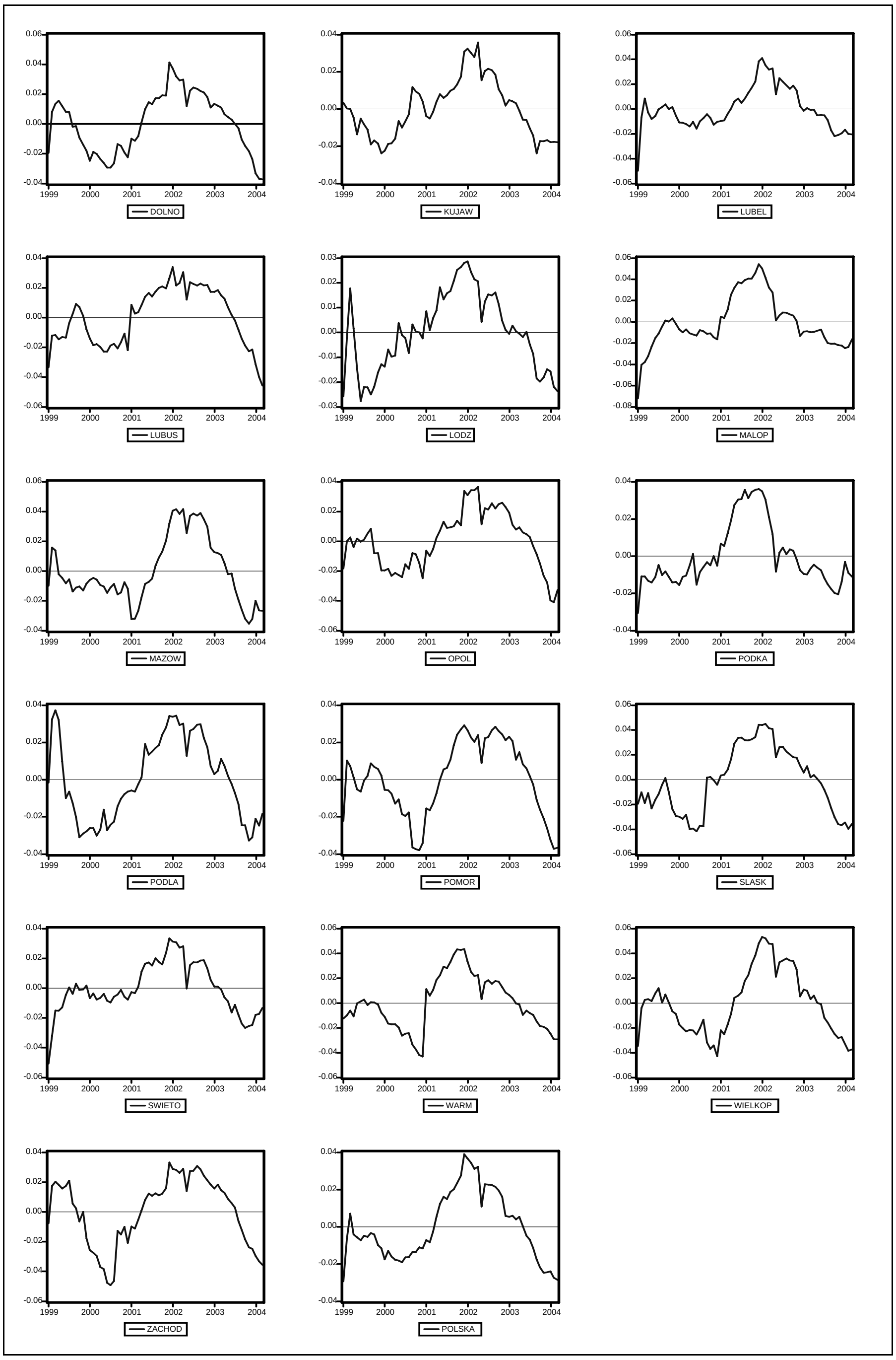


Figure A2— Deviations from trend in voivodships and Poland

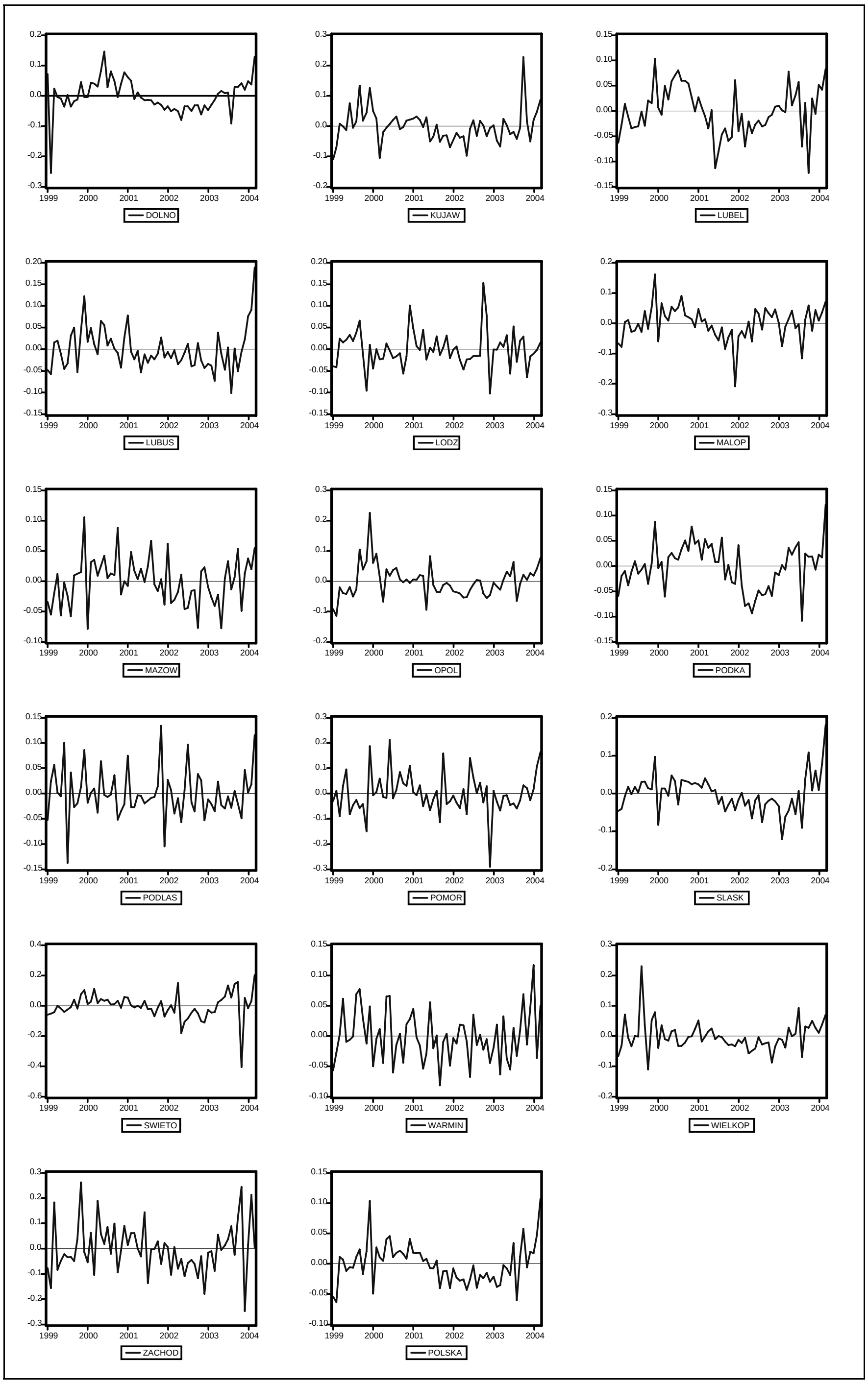


Table A1. Unemployment — Correlation coefficients

	Dolnośląskie	**Kujawsko - pomorskie**	**Lubelskie**	**Lubuskie**	**Łódzkie**	**Małopolskie**	**Mazowieckie**	**Opolskie**	**Podkarpackie**	**Podlaskie**	**Pomorskie**	**Śląskie**	**Świętokrzyskie**	**Warmińsko - mazurskie**	**Wielkopolskie**	**Zachodniopomorskie**	**Poland**
Dolnośląskie	1,000																
Kujawsko – pomor.	0,769	1,000															
Lubelskie	0,852	0,786	1,000														
Lubuskie	0,874	0,712	0,854	1,000													
Łódzkie	0,706	0,838	0,782	0,721	1,000												
Małopolskie	0,592	0,606	0,789	0,743	0,734	1,000											
Mazowieckie	0,810	0,796	0,831	0,730	0,675	0,454	1,000										
Opolskie	0,953	0,785	0,863	0,919	0,689	0,606	0,826	1,000									
Podkarpackie	0,589	0,681	0,710	0,649	0,814	0,916	0,429	0,550	1,000								
Podlaskie	0,888	0,835	0,755	0,708	0,799	0,472	0,802	0,815	0,606	1,000							
Pomorskie	0,870	0,559	0,792	0,873	0,589	0,546	0,818	0,873	0,446	0,710	1,000						
Śląskie	0,862	0,862	0,829	0,888	0,806	0,777	0,691	0,872	0,797	0,812	0,691	1,000					
Świętokrzyskie	0,701	0,726	0,904	0,816	0,764	0,908	0,681	0,743	0,817	0,596	0,660	0,814	1,000				
Warmińsko – mazur.	0,828	0,617	0,766	0,858	0,690	0,796	0,600	0,790	0,775	0,706	0,817	0,823	0,764	1,000			
Wielkopolskie	0,925	0,742	0,919	0,872	0,678	0,676	0,881	0,915	0,604	0,794	0,917	0,796	0,794	0,845	1,000		
Zachodniopomorskie	0,949	0,680	0,743	0,808	0,549	0,425	0,745	0,899	0,437	0,849	0,809	0,803	0,570	0,737	0,838	1,000	
Poland	0,932	0,849	0,949	0,921	0,844	0,787	0,847	0,924	0,765	0,859	0,849	0,926	0,878	0,869	0,940	0,833	1,000

Table A2. Industry — Correlation coefficients

	Dolnośląskie	**Kujawsko - pomorskie**	**Lubelskie**	**Lubuskie**	**Łódzkie**	**Małopolskie**	**Mazowieckie**	**Opolskie**	**Podkarpackie**	**Podlaskie**	**Pomorskie**	**Śląskie**	**Świętokrzyskie**	**Warmińsko - mazurskie**	**Wielkopolskie**	**Zachodniopomorskie**	**Poland**
Dolnośląskie	1,000																
Kujawsko – pomor.	0,314	1,000															
Lubelskie	0,439	0,134	1,000														
Lubuskie	0,547	0,395	0,532	1,000													
Łódzkie	0,088	0,221	-0,094	0,222	1,000												
Małopolskie	0,469	0,568	0,396	0,539	0,124	1,000											
Mazowieckie	0,341	0,297	0,232	0,329	-0,066	0,410	1,000										
Opolskie	0,428	0,539	0,397	0,617	0,084	0,546	0,399	1,000									
Podkarpackie	0,520	0,486	0,461	0,482	0,149	0,403	0,442	0,487	1,000								
Podlaskie	0,092	0,246	0,170	0,534	0,297	0,383	0,146	0,253	0,236	1,000							
Pomorskie	0,291	0,226	0,267	0,423	0,206	0,306	0,180	0,253	0,274	0,256	1,000						
Śląskie	0,529	0,604	0,293	0,639	0,225	0,624	0,515	0,485	0,581	0,389	0,410	1,000					
Świętokrzyskie	0,307	0,316	0,183	0,349	0,188	0,351	0,390	0,391	0,334	0,279	0,211	0,444	1,000				
Warmińsko – mazur.	0,274	0,388	0,082	0,484	0,298	0,332	0,293	0,358	0,215	0,336	0,263	0,487	0,310	1,000			
Wielkopolskie	0,302	0,396	0,267	0,535	0,203	0,296	0,147	0,370	0,495	0,306	0,103	0,481	0,233	0,418	1,000		
Zachodniopomorskie	0,372	0,303	0,145	0,190	-0,043	0,041	0,131	0,268	0,289	-0,042	0,057	0,172	0,003	0,058	0,240	1,000	
Poland	0,656	0,646	0,461	0,749	0,261	0,704	0,639	0,658	0,706	0,402	0,479	0,886	0,518	0,530	0,591	0,316	1,000

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 309—323

THE USE OF BANKING BTS DATA IN DIAGNOSING AND FORECASTING SITUATION OF INDUSTRY AND CONSTRUCTION IN POLAND

Józef Garczarczyk[1] **Robert Skikiewicz**[2]

ABSTRACT

The aim of the paper is to construct econometric models describing relationships between the economic situation in banking sector (qualitative data) and macroeconomic indicators showing economic situation in other sectors of the Polish economy (quantitative data from the Central Statistical Office).

Qualitative data on the economic situation in Polish banking sector used in analyses has been obtained from business test surveys (BTS) conducted by the Department of Marketing Research of the University of Economics in Poznań. The survey, which is carried on quarterly in Polish banks since 1992, is one of the earliest applications of the BTS in the financial market.

Analyses will be conducted with reference both to the whole banking sector (synthetic indicator Poznań Index of Banking Business Conditions) and to the particular groups of bank services (deposits, credits). Econometric analyses will be preceded by cross-correlation analysis and Granger causality tests.

1. Introduction

With the strengthening of the market economy in Poland — after the change of the economic system in 1989 — the increase in demand for information concerning the assessment of the economic situation both on the national scale and with respect to particular branches, the estimation of their attractiveness level and their development abilities, has been noticed. Simultaneously, the increase in the interest in short-term forecasts of the basic economic categories, which are the basis of the decision-making for the participants of economic processes, is clearly

[1] Department of Marketing Research, Poznan University of Economics,
e-mail: j.garczarczyk@ae.poznan.pl
[2] Department of Marketing Research, Poznan University of Economics,
e-mail: r.skikiewcz@ae.poznan.pl

visible. Stabilisation of the market economy in Poland, its opening as a result of joining the EU structures and the liberalisation of economic activities contribute to the fact that in the managers' decision-making process the importance of macroeconomic analyses and the forecasts of those numbers, achieved also by unconventional means, is growing.

One of the sources providing the abovementioned information, apart from the data from market statistics, are the results of the business tendency surveys. Such research has a long tradition in fully-grown, mature market economies and the results are commonly used by the participants of the economic life for the assessment and forecasting of the directions of changes in economic activities on a national or regional scale or with respect to particular branches. In Poland, the interest in the business tendency surveys on a wider scale has appeared only after the change of the economic system.

One of the most popular methods of examining the changes in economic activities is the business tendency survey. This method has many advantages[1], and the most important of these include its speed and relative ease of achieving information and the possibility of achieving not only the assessment of the accomplished economic processes (ex post), but also their forecast for the coming period (ex ante). An indisputable advantage of this method is its universality, enabling the researchers to diagnose and forecast the business tendencies both on the level of the whole economy as well as in particular branches and sectors. Despite the widespread use of the business tendency surveys to study the situation in industry, construction and foreign trade, this method has been relatively rarely used by research institutes to assess and forecast business performance tendencies in the service sector, and especially in the financial service sector.

One of the first centres in Europe to use the business tendency survey for the assessment of the situation on the financial service market was the Department of Marketing Research of the University of Economics in Poznań (DMRP)[2]. Such research has been conducted in the banking (since 1992) and insurance (since 1993) sectors, in a quarterly cycle on a Polish, nation-wide constant sample consisting of 500 units in both cases. The panel comprises of both the head-offices and branches of banks and insurance companies. Information from the units par-

---

[1] A more extensive characteristics of the advantages and efficiency of the business test survey can be found, among others, in: E. Adamowicz, Badania koniunktury gospodarczej i ich skuteczność (Business conditions research and its effectiveness), in: (ed.) J. Garczarczyk, Rynek usług finansowych w Polsce w latach 1990—2001. Diagnozowanie i prognozowanie koniunktury (Financial services market in Poland 1990—2001. Diagnosing and forecasting economic situation), Akademia Ekonomiczna w Poznaniu, Poznań 2003; M. Bieć, Test koniunktury. Metody, techniki, doświadczenia, SGH, Warszawa 1995.

[2] A detailed description of the methodology and research organisation can be found in: J. Garczarczyk, R. Matusewicz, M. Mocek, Koniunktura na rynku bankowym i ubezpieczeniowym w Polsce (Business conditions in the Polish banking and insurance sector), Akademia Ekonomi-czna w Poznaniu, Poznań 2001.

ticipating in the survey is received by a mail survey and the achieved return is on the level of 40—50%.

The need of research, diagnose and forecasting of the business tendency fluctuations in the financial sector stems from the fact that this sector is the "pacemaker of the economy's bloodstream", which makes the information about the business performance tendencies indispensable for the state's decision-making centre and for the all economic subjects. Moreover, in Poland, as a result of the systemic transformation, the rank of the financial sector has risen decisively and banks have found a proper place for their function.

Considering the abovementioned premises, the authors have tried to answer the following questions:

- Do the results of the business tendency survey in the banking sector (banking BTS) reflect the dynamics of changes in other sectors of Polish economy?
- Can the qualitative data from the surveys in the banking sector (banking BTS) be used in econometric models describing the activity changes in the particular sectors of Polish economy?
- Can such models serve as the basis of creating the economic development forecasts, what is the accuracy of such forecasts and the reasons of their errors?

However, in this article the issue of evaluating the applicability of the banking BTS data for diagnosing and forecasting the situation in the banking sector has been omitted, as this problem was the subject of analyses in previous publications[1].

2. Research Methodology

The basis for the analyses and comparison with qualitative data from the banking BTS (DMRP) comprised of the quantitative (statistical) data collected and published by the Central Statistical Office (CSO). For the analyses the following variables characterising the dynamics of the quarterly changes in Poland in 1995—2003 were used:

- PPI – the sold industry production value,
- PBI – the construction and assembly production value,

[1] The description of the results of this research is to be found in: J. Garczarczyk, R. Matusewicz, Economic Performance in the Polish Banking Sector — Accuracy Evaluation of the Results Obtained from the Business Tendency Survey, Selected Papers submitted to the 26th Ciret Conference, Taipei, pp. 20—24, R. Matusewicz, Zastosowanie testu koniunktury w ilościowych prognozach na rynku usług bankowych w Polsce (The use of business test survey in quantitative forecasts of the market of banking services in Poland), w: Badanie gospodarki polskiej — stan bieżący i perspektywy (Researches of Polish economy — actual state and future) , Instytut Rozwoju Gospodarczego SGH, Warszawa 2002, pp. 199—208.

Statistical data were present in the analyses as referential variables. In order to provide the comparison with the qualitative data, the quantitative data have been transformed into indexes, with the use of two formulas:

- Y – analogical quarter of the previous year = 100,
- Q – the previous quarter = 100

(the symbols Y, Q are present as the last letter in the variable description).

Among many qualitative variables achieved during the survey (BTS) characterising the changes in the banking sector, the following were subject to comparative studies[1]:

- OSF – general assessment of the banks' financial situation,
- OPS – obtained means value,
- OWK – given loans value,
- PIK – the composite indicator (DMRP banking indicator) — Poznań Index of Banking Economic Performance — PIKBANK, characterising the economic situation in the banking sector.

The synthetic index PIKBANK is calculated as the weighted average of five BTS balances according to the following formula:

$$PIKBANK = \frac{OPS + OWK + (-OKN) + OSF + PSB}{5}$$

- OPS – balance of obtained financial means in t period (assessment),
- OWK – balance of given loans in t period (assessment),
- OKN – balance of doubtful and loss loans in t period (assessment),
- OSF – balance of financial situation in t period (assessment),
- PSB – balance of total assets in t+1 period (prediction).

PIKBANK indicator is estimated for the whole banking sector and for the bank segments on the basis of the ownership criterion and localisation criterion[2] .

The qualitative variables used in the analyses are balances of the assessment of the studied phenomena, formulated ex post. They were analysed as modified BTS balances.

While comparing the adjustment of qualitative and quantitative variables, the leads and lags of the qualitative variables from one to eight quarters has been studied. In the analyses presented in the text the following methods have been employed in consecutive steps:

- cross-correlation analysis,
- Granger causality test,

---

[1] A more extensive description of the banking tendency surveys can be found in: J. Garczarczyk, Analiza wahań koniunktury na rynku usług bankowych w latach 1992—2002 (The analysis of business conditions in the Polish banking sector 1992—2002), in: Badanie gospodarki polskiej — stan bieżący i perspektywy (Researches of Polish economy – actual state and future), Instytut Rozwoju Gospodarczego SGH, Warszawa 2002, pp. 17—32.

[2] J. Garczarczyk, R. Matusewicz, Economic Performance in the Polish Banking Sector … op. cit., pp. 4—5 and J. Garczarczyk, M. Mocek, R. Matusewicz, Konunktura ... op. cit., p. 68.

- econometric modelling.

In the correlation analysis, studying the interdependence between the quantitative (CSO) and qualitative variables (DMRP), Pearson's coefficient of correlation has been used. It allowed for establishing the strength of the interdependence relations between these variables depending on the lead or lag period. Employing Granger causality test aimed at preliminary verification of whether there exist causal relationships between the variables from the banking BTS and statistical data and what their direction is. On the basis of the abovementioned analyses, an attempt has been made to construct econometric models aiming at describing the quantitative changes of referential variables (sold industry production, and construction and assembly production) with the use of qualitative variables (banking BTS).

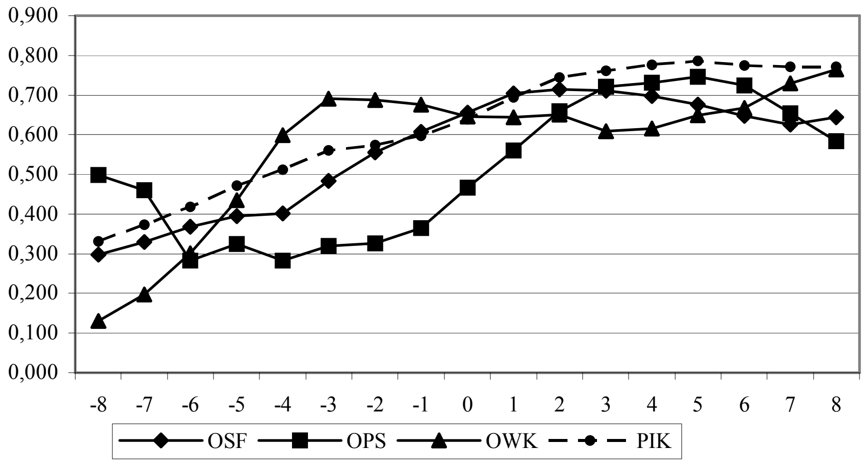
3. Correlation Analysis

The aim of this part of analyses is a diagnosis of the existence of a relation between the business fluctuations on the banking market and fluctuations in chosen economy sectors. Consequently, an analysis of the interdependence between banking BTS and statistical data indexes, characterising the changes in sold industry production and construction and assembly production has been conducted. Eight lead and lag periods have been used.

Correlation coefficients between the sold industry production (PPIY) and the banking BTS balances achieved maximum values from 0.715 to 0.786, while it is worth noticing that such values have been observed for the periods t+2, t+5 and t+8, which means that the strongest relation was present between the current economic situation in industry and the future banking BTS balances. It indicates an anticipatory character of the changes in the industry sector in relation to the banking sector. In turn, the relations between quarterly sold industry production index (PPIQ) and banking BTS balances proved to be noticeably weaker, which is confirmed by the maximum correlation coefficient of only 0.352.

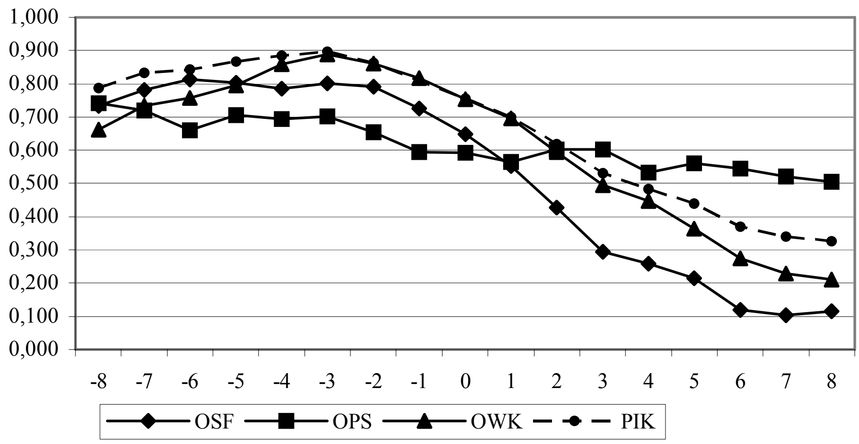
The yearly construction and assemby production index (PBIY) shows even a slightly stronger relation to banking BTS balances. It is also symptomatic that the highest correlation coefficient values have been observed this time for the banking BTS balances from the periods earlier than t, which shows that the changes in the construction operate as a consequence of the banking sector changes. The strongest correlation has appeared between PBIY and the OWK (0.889) and PIK (0.897) variables from the t-3 period. It means that the changes in banking BTS balances show the strongest relation to the construction economic situation changes operating 3 quarters later. In the case of construction and assembly production, similarly to the case of industry production, the relations between the quarterly index (PBIQ) and banking BTS balances are noticeably weaker, which is confirmed by the maximum value of Pearson's correlation coefficient of only 0.196.

Figure 1. Correlation coefficients between the sold industry production index (PPIY) and banking BTS balances



Source: own calculations

Figure 2. Correlation coefficients between the construction and assembly production index (PBIY) and banking BTS balances



Source: own calculations

4. Granger causality analysis

In order to verify the conclusions introduced earlier and based on the correlation analysis regarding the relation between banking BTS balances and statistical data, a causal relationships analysis has been conducted, based on Granger causality test. In the case of this test, the number of lags introduced into the econometric models on which the test is is based is decisive for the indication of a variable influencing the shaping of the second variable. Therefore, conducting the test for various sets of lags is advisable, as it can confirm or question conceivable influence of one variable on the other[1]. In this text, a study of causal relationships has been conducted solely in order to establish preliminarily the legitimacy of constructing ADL models based on the variables from the banking survey tests. However, the indications stemming from this test should be treated only as preliminary as a result of short time series on which the analysis has been based and the nonstationarity of the variables[2].

Table 2 presents the results of the tests conducted in order to discover the variables influencing the sold industry production. From the set of variables coming from the banking BTS only the balance of the bank's financial situation does not influence the sold industry production. In turn, the given loans balance cause the yearly sold industry production index for the lag set of 2, 3 and 8 periods and for the lags from 4 to 7 a mutual causal relationships can be found between the variables. It is interesting that there appear similar in character causal relationships for particular lags between the quarterly sold industry production index and the given loans balance. This time, however, the mutual causality of the aforementioned variables has been observed for the lags of 3 to 7 periods. For the lags amounting to 8 periods, a one-sided causality of the change of the values of given loans on the quarterly sold industry production has been noted. It seems, therefore, justified to draw a conclusion about the influence of the value of given loans on the level of industry production.

[1] J. Shan, F. Sun, Domestic saving and foreign investment in Australia: a Granger causality test, in: International Economic Journal, Volume 12, Number 4, Winter 1998, pp. 83—84.

[2] For more see: W. Charemza, D. F. Deadman, Nowa ekonometria (New econometrics), PWE, Warszawa 1997,pp. 161—162

Table 1. Causal relationships between the indexes of the changes in the sold industry production (PPI) in the light of Granger causality test

PPI	BTS	PPI ↔ BTS	PPI → BTS	BTS → PPI
PPIY	OSF	–	–	–
PPIY	OPS	–	–	–
PPIY	OWK	+	–	+
PPIY	PIK	–	–	+
PPIQ	OSF	–	–	–
PPIQ	OPS	–	–	+
PPIQ	OWK	+	–	+
PPIQ	PIK	–	+	–

„+” means there is a relationship between variables
„–” means there is no relationship between variables
Source: own calculations

In turn, the PIKBANK indicator in Granger's sense cause the sold industry production (PPIY) for the lag of 4. For the same lag, the quarterly sold industry production index causes in Granger's sense PIKBANK, so there exists a causal relationship between the variables in opposite direction. The value of the means obtained, in turn, causes in Granger's sense the quarterly index of sold industry production for the lag of 7 periods.

Table 2. Causal relationships between the indexes of construction and assembly production changes (PBI) and the banking BTS balances in the light of Granger causality test

PBI	BTS	PBI↔ BTS	PBI→ BTS	BTS → PBI
PBIY	OSF	–	–	+
PBIY	OPS	+	+	+
PBIY	OWK	–	–	+
PBIY	PIK	–	–	+
PBIQ	OSF	+	–	–
PBIQ	OPS	–	–	+
PBIQ	OWK	+	+	–
PBIQ	PIK	+	+	–

„+” means there is a relationship between variables
„–” means there is no relationship between variables
Source: own calculations

Even a greater number of causal relationships can be found while analysing the results of the Granger causality test between construction and assembly production and banking BTS balances. In this case a decisive majority of causal relationships appears for slight lags, usually not exceeding three periods. Each variable coming from the banking BTS causes the yearly construction and assembly production index. Additionally, the value of the means obtained both causes and is caused by construction and assembly production for the lags of 3 to 7 periods. Moreover, for the lag of 8 periods, construction and assembly production in Granger's sense causes the obtained means value.

A slightly different character of causal relationships has been observed between the quarterly construction and assembly production index and banking BTS balances. Only the value of the means obtained caused in Granger's sense construction and assembly production. The bank's financial situation simultaneously both caused and was caused by construction and assembly production for the lag of 3 periods. Two other categories coming from the banking BTS, i.e. the value of the loans given and the synthetic PIKBANK indicator for the number of lags of 3 periods showed mutual causal relationships with construction and assembly production and for a higher number of lags they were caused by construction and assembly production.

5. Econometric models

In order to construct econometric models it has been decided to use autoregressive models with distributed lags (ADL), in which the delayed quantitative variables (statistical data) and delayed qualitative changes from the banking BTS have been used as explanatory variables. For the models the approach from the general to the particular has been employed, assuming 8 periods as the primary number of lags[1] and a lower number of lags if the earlier Granger causality test indicated for such a number of lags the existence of causal relationships between the variables considered in the model. For the estimated models a Breusch-Godfrey serial correlation LM test has been conducted in order to verify whether there exists a first order serial correlation of the residuals. On the basis of the models, characterised with the adjustment of empirical data to theoretical data surpassing 50% and without serial correlation in the residuals, the forecasts for the 4 following quarters of the year 2003 have been made. Later, the accuracy of the forecasts has been verified[2]. In further stages of the procedure only those models have been taken into consideration for which the mean absolute percentage error of the forecast was not higher than 5%[3]. For the models chosen on the basis of the aforementioned criteria, the causes of their errors have been estab-

[1] W. Charemza, D. F. Deadman, op. cit., pp. 75—77.

[2] In the forecasts verification the following measures have been employed: root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

[3] Forecasts for which the MAPE is between 3% and 5% are regarded as good and those for which the MAPE is lower than 3% are regarded as very good. For more, see: A. Zeliaś, B. Pawełek, S. Wanat, Prognozowanie ekonomiczne (Economic forecasting), PWN, Warszawa 2003, pp.49—50.

lished with the use of Theil's coefficient components[1]. Additionally, the Janus quotient has been established in order to state whether the model can be used for making the forecast for another period.

Table 3. Chosen models for the sold industry production (PPI) and construction and assembly production (PBI)

Model A	Variables	Coefficient	Std. Error	t-Statistic	Prob.
PPIY	C	96,992	6,450	15,038	0,000
	PPIY(-4)	-0,371	0,079	-4,711	0,000
	OWK(-2)	0,243	0,063	3,862	0,001
	OWK(-3)	0,251	0,072	3,491	0,002
	OWK(-6)	-0,113	0,048	-2,368	0,028
	R-squared	Adjusted R-squared	S.E. of regression	Breusch-Godfrey Serial Correlation LM Test (Probability)	
	0,913	0,897	2,583	0,526	
Model I	Variables	Coefficient	Std. Error	t-Statistic	Prob.
PBIY	OWK(-1)	0,299	0,141	2,121	0,043
	OWK(-3)	0,455	0,139	3,277	0,003
	R-squared	Adjusted R-squared	S.E. of regression	Breusch-Godfrey Serial Correlation LM Test (Probability)	
	0,789	0,781	8,482	0,101	

Source: own calculations

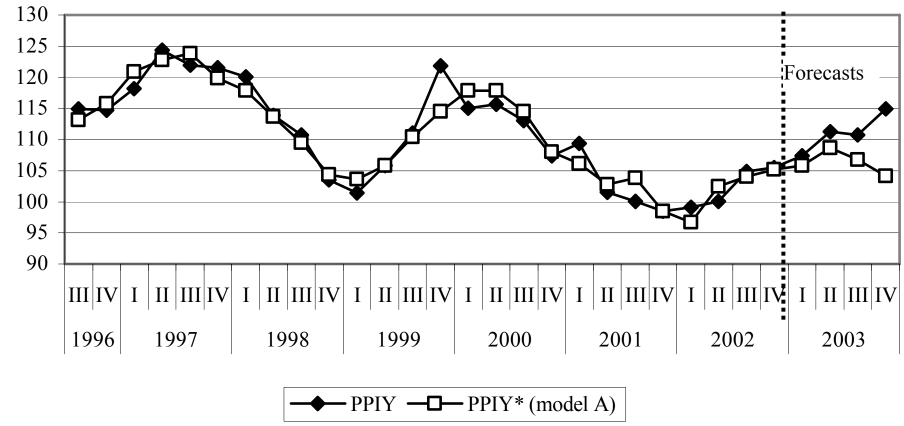
Various forms of econometric models have been tested, in which as the explanatory variables the indexes of referential variables (PPI, PBI) were employed and as the explanatory variables the qualitative variables groups and delayed explained variables.

In table 3, the characteristics of the models with the best adjustment to empirical data have been shown, one in each of the categories of statistical data, i.e. for the sold industry production and construction and assembly production. Simultaneously, they also meet the demands in the range of the mean absolute percentage error. Among the models presented, the A model is the best adjusted to the actual data for the sold industry production. The fact that in every model only one variable from the banking BTS appears — the given loans balance — is worth emphasising. It also seems important that in the case of construction and

[1] Three Theil's coefficient components may be distinguished. The first of these, I1 informs us about the extent in which the forecast error was caused by the improper anticipation of the phenomenon's average level. The second coefficient, I2 states how much the predictions elasticity was adjusted to the actual fluctuations of the prognosed variable. The third Theil's I3 coefficient allows us to state to which extent the forecast error was influenced by the lack of agreement of the prognosis change direction with the actual direction changes of the prognosed variable. For more, see: A. Zeliaś, Teoria prognozy (Theory of forecasting), PWE, Warszawa 1997, pp. 230—233.

assembly production the only explanatory variables in the model are the qualitative data from the banking BTS, which turned out to be sufficient for the explanation of the shaping of these economic phenomena.

Figure 3. Sold industry production — statistical data and model A data



Source: own calculations

Figure 3 presents the actual values of the sold industry production and the model ones (II 1996—IV 2002) and the forecasts and their realisation (I 2003—IV 2003). Unfortunately, despite the small differences between the theoretical and empirical values in the period for which the model's parameters have been estimated, the forecast for the fourth quarter of 2003 is significantly different from the actual value of the phenomenon. Consequently, good adjustment of the presented model A, expressed by the R squared of 0. 913 allowed for accurate forecasts only 3 quarters in advance. It means that the relations between the variables considered in the model have not remained up-to-date as early as in the next, fourth quarter.

In the following part of the article, for all models meeting the adjustment and the mean absolute error percentage criteria, the causes of errors in forecasts created on their basis will be presented, together with the verification of their usefulness in forecasting for the consecutive period.

As much as eight models characterising the shaping of the sold industry production have met the total adjustment to empirical data and forecast accuracy criteria. Five of them allowed for the forecasting of the yearly sold industry production changes and three of them of the quarterly changes. On the side of explanatory variables in the seven models there were the given loans balances and only in one case the means obtained balances.

Table 7. The ex post errors of forecasts and the reasons of their appearance for the ADL models — the sold industry production (PPI)

	R-squared	RMSE	MAE	MAPE	I	I_1	I_2	I_3	Janus quotient
Model A	0,91	5,95	4,73	4,19	0,03	63%	3%	34%	6,58
Model B	0,91	6,12	5,58	4,97	0,03	83%	11%	6%	7,37
Model C	0,90	6,81	5,38	4,77	0,03	62%	0%	38%	7,46
Model D	0,88	6,21	5,24	4,65	0,03	71%	5%	25%	2,90
Model E	0,91	5,36	4,1	3,62	0,02	47%	6%	47%	3,39
Model F	0,83	3,58	2,79	2,63	0,02	42%	23%	35%	3,62
Model G	0,87	3,58	2,86	2,7	0,02	34%	17%	49%	3,19
Model H	0,54	4,01	3,15	3	0,02	44%	5%	51%	2,85
Model I	0,79	6,01	4,14	4,96	0,03	10%	87%	3%	0,54

A: PPIY= C+PPIY(-4)+OWK(-2)+OWK(-3)+OWK(-6)

B: PPIY= C+PPIY(-3)+PPIY(-6)+OWK(-1)+OWK(-3)

C: PPIY= C+PPIY(-4)+OWK(-2)+OWK(-3)+OWK(-5)

D: PPIY= C+PPIY(-1)+PPIY(-4)+OWK(-1)

E: PPIY= C+PPIY(-1)+PPIY(-4)+PPIY(-6)+OPS(-6)+OPS(-7)

F: PPIQ= C+PPIQ(-1)+PPIQ(-2)+PPIQ(-3)+OWK(-2)+OWK(-3)+OWK(-5)

G: PPIQ= C+PPIQ(-1)+PPIQ(-2)+PPIQ(-3)+PPIQ(-6)+OWK(-1)+OWK(-3)+OWK(-5)

H: PPIQ= C+PPIQ(-1)+PPIQ(-2)+PPIQ(-3)+OWK(-1)

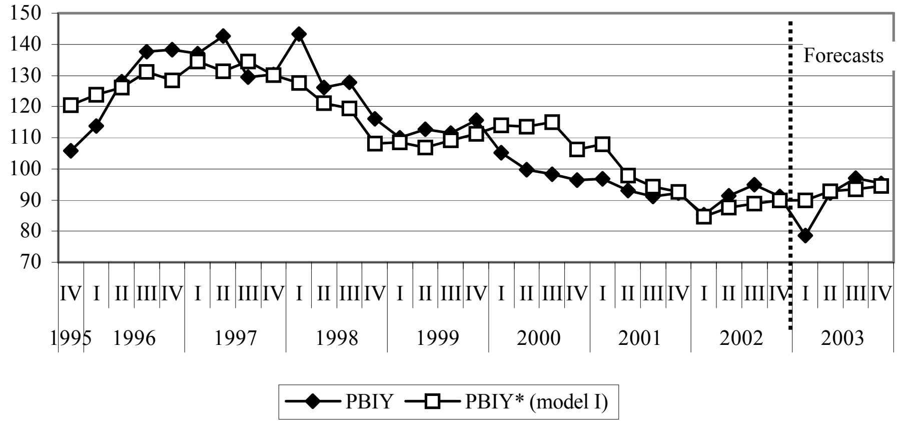
I: PBIY= OWK(-1)+OWK(-3)

Source: own calculations

In the case of five models characterising the shaping of the sold industry production, the main cause of the forecasts' errors was the improper forecasting of the phenomenon's average level. At the same time, what is indicated by the I_2 coefficient, it stemmed only in a slight degree from the differences in the scope of changeability of the forecasts created on the basis of the model and changeability of the phenomenon in the period for which the forecast were created. Only in the case of the B model, did the error stem from the improper anticipation of the phenomenon's direction change, and to an insignificant degree. However, in the case of two models (G and H), the insufficient congruity of the direction of actual changes with the anticipated values constituted the fundamental reason of the presented forecast errors.

Unfortunately, despite low values of the mean absolute percentage error of the ex post forecast, the Janus quotient indicates that the models shown in Table 8 cannot be used for forecasts for the consecutive quarter. It results from the fact that the forecasts created for four quarters in advance were characterised with various accuracy for particular quarters. Usually, for the first and second quarter of 2003, the mean absolute percentage error of the forecasts obtained did not exceed 2% but for the third and fourth quarter the errors were, unfortunately, higher. Therefore, the more distant the forecast period was, the higher the error during the forecasting of the phenomenon on the basis of the model presented. It may, therefore, be stated that the relations between variables in ADL models characterising the shaping of the sold industry production do not retain stability for too long in the future periods and, in consequence, allow for the forecasts to be created only for one or two periods in advance.

Figure 4. Construction and assembly production (PBIY)— statistical data and model I data



Source: own calculations

Among the estimated models describing the shaping of construction and assembly production, only the model considering on the side of delayed variables the given loans balances lagged by one and three periods, was at the same time characterised by the adjustment exceeding 50% and allowed for achieving for the four 2003 quarters forecasts with the mean absolute percentage error of less than 5%. The divergence between the actual phenomenon level and the forecasts created resulted primarily (in as much as 87%) from different changeability of the forecasts and the phenomenon's actual values. This model allowed for the creation of forecasts for four quarters in advance, in the case of which only in 3% did the errors result from the divergence between the forecast values change direction and the actual values. It is also immensely important that this model furtherly retains its up-to-dateness and allows for the forecast for the following, fifth quarter.

6. Conclusions

The analyses conducted with the use of various methods proved that the qualitative data from the banking BTS describe well the changeability of the activity process in other sectors of economy and allow for the formulation of short-term quantitative forecasts of their further course. This general statement allows for drawing subsequent, more detailed conclusions:

1. The qualitative data collected in the banking BTS are a valuable source of information not only about the performance in the banking sector but also in other sectors, as well as about the whole economy. They allow for establishing trends and directions of changes in the economic activity in the most important economy sectors.
2. Correlation analysis proved that the qualitative data concerning banking economic situation (banking BTS) reflect well the dynamics of quantitative changes in the selected economy sectors (industry, construction). A significantly higher convergence of the qualitative data changes (banking BTS) was obtained with the yearly rather than the quarterly quantitative data index.
3. From the analyses conducted it ensues that the fluctuations in the banking sector (in the light of the data from the business tendency survey) are delayed in relation to the fluctuations in the industry. In the case of the construction they are anticipatory.
4. The estimated econometric models with the qualitative data as the explanatory variables describe well the cyclic course of the changes in the economy sectors under analysis and also project the turning points. They may, therefore, be useful for diagnosing and forecasting economic situation and, at the same time, for conducting current economic policy.
5. Among the four qualitative variables describing situation in the banking sector, the highest usefulness in econometric modelling is shown by the given loans balances. The dim prognostic value of the PIKBANK index, characterising the general situation in the banking sector, is surprising.
6. The main reason for the forecasts errors was incorrect anticipation of the average level of a phenomenon, while the prediction's flexibility was usually adjusted to actual fluctuations of the variable being forecasted.

The results obtained are of diagnostic character so they require further improvement and testing. It is, therefore, necessary to continue the presented research direction, to collect even longer qualitative and quantitative data series, to test other variants of regression models and to expand the research into other economy sectors. However, it may be already stated that the qualitative data on the banking situation (banking BTS) may be a source of information about the changes in other sectors of economy.

REFERENCES

ADAMOWICZ, E (2003), Badania koniunktury gospodarczej i ich skuteczność (Business conditions research and its effectiveness) w: red. J. Garczarczyk, Rynek usług finansowych w Polsce w latach 1990—2001. Diagnozowanie i prognozowanie koniunktury (Financial services market in Poland 1990—2001. Diagnosing and forecasting economic situation), Akademia Ekonomiczna w Poznaniu.

BIEĆ, M (1995), Test koniunktury. Metody, techniki, doświadczenia (Business test survey. Methods, techniques and experiences), SGH.

CHAREMZA, W., DEADMAN, D,F. (1997), Nowa ekonometria (New econometrics), PWE.

CIEŚLAK, M (2001), Prognozowanie gospodarcze. Metody i zastosowanie, gospodarcze (Economic forecasting. Methods and applications), PWN.

GARCZARCZYK, J (2002), Analiza wahań koniunktury na rynku usług bankowych w latach 1992—2002 (The analysis of business conditions in the Polish banking sector 1992—2002), w: Badanie gospodarki polskiej — stan bieżący i perspektywy (Researches of Polish economy — actual state and future), Instytut Rozwoju Gospodarczego AGH.

GARCZARCZYK, J., MOCEK, M., MATUSEWICZ, R (2001), Koniunktura na rynku bankowym i ubezpieczeniowym w Polsce (Business conditions in the Polish banking and insurance sector), Akademia Ekonomiczna w Poznaniu.

GARCZARCZYK, J., MATUSEWICZ, R (2002), Economic Performance in the Polish Banking Sector — Accuracy Evaluation of the Results Obtained from the Business Tendency Survey, Selected Papers submitted to the 26th Ciret Conference, Taipei.

MATUSEWICZ, R (2002), Zastosowanie testu koniunktury w ilościowych prognozach na rynku usług bankowych w Polsce (The use of business test survey in quantitative forecasts of the market of banking services in Poland), w: Badanie gospodarki polskiej — stan bieżący i perspektywy (Researches of Polish economy — actual state and future), Instytut Rozwoju Gospodarczego SGH.

SHAN, J., SUN, F (1998), Domestic saving and foreign investment in Australia: a Granger causality test, International Economic Journal.

ZELIAŚ, A (1997), Teoria prognozy (Theory of forecasting), PWE.

ZELIAŚ, A., PAWEŁEK, B., WANAT, S (2003), Prognozowanie ekonomiczne (Economic forecasting), PWN.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 325—363

MEASURING THE USEFULNESS OF CONSUMERS' INFLATION EXPECTATIONS IN FINLAND

Pertti Kangassalo[1] **and Kari Takala**[2]

ABSTRACT

Since October 1995 in a Consumer Survey by Statistics Finland, the respondents have been asked monthly to give both their assessment of current inflation and expectations of future inflation in percentage points. We find out that expectations do have predictive power for inflation, but the forecast horizon of this information does not carry 12 months ahead. Inflation expectations are an unbiased predictor for the median and core CPI, but not for the ordinary CPI. Evidence was found that consumers react excessively to commodity groups that are familiar to them, such as food, housing and restaurant services. This phenomenon could also explain the recent divergence between consumers' inflation assessments and expectations after the changeover to euro. Apart from the clear evidence in favour of the usefulness of the mean of expectations, we find weak evidence that volatility (standard deviation) and skewness could be helpful in making more precise forecasts.

Key words: inflation expectations, monetary policy, surveys, forecasting
JEL Classification: D12, D84, E31, E52

1. Introduction

In this paper we consider the accuracy of consumers' assessments and expectations about the past inflation and about future inflation. The data from the Finnish Consumer Survey are used as indicators of inflation assessments, i.e. valuations concerning the past inflation, and of inflation expectations. These data are available monthly since October 1995. In the Consumer Survey consumers state the inflation rate as a direct percentage change from the previous year's corresponding month. By analysing the inflation assessments we can first evaluate how precisely consumers know the current inflation situation. The actual inflation

[1] Statistics Finland, P.O. Box 6A, FI-00022 Statistics Finland, Helsinki; pertti.kangassalo@stat.fi.
[2] The Bank of Finland, P.O. Box 160, FI-00101 Helsinki; kari.takala@bof.fi.

is commonly measured by means of Statistics Finland's Consumer Price Index (CPI), so in principle the estimates measure the annual change in consumer prices, which is widely in the public eye, too. However, we also study whether some other measure of inflation, such as the Harmonised CPI (HICP), core inflation or median CPI would give a closer relationship with the reported inflation estimates.[1]

First we study how well the inflation assessment estimates match the actual inflation. In this context we can also examine the sensitivity of the assessments with respect to the inflation rates of different consumption expenditure groups. Compared with the real consumption weights, consumers may adopt a divergent attitude towards the categorical inflation rates, i.e. familiar or frequently bought consumption items, such as foodstuffs or petrol, may be over-represented in consumers' thoughts. This phenomenon may partly explain why inflation assessments have diverted from the measured inflation in a number of the eurozone countries from the beginning of 2002. The changeover to euro raised the perceived inflation in many eurocountries, including Finland, while at the same time expectations concerning the rate of inflation in 12 months' time declined significantly, which needs an explanation.

As a second step we study whether consumers can forecast the future inflation with any accuracy. This is not an easy task, since inflation is widely known to be close to a random walk process, i.e. probably containing a unit root, which implies that changes in inflation should be unpredictable. If inflation is to any extent predictable, we are interested in how inflation expectations are formed. The accuracy of inflation forecasts is analysed by means of weak rationality regression tests, testing for bias and (linear) efficiency.

Finally, we will also carry out an analysis concerning the distribution moments of inflation expectations by using individual survey interview data. It is interesting to try to find out whether the uncertainty of inflation expectations has somehow changed over time and whether this has had any interesting consequences. We calculated the distribution parameters of the cross-sectional survey data samples and examined their variation and the predictive power of different distributional moments of inflation expectations on actual inflation. It is especially interesting to see whether the volatility, skewness or kurtosis of inflation expectations can give us any information about future inflation. A

[1] The concept of core inflation is not perfectly clear. The core or the so-called core inflation measures the permanent component of inflation by excluding volatile energy and non-processed fresh food out of the inflation concept, therefore core inflation is usually measured as harmonised (HICP) inflation excluding unprocessed food and energy. Unprocessed food and energy prices are typically out of normal control since the prices of unprocessed food, such as vegetables are subject to climate conditions and lately also animal diseases that may affect price formation in a exogenous way. Energy prices are heavily affected by oil prices, which again are frequently influenced by Middle-East political turmoils and are hard to forecast.

related study using skewness of CPI commodity group data has been conducted by Suvanto and Hukkinen (2001).

During the past decade a lot of attention has been paid to consumers' expectations because they have been noticed to contain important information on people's economic behaviour in the future. The expectations also possess predictive power concerning macro-economic phenomena, e.g. indebtedness, unemployment and inflation (e.g. Berg & Bergström, 1996, and Djerf & Takala, 1997). If inflation expectations carry information about future inflation, then monetary policy authorities should already react to rising inflation expectations to neutralize their self-fulfilling nature. This is quite obvious, because households make plans and decisions to purchase all the time on the commodity market.

2. Data description

2.1. About the consumer survey

Statistics Finland introduced the Consumer (Barometer) Survey in November 1987. From October 1995, the data have been collected monthly in accordance with a harmonised data collection method of the European Union (European Commission, 1997). The Consumer Survey examines Finnish consumers' evaluations and expectations on the general economic situation in Finland and on that of their own households (i.e. consumer sentiment). Furthermore, it also considers consumers' intentions to make purchases, save and raise loans.

Until December 1999, the data for the Consumer Survey were collected together with the data of the Labour Force Survey. Six rotating panels were then used in the Consumer Survey. The same person was asked the same questions three separate times at six-month intervals, and each month a third of the target persons were new. Statistics Finland's field interviewers conducted the telephone interviews around the country.

From January 2000 on, the Consumer Survey has had a totally new individual sample each month. All the interviews are now conducted from Statistics Finland's CATI Centre by about 35 interviewers. The target area is still the whole country, and the respondents represent the 15 to 74-year-old population in Finland, according to age, gender, region and native language.[1] The change in the data collection method has had some effect on the results of the survey

[1] Data collection for the Consumer Survey takes place mainly during the first fortnight of each month, and the results are published on the last working day of the same month. The gross sample size of the Consumer Survey is 2,200 persons monthly. Systematic random sampling (SYS) is used to extract the sample from the updated Central Population Register. The sorting system of the sampling frame is based on geographical population density. Non-response being currently as high as 25 per cent, a little more than 1,600 persons usually respond to the survey each month.

(Kangassalo and Heiskanen, 2001), but it looks apparent that no clear change can be seen in the distribution moments of inflation expectations.

The results of the Consumer Survey are weighted up to the total population by means of sample weights. Weighting improves the accuracy of the data. The weights are established by using the probability of each observation to be included in the sample. Individual weights are computed by CALMAR software, so that the estimated marginal distributions of certain background variables (region, person's age group and gender) correspond to the marginal distributions in the whole population, i.e. the population structure.[1]

2.2. How the questions about inflation are asked in the consumer survey

There are four questions in the Finnish Consumer Survey that deal with price changes in the past and future. The two EU harmonised questions with answer categories precede the two percentage questions, so that the questions dealing with assessment come first and these are then followed by those about expectations. The questions about inflation assessment (past inflation) are:

EUp. *"Compared to 12 months ago, do you think consumer prices:"*
1 have risen a lot
2 have risen moderately
3 have risen slightly
4 have hardly changed
5 have fallen
9 don't know.

SFp. *"By how many per cent do you think consumer prices have changed during the past 12 months? Please give your answer at the accuracy of one decimal."*

Correspondingly, inflation expectations are asked with the following:

EUf. *"Compared to now, do you think that in the next 12 months:"*
1 prices will increase more rapidly
2 prices will increase at the same rate
3 prices will increase at a slower rate
4 prices will stay about the same
5 prices will fall
9 don't know.

[1] Weighting generally has only a minor effect on the actual values of the sentiment scales. There are several reasons for this. First, the sample design (SYS) produces a self-weighting sample. Second, non-response has been relatively constant in different population groups and, therefore, has not caused considerable bias in the material. Third, questions concerning opinions and the extracted indicators have shown only a weak correlation with the variables used in weighting. Opinions tend to vary quite considerably in different population groups.

SFf. *"By how many per cent do you think consumer prices will change during the next 12 months? Please give your answer at the accuracy of one decimal."*

In respect of both assessments and expectations, if the respondent chooses alternative 4 (no change in prices) in an EU question, the corresponding SF question is not asked and the system automatically records a zero against it. If alternative 6 (don't know) is chosen, the percentage question is skipped.[1] In general, no perfect harmony is required between assessments and expectations, but technically these four questions have worked reasonably well together. However, according to the interviewers' experiences, the questions on price changes, especially the one on the per cent change and on the decimal point are among the most difficult ones on the questionnaire. Nevertheless, they very rarely totally interrupt the interviewing even though they are placed at the beginning of the questionnaire just after the opening questions about general economy and unemployment.

In the following we use as measures of inflation assessments and expectations the monthly observations in percentage points from the Finnish Consumer Survey between October 1995 and December 2002. The monthly cross-sectional data sets were available from October 1995 to December 2002. The monthly number of respondents varied between 1,485 and 2,095.

Consequently, the main aim is to measure the same concept of inflation that is published and generally known via media like newspapers, television, etc., so consumers are able to obtain information on this phenomenon almost free of charge. Yet it seems quite clear that only some consumers follow this information closely. However, it should be borne in mind that the actual inflation rate of the current month is not known at the time of the interview and that in practice consumers only observe direct the prices of a limited number of goods.

2.3. About the consumer price index

The consumer price index (CPI) is used below as the main measure for the actual inflation, instead of e.g. the harmonised or core inflation, because inflation expectations are probably mostly connected with the consumer price inflation published widely in the media.[2] Technically, we can justify choosing the CPI for this purpose by the fact that categorical price indices are available for the CPI.

---

[1] There are also some consistency checks between the questions, e.g. if a minus signed assessment is given in the future percentage change question even though it has been previously stated that prices will increase, the system always gives a note about this contradiction and the respondents can change their answers although this is not obligatory. Large outliers are also excluded from influencing the reported results.

[2] It is also possible and problematic that for our purposes consumers may think about the HICP inflation or some other measure of inflation calculated with the EU method. In practice, all consumers have their own consumption bundles of goods and probably form their inflation assessments basing on these individual bundles.

The 2000-based CPI contains 12 different consumption groups, every one of which gets a weight that corresponds to the value of the goods in the consumer's commodity basket. In fact, at the beginning of our sample, the CPI was calculated with weights from base year 1990, and between 1998 and December 2001 the CPI was calculated and reported with 1995 weights. The CPI commodity groups and the weights for base year 2000 are presented in Table 1.

Table 1. CPI commodity groups and weights

CPI commodity groups and weights	Weights, % 2000	Weights, % 1995
01. Food and non-alcoholic beverages	13.83	15.76
02. Alcoholic beverages and tobacco	5.97	6.16
03. Clothing and footwear	4.99	4.60
04. Housing, water, electricity, gas and other fuels	19.48	22.10
05. Furnishing, household equipment and routine maintenance of the house	5.03	4.67
06. Health	4.29	4.12
07. Transport	15.51	13.58
08. Communication	3.48	2.21
09. Recreation and culture	12.41	11.93
10. Education	0.57	0.15
11. Restaurants and hotel	7.58	6.48
12. Miscellaneous goods and services	6.86	8.23

Source: Statistics Finland, Consumer Price Index

3. Basic analysis of inflation assessments and expectations

3.1. Inflation assessments

Our starting point is how well consumers' valuations concerning the past inflation match the actual measured inflation. First of all, it should be noted that, on average, consumers seem to be able to assess the current inflation rate with certain accuracy, since the correlation is the highest between current inflation assessments and current inflation and not, for instance, the known inflation rate of the previous month. The inflation assessment figure is published at the end of the month, while the actual inflation of rate of the previous month calculated by Statistics Finland is published in the middle of the following month.

Figure 1 shows that from the end of 1995 to June 1996 the consumer assessment of the CPI was negative while the actual CPI had already become positive. This negative assessment period probably occurred partly because actual inflation was still slightly negative in the last quarter of 1995. However, the obvious reason may be the fast decline in food prices in Finland during 1995 due to increased competition as Finland joined the EU. This 10 per cent fall in food prices may have made consumers believe that the effects could extend to the rest of the economy as well (Figure 2). However, food only accounted for about 16 per cent of expenditure, so it cannot explain the entire bulk of the perceived deflation. No decline of prices happened in other expenditure except for that on communication (weight 3.5), It was noticed rather soon that the fall in food prices did not continue, so also the assessment estimates became smoother, changed into positive and reached the level of actual inflation. Since then the shape and the level of the assessment have been corresponding fairly well with the experienced inflation for many years, i.e. consumers have been able to assess the measured inflation quite accurately. Towards the end of the year 2000 and at the beginning of 2001 inflation in Finland accelerated to as high as four per cent, mostly due to higher oil (energy) and housing prices, but the assessment only followed this sluggishly. Since the adoption of the euro currency in January 2002, consumers' inflation assessment estimates have clearly exceeded the actual measured inflation rates. One plausible reason for this might be the inflation assessment's sensitivity to inflation in services and other menu-list price changes.

Figure 1 tells us already that something very exceptional has happened during the changeover to euro in Finland among the euro countries. Even though consumers have been surely informed by newspapers and the media that inflation was not greatly affected by the changeover, consumers do not seem to believe it or they only pay attention to some particular items of the CPI. According to surveys conducted by the Consumer Research Institute the price awareness of consumers has not deteriorated significantly because of the euro changeover.[1] Thus, this phenomenon seems to be at least partly psychological. Price increases have been accused of having arisen from the changeover to euro, although stable prices for goods have been clearly marked in both national currencies and euros. The price decreases that have also occurred have been argued to have been caused by other reasons. For instance, the most clearly observed price decreases that have been seen in fast food (mainly hamburgers) have been understood to have happened as a consequence of competition, not because of the changeover.

The introduction of the euro as the payment instrument did however change the pricing habits of service producers, since 80 per cent of service prices are currently being rounded to even zero cent prices. It should also be noted that Finland was the only country that did not issue the smallest one and two cent euro coins into circulation. Prices were rounded to 5 cent because of this, which also increased the number of roundings in prices. During the Finnish markka period these roundings were used much less. This effect could be interpreted partly as natural menu-cost pricing behaviour if it turns out that the current price level will prevail for a longer time than usual. Probably This is probably also the way consumers see the phenomenon since inflation expectations declined at the same moment as past inflation assessments increased. One theory proposed to explain high inflation assessment is that consumer have used proxy exchange rates in

[1] Kuluttajatutkimuskeskus (Consumer Research Institute).

their minds while making price comparisons (D' Elia, 2003). For instance, in Finland the official exchange rate was 1 euro = 5.94573, while consumers may have used simply six as conversion rate, which is 0.9 percent higher. In Germany official exchange rate was 1.956, while using two consumers may have made 2.26 percent error. Based on difference in official and proxy exchange rate, this effect might have been strong (above 2 %) in Italy, Greece, Germany, moderate (about 1.5 %) in Austria, France, Ireland and probably negligible or negative in Belgium, Netherlands, Portugal and Spain.

However, there have to be also other reasons for this kind of "common misperception". A quite natural explanation arises from the fact that the measured consumer price index???flation (CPI) has a bundle of goods that represent total aggregate consumer expenditure weights, while consumers typically make inflation assessments from a narrower perspective. For instance, falls in housing costs due to a decline in interest rates might not be regarded as lower inflation, while a rise in petrol prices will surely be registered as rise in inflation. The familiarity of prices is also certain to affect greatly the perceived rate of inflation. At the beginning of 2002, the prices of unprocessed food (fresh vegetables and meat) increased significantly throughout the euro area, which was probably also related to the introduction of the euro, although the actual connection was weak.

Table 2 tests for evidence about the consumption items that affect most the inflation assessment. According to the performed regression analysis, inflation assessment is significantly influenced by food and beverages prices, housing costs, health expenditure prices, and restaurant and hotel service prices. Due to stationarity of the commodity group inflation, spurious correlation should not be a problem. Therefore, the results seem to indicate that consumer inflation assessment is clearly influenced by the prices on which consumers receive a lot of information daily. Basing on the regression analysis, the most important commodity group affecting past inflation assessment is food and beverages prices. It is quite probable that consumers follow food prices closely while making their daily shopping. The same applies to restaurant and service prices. The importance of housing costs might also relate to their large share of the total expenditure that is reported.

Another feature already mentioned in connection with the changeover to euro has been the price awareness of consumers. Price increases happen more often at the shift of the year, and it seems likely that at least temporarily price changes were looked at more closely in this special case of change in currency. It is clear that altogether more prices were changed than normally due to the euro changeover. This seems also clear while comparing the CPI with the median inflation that had been close to each other from beginning of 2002 (Figure 3).[1] Therefore, it may be that the few increases were thought to indicate even higher inflation than was actually measured by their consumption expenditure weights.

[1] Median inflation is calculated here as the 50th percentage point of the cumulative distribution of ordered frequency distribution of the price increases of 492 commodities (goods and services).

In close relation to this, we also calculated the correlation of the inflation assessment series with various different measures of inflation. Although the correlation was significant with all the measured inflation concepts, there seems to be some support to the assumption that during the estimation period consumers' inflation assessment could measure the core or harmonised inflation more closely than the ordinary CPI inflation. This does not seem to be entirely surprising since, for instance, energy prices are quite volatile and, therefore, hard to follow and difficult to forecast. The same applies largely to unprocessed food prices. In Finland, the CPI inflation measures the housing costs of owner-occupants from the basis of house prices and housing loan interest rates rather than from rents. Housing costs form a significant part of the services inflation that accounts for over 40 per cent of the HICP.

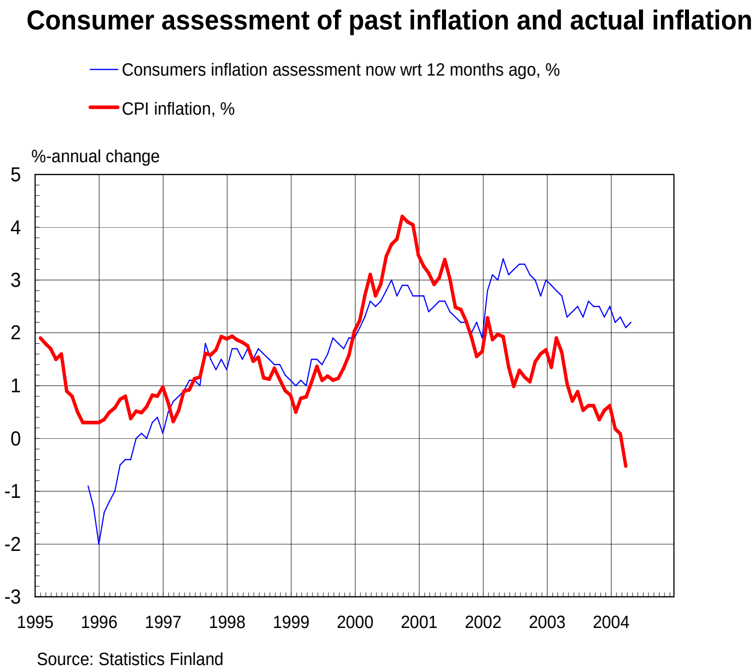


Figure 1.

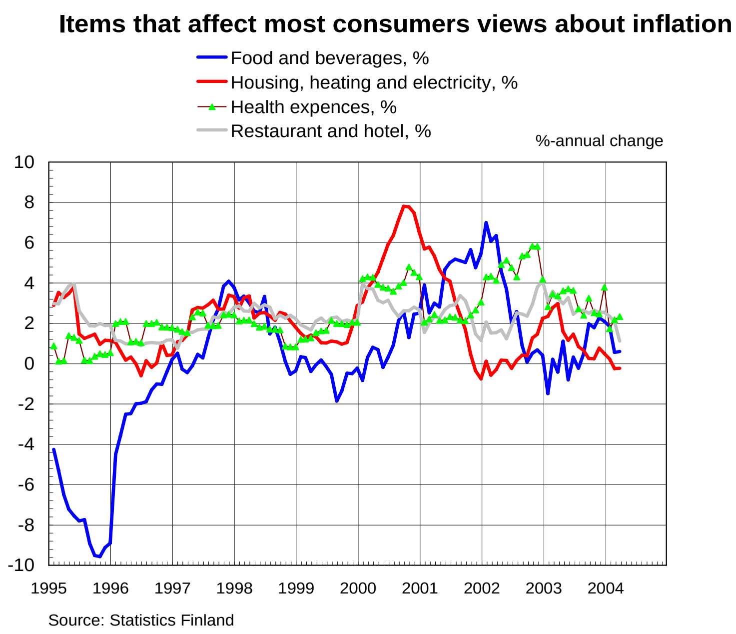

Items that affect most consumers views about inflation
Food and beverages, %
Housing, heating and electricity, %
Health expences, %
Restaurant and hotel, %
%-annual change
10
8
6
4
2
0
-2
-4
-6
-8
-10
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
Source: Statistics Finland

Figure 2.

Table 2. Dependence of inflation assessments on commodity group inflation

Consumers inflation assessments and expections regressed wrt CPI commodity groups
Estimation period: 1995/m10 - 2002/m12

	Inflation assessments			Inflation expectations		
Variable	Coefficient	t-value	Instability test	Coefficient	t-value	Instability test
Food	0.22	6.85	0.35	0.08	4.20	0.10
Alcohol	-0.30	-3.10	0.77 **	0.48	8.50	0.08
Clothing	-0.03	-0.63	0.55 *	0.01	0.23	0.43
Housing	0.10	2.59	0.09	0.12	5.48	0.10
Furnishing	0.24	2.27	0.38	0.14	2.23	0.09
Health	0.42	7.44	0.64 *	-0.06	-1.83	0.05
Transport	0.06	0.93	0.64 *	-0.05	-1.26	0.21
Communication	0.03	1.13	0.87 **	-0.01	-0.57	0.16
Culture	-0.11	-0.92	0.20	0.06	0.93	0.03
Education	-0.07	-1.40	0.24	-0.06	-2.17	0.05
Restaurants	0.29	3.43	0.24	0.37	7.49	0.09
Miscellaneous	-0.02	-0.68	0.06	0.01	0.47	0.12

Performance:
$R^2 = 0.949$ \ $\sigma = 0.489$ DW = 0.735
Instability tests, variance: 0.284 joint: 5.02203**

Performance:
$R^2 = 0.983$ \ $\sigma = 0.282$ DW = 1.36
Instability tests, variance: 0.470 joint: 2.919

Diagnostics:	p-value	Significance	Diagnostics:	p-value	Significance
AR 1- 6 F(6, 69)	[0.0000]	**	AR 1- 6 F(6, 69)	[0.0019]	**
ARCH 6 F(6, 63)	[0.0744]		ARCH 6 F(6, 63)	[0.0034]	**
Normality $Chi^2(2)$	[0.0391]	*	Normality $Chi^2(2)$	[0.0000]	**
Xi^2 F(24, 50)	[0.0000]	**	Xi^2 F(24, 50)	[0.0375]	*
RESET F(1, 74)	[0.0016]	**	RESET F(1, 74)	[0.0000]	**

Correlations of assessment and expectations between inflation measures
Period: 1995/m10 - 2002/m12, 84 observations (all correlations signif. at p < 0.01)

Variable	KB5 assessment	KB6 expectations	CPI inflation	HICP inflation	HICP excl. fresh food and energy	CPI core inflation	CPI median inflation
KB5 (assessment)	1.00	0.66	0.70	0.73	0.78	0.81	0.73
KB6 (expectations)	0.66	1.00	0.89	0.84	0.64	0.73	0.69
CPI inflation	0.70	0.89	1.00	0.95	0.71	0.77	0.66
HICP inflation	0.73	0.84	0.95	1.00	0.75	0.79	0.71
HICP excl. food, energy	0.78	0.64	0.71	0.75	1.00	0.97	0.95
CPI core inflation	0.81	0.73	0.77	0.79	0.97	1.00	0.94
CPI median	0.73	0.69	0.66	0.71	0.95	0.94	1.00

*) Median inflation is calculated as median from 492 commodity inflation

3.2. Inflation expectations and the changeover to euro

Forecasting inflation with reasonable precision is already in principle a difficult task, since the statistical process of the inflation percentage is quite close to a random walk process and inflation changes are, therefore, very unpredictable.[1] The intuitive interpretation of a random walk process in this case relates to the function of prices to equate and clear the demand and supply towards market equilibrium. On an efficient market, the price changes around the equilibrium point are also random. It is quite impossible to find any local or global trend from the published inflation rates. In the medium term it is to some extent possible to forecast inflation with some accuracy, for instance by means of money supply, economic activity, interest rates and other types of fundamentals. The inflation target of the ECB sets the upper limit for the HICP inflation close to 2 per cent, but preferably from below. If consumers believe this is credible, it certainly should also affect inflation expectations and make them stationary in the long run. In practice, there seems to be some room for inflation forecasts based on the origins of inflation arising from different sources of costs increases and demand pressures affecting prices of commodity groups. These factors include demand indicators, wages and other cost variables that could be regarded as exogenous sources of inflation.

It is important to consider what information inflation expectations are based upon. As we saw from the strong correlation between inflation assessments and expectations, there is good ground to think that the information set behind these two evaluations is very similar (Figure 4). However, inflation expectations are also influenced by other background factors that are only related to forward-looking behaviour, such as interest rates. If inflation target is credible, interest rate changes influence future inflation. Since the beginning of the euro changeover, inflation expectations have declined drastically, while at the same time perceived inflation has increased and, even more surprisingly, stayed stubbornly high while actual inflation has become lower. This is certainly hard to interpret. During the changeover once-for-all price increases were made, which probably affected perceived inflation due to price raises in food, restaurant and some traffic expenses. As at the same time, inflation expectations declined significantly: these price raises were not expected to continue or be repeated within the forthcoming year. Tthis decline was probably not so much an indication of expectations of a forthcoming slowdown in activity in the next 12 months but should rather be interpreted in terms of the cost factors that occured in connection with the euro changeover.

Therefore, consumers have naturally found it difficult to estimate future inflation. Because of the nature that reminds the random walk process described above, the best inflation forecast in the short run is close to the current observed inflation. Thus, we can ask how accurately consumers’ expectations of inflation follow the past assessments of inflation. As Figure 4 shows, these estimates and expectations clearly follow each other, even though there was obvious bias

[1] Although the autoregressive coefficient of the consumer price index does not differ significantly from unity, the order of integration with ML estimation is about 0.5 for the index level and for the annual change. A lot of discussion has been raised on whether the CPI inflation has a unit root and, therefore, the CPI index should be an I(2) -variable.

downwards in the inflation estimates from the end of 1995 until 1997 and, correspondingly, some bias upwards since January 2002. At a certain point of time, estimates of past inflation and expectations of future inflation are very similar. Most likely this relates heavily to the same information set behind both of these variables.

Figure 5 shows that the expectations of actual inflation lagged by one year, i.e. to the point of time where inflation expectations are targeted. Actual inflation and inflation expectations followed each other reasonably well until the end of 1997. However, consumers subsequently undervalued the rise in consumer prices and, respectively, from May 1998 on, overestimated the rise in prices. Again, during the year 2000 the expectations proved to be far too low. During that time, measured inflation was greatly influenced by an increase in house prices that was not assumed to raise future housing costs that much.[1] High oil prices also raised the inflation contribution of energy prices to one percentage point during 2000.

Inflation expectations lagged by one year do not seem to be particularly accurate. However, the performed Granger causality tests indicate that inflation expectations do have certain predictive power over future inflation as shown below in Chapter 5. However, this leading indicator effect does not last for the entire forecast horizon. According to dynamic regressions the effect is at its best at around 4 to 6 months. Nevertheless, it is undeniably significant. Similar results were also found by Pursiainen (1999) using spectral analysis.

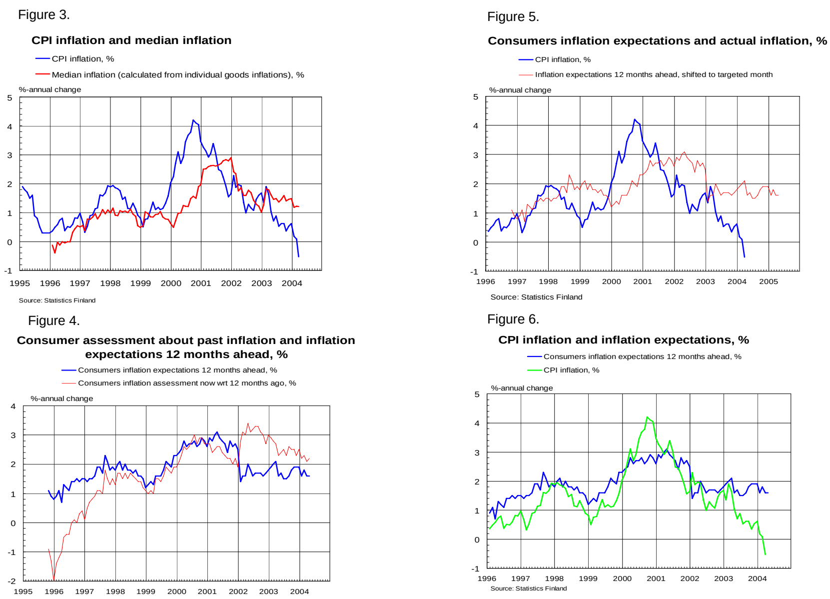
On the other hand we cannot think that there should be any dramatic contradiction between the expectations and the experienced inflation, because the changes in consumer prices are quite small and the forecast error between the expectations and the actual inflation is about two percentage points at worst. Compared with the general level of inflation the difference is quite big, though. Furthermore, we have to state that the current low level of inflation probably makes the estimating task more difficult for consumers. When inflation is nearly zero, the directions of price changes may be hard to forecast, because there are typically movements to both directions in relative prices.

We can also clearly see from Figure 6 that inflation expectations follow the actual inflation at the time forecasts were made. Thus, inflation expectations are heavily affected by the recently experienced inflation due to the inertia of inflation and the random walk behaviour of the inflation process. In addition, as future contains inherent uncertainty, consumers cannot know the surprises forthcoming in the inflation process. The confidence bounds of the inflation process, i.e. the uncertainty in the forecast, grow very rapidly during just a few months.

We saw that from the beginning of the changeover to euro in January 2002, inflation assessments and expectations have diverted greatly. In addition to the

[1] In Finland, UK and Sweden house prices affect consumer price inflation direct as capital costs of owner-occupied housing is affected by a capital user cost calculations. Ordinary rents behave more smoothly and do not contribute inflation that much.

other reasons already mentioned, it is likely that the large number of price changes due to the introduction of the euro currency affected this divergence. However, as we saw already when comparing past inflation and inflation assessments, the divergence was apparent and interesting.



Figures 3—6

4. Factors affecting inflation expectations

When we study consumers' expectations of inflation it is also interesting to examine which are the commodity groups, e.g subgroups in CPI where price changes have the greatest influence on the formation of expectations. On the other hand, we are also interested in discovering the macro-economic variables that are reflected in people's estimates of past development of inflation. It is also possible to identify which macroeconomic variables have not been included in the forecast information set by regressing forecast errors with lags of these macroeconomic variables. However, this sort of reasoning does not exactly say whether it is possible to predict these variables. It simply says that these variables are missing from the information set or have not always been properly taken into account.

We may think that inflation assessments and expectations are largely affected by price changes in expenditure groups such as food and housing, on which a major part of income is consumed. It is also possible that households follow especially price trends in such commodity groups that are purchased almost daily. Consumer durables and dwellings are purchased more infrequently and knowledge about their price changes may not be accurate. Consumers also tend to follow the prices of certain groups of durables only when they are contemplating purchasing those in the near future and not otherwise. Therefore, current knowledge of these goods is not very widespread. One fact supporting this view is that information on the price development of restaurant services seems to receive over-emphasised significance.

It seems quite reasonable to conclude that the same factors that affect inflation assessments are also behind the formation of inflation expectations (table 2). As we have seen, few special commodity groups receive great deal of attention in the formulation of inflation assessments and expectations. From the beginning of the euro changeover in January 2002, inflation assessments and expectations have diverted quite radically.

4.1. Correlation of expectations with different inflation measures

We have seen that although consumers may be informed most by the published overall CPI inflation, they seem to react in an uneven manner to the different components of inflation. To be more specific, this is not necessarily very odd since in practice every consumer has a different expenditure bundle of himself or herself. Consumers also get new information on different consumption groups infrequently. Food and beverages are shopped almost daily, while few other items like durables or most services are bought seldom. Car owners certainly pay attention to petrol prices, while persons driving leased company cars do not necessarily consider them at all. The motives for monitoring prices may, thus, also differ. Therefore, it may be worthwhile to investigate different measures of inflation to see whether any differences can be seen in assessment and expectation reactions to these measures. [1]

There could also be other reasons for different reactions to different consumption items. The transparency of different services varies, as many public services are calculated as expenditure based on imputed costs. Public goods prices are changed only at specific times, like the beginning of the year, and if these services are used rarely they are not remembered so clearly. On the other hand fresh food prices may vary due to crops and are, therefore, remembered more clearly. Some consumer goods, such as clothes or durables, are only bought

[1] The ordinary CPI is the most common and widely used inflation measure supposed to measure inflation weighted with private consumption goods expenditure. In Finland, as well as in the UK and in Sweden, house prices and housing loan interest rates affect the housing costs of owner-occupier dwellers. The harmonised CPI (HICP) is the unified ECB measure of inflation for monetary policy purposes, which excludes owner-occupied housing expenditure (weight 6.8%), consumer credit interest payments (weight 0.53%) and lottery games (weight 2.1%) and a few items of social and health expenditure. Core inflation is further reduced from the HICP excluding energy and unprocessed food prices.

during (seasonal) sales. Therefore, changes in their normal prices are not investigated frequently.

It is also not clear whether different measures of inflation correlate differently with the moment generating function behind inflation expectations. Table 3 shows the correlation between different inflation measures and the moments of inflation expectations. The first observation from these correlations is that the CPI and HICP inflations correlate heavily with each other, while core inflation and median inflation also seem to measure more closely the same inflation components. Second, with the central moments of expectations skewness seems to be quite instantaneously uncorrelated with inflation measures as well as with the other moments of expectation distribution. With respect to distribution moments other than asymmetry (skewness) and with the mean of expectations, standard deviation and kurtosis may occasionally contain some information about future inflation, but this information is not particularly stable. It should also be noted that when inflation is low, kurtosis is high and distribution of expectations very concentrated, and vice versa when inflation is high.

Table 3. Correlations between inflation measures and moments of expectations

Variable	Correlations between inflation measures and expectation moments Marked correlations are significant at p<.01000, N=81							
	CPlg	HICPg	CPlund	CPlmed	Mean	Stdev	Skewness	Kurtosis
CPlg	1.00	0.95	0.78	0.66	0.91	0.77	-0.18	-0.71
HICPg	0.95	1.00	0.79	0.71	0.87	0.82	-0.24	-0.69
CPlund	0.78	0.79	1.00	0.94	0.77	0.67	-0.15	-0.58
CPlmed	0.66	0.71	0.94	1.00	0.73	0.64	-0.15	-0.56
Mean	0.91	0.87	0.77	0.73	1.00	0.80	-0.07	-0.81
Stdev	0.77	0.82	0.67	0.64	0.80	1.00	-0.44	-0.65
Skewness	-0.18	-0.24	-0.15	-0.15	-0.07	-0.44	1.00	0.09
Kurtosis	-0.71	-0.69	-0.58	-0.56	-0.81	-0.65	0.09	1.00

4.2. Dynamic correlation between inflation and expectations

After a static correlation analysis, we must turn to dynamic correlations to investigate the forecasting properties of expectations. Dynamic relationship between inflation and inflation expectations can be analysed by investigating cross-correlation functions. Plotting the cross-correlation function between the CPI inflation and inflation expectations at the time they are published reveals that the highest correlation is attained at lag zero. Therefore, inflation expectations are heavily dependent upon the time the expectations are formed (Figure 7). Of course, it would be nice to have expectations that correlate strongly with exactly the one-year ahead inflation, but for various reasons, such as the autoregressive nature of the inflation process and cumulative innovations (or even white noise errors), this is in practice impossible. The relevant question therefore, is whether inflation expectations made this month contain any relevant information about future inflation process, not necessarily for exactly 12 months, but for a shorter horizon.

The estimated cross-correlation function is clearly asymmetric and expectations are significantly correlated up to the 12th month as they were supposed to do. On the other hand, actual inflation only correlated with lagged expectations only up to 9th lag. The same observation can be made from the different presentations in Figure 8.

The ordinary CPI inflation includes a few inflation components that are difficult to forecast, such as owner-occupied housing costs including house price effects and effects from interest rates, in addition to energy (oil) prices, so it is no wonder that consumers cannot predict these items. Qualitatively the same interpretation can be made even more clearly from the cross-correlation functions between inflation expectations and core inflation and also median inflation (Figures 9-10). Basing on these measures of inflation, inflation expectations seem to correlate mostly with future inflation 4 to 5 months ahead. The trouble with cross-correlation function analysis is that the series under investigation have similar time series processes.[1] A bivariate VAR(4) model estimated for the CPI inflation and inflation expectations showed clearly the predicting power of expectations for the next few months. A Wald test imposed on the lagged (lags 1-4) expectation coefficients turns out to be very significant in indicating the explanatory power of expectations.

---

[1] Therefore, the series could either be pre-filtered and the cross-correlations of the resulting innovations could be studies, or a bivariate VAR system could spesified to study the predictive content of inflation expectations.

Figure 7.

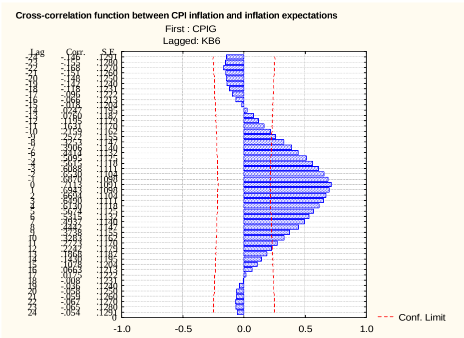


Figure 8.

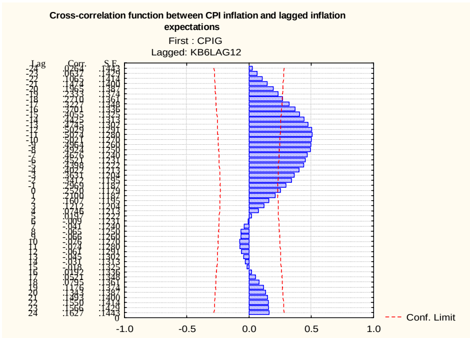


Figure 9.

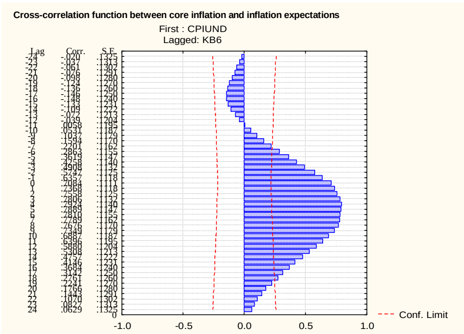
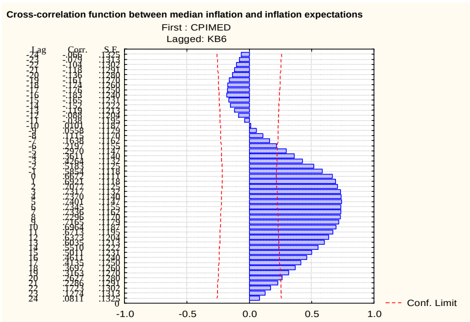


Figure 10.



5. Can we forecast inflation with inflation expectations?

During the past couple of decades, inflation expectations have been taken increasingly into consideration in examinations of inflation processes in many countries. Furthermore, a lot of interest has been paid to the influence of different economic policy decisions, particularly interest rate policy, on the formation of inflation expectations. Several studies have proved that consumers' expectations on inflation do have predictive power over the inflation process, and that there is an obvious connection between interest rate formation and the development of inflation expectations. Next, we will examine whether actual inflation is predictable with inflation expectations. In addition, we will try to find out the forecast horizon of the expectations and the strength of their predictive power with respect to inflation.

5.1. Granger causality tests on the time series

Although we have already analysed the cross-correlation functions between the Consumer Survey variables, it must be remembered that correlation in itself does not necessarily imply causation in any useful sense. Granger causality tests can be helpful in establishing the degree of usefulness of expectations in forecasting models. Granger causality tests pose the question of causation from the point of predictive ability. The question of whether expectations “Granger-cause” inflation is approached by looking at how much of the current inflation can be explained by past values of inflation and then seeing whether the lagged values of inflation expectations can improve this predictive explanation.[1]

By using Granger causality tests we can also examine whether variables are exogenous or endogenous with respect to other variables, i.e. can variable y be explained by the history of variable x. Granger causality can also be two-sided, in which case past inflation will also predict future expectations. In this case we mainly aim to find out whether actual inflation can be forecast by the history of inflation expectations. If not, at least for this purpose inflation expectations would be useless to the forecasting of inflation.

Table 4 reveals few interesting findings about estimated Granger causality tests. First of all, as it is clear that recently observed inflation affects inflation expectations, it also turns out that even if consumers knew exactly the past inflation process, their inflation expectations include additional predictive contents over future inflation up to 12 months ahead at 5 per cent significance level. Quite surprisingly this result holds true with the CPI inflation, but not with the median inflation. However, the kurtosis of expectations is useful for median inflation. Basing on performed tests, median inflation also appears to affect inflation expectations strongly. Another interesting observation is that the skewness of consumers' inflation expectations can predict actual inflation within 3 to 6 months. This is in accordance with the idea that arising inflationary pressures may show out in the right hand side tail of expectations even before they spread more widely to the overall inflation.

The causality tests showed that actual inflation could be effectively foreseen from the history of inflation expectations. The test results depend on the number of lags used, but between the lags of 2 to 12 months inflation expectations are, on average, significant in predicting the CPI inflation. Although the accuracy of consumers’ point estimates for inflation expectations does not always seem to be

[1] Granger causality tests are performed by two linear regressions where an autoregressive model for inflation is complemented either with the lagged values of the consumer barometer variable in the equation and *whether* equation *without?*. Granger causality test is then based on an F-test for the lagged barometer response variables as a group. If this F-test is significant, we can conclude that the expectation variable reduces the forecast error of the inflation and is, therefore, useful at least in an autoregressive model. It should be borne in mind that Granger causality does not mean any strict philosophical relation between cause and effect, but merely that one variable can be helpful in predicting another. In this sense forecasting ability is a stronger property than pure explanatory coincidence, since it requires controllable predictability between variables when all the “third variable” effects are eliminated from the framework. Granger non-causality, i.e. absence of feedback from the dependent variable back to the independent variable, has, therefore, an important role in econometric model building (see e.g. Hamilton, 1994).

convincing, the expectations have predictive power for actual inflation according to the performed Granger causality tests. This emphasises the role of the expectations as a process.

One of the basic ideas we had in mind was that the kurtosis of inflation expectations may have some power in predicting inflation, as concentration of expectations might indicate reduced variation and uncertainty, and then slowing down of inflation pressures. The kurtosis seemed to have predictive power over the median inflation in the very near future, but evidence of this with the ordinary CPI inflation was rather weak.

Table 4 also presents test results related to short-term market interest rates and inflation variables. Short-term rates affect Finnish inflationary pressures very clearly before 1999, but after 1999 the eurosystem interest rates have not been influenced by the Finnish inflation history. This finding is quite understandable, especially during the euro period after 1999. However, it is more difficult to interpret why consumers' inflation expectations are not affected by interest rates at all after 1999. Producer prices seem to affect inflation expectations strongly according to these Granger causality tests. Of course, inflation expectations do not predict producer prices. According to the performed tests the relationship between the growth of GDP and inflation expectations is not particularly strong. There is only weak evidence that expectations may have some power in predicting GDP one year ahead.

We may take a closer look at Granger causality between interest rates and inflation expectations as well as inflation assessments to see whether monetary policy reacts on inflation and whether monetary policy affects the future inflation rate. In particular, we may ask whether the interest policy of the ECB has affected inflation in Finland since the establishment of the monetary union at the beginning of 1999. Of course, interest policy is not activated for Finnish inflation purposes (only), but we could expect that interest rates might also affect domestic (expected) inflation pressures. Table 5 presents additional evidence that the interest rate effect of the short-term rate (3 month Helibor/euribor rate) on inflation expectations may be different before and after the fixing of the exchange rate to euro and during the changeover from national monetary policy to ECB interest rate policy. Using the period 1995/m10 to 1998/m12, it is the case that higher interest rate reduces inflation expectations during the next 6 months, while from beginning of 1999 this relationship turns to positive. Even though the estimation period is very short, it may refer to change in the effectiveness of monetary policy to affect inflation expectations during the euro period.

Table 4. Granger causality tests for inflation and inflation expectations

Granger causality tests between inflation and inflation expectations
Estimation period: 1995/m10 - 2002/m12

		Granger causality test p-values				
Causation		Lag length				
From	To	1	2	3	6	12
KB6	CPIg	0.296	0.010	0.010	0.020	0.048
CPIg	KB6	0.001	0.056	0.204	0.297	0.069
KB6	CPImedian	0.690	0.438	0.096	0.124	0.636
CPImedian	KB6	0.004	0.007	0.014	0.161	0.124
SKEWNESS	CPIg	0.723	0.856	0.005	0.016	0.350
CPIg	SKEWNESS	0.220	0.572	0.522	0.162	0.542
KURTOSIS	CPIg	0.296	0.279	0.059	0.183	0.227
CPIg	KURTOSIS	0.001	0.009	0.022	0.146	0.443
KURTOSIS	CPImedian	0.003	0.002	0.023	0.151	0.112
CPImedian	KURTOSIS	0.112	0.615	0.637	0.590	0.291
EHEL3	CPIg	0.563	0.035	0.017	0.034	0.019
CPIg	EHEL3	0.439	0.141	0.522	0.508	0.163
EHEL3	KB6	0.941	0.681	0.240	0.399	0.605
KB6	EHEL3	0.000	0.033	0.065	0.154	0.510
PRODPRI	KB6	0.006	0.015	0.017	0.016	0.056
KB6	PRODPRI	0.631	0.466	0.569	0.078	0.115
GDP	KB6	0.244	0.294	0.313	0.801	0.574
KB6	GDP	0.369	0.197	0.108	0.120	0.049

Variables:

KB6 = Consumers inflation expectations for 12 months ahead, %
CPIg = Consumer price inflation (CPI), %
SKEWNESS = Skewness (3. moment) of the individual inflation expectations
KURTOSIS = Kurtosis (4. distribution moment - 3) of the individual inflation expectations
CPImedian = Median of CPI commodities inflation rates from individual goods (492 items) data, %
EHEL3 = Short-term (3 month) euribor market interest rates, %
GDP = GDP monthly indicator (Statistics Finland), %-annual growth
PRODPRI = Producer prices in industry, %-annual change

Table 5. Interest effect on inflation expectations before euro and during euro

Before euro connection

Modelling KB6 by OLS

Estimation period: 1995/m10 - 1998/m12

Variable	Coefficient	t-value	PartR^2	Instab
Constant	2.2055	9.477	0.7434	0.11
EHEL3M	0.8700	3.111	0.2379	0.10
EHEL3M_1	-1.0881	-2.465	0.1638	0.10
EHEL3M_2	0.6765	1.423	0.0613	0.10
EHEL3M_3	-0.2483	-0.523	0.0088	0.10
EHEL3M_4	0.0504	0.105	0.0004	0.10
EHEL3M_5	-0.0586	-0.133	0.0006	0.09
EHEL3M_6	-0.3404	-1.543	0.0713	0.09

R^2 = 0.824 F(7,31) = 20.742 [0.0000] \σ = 0.177 DW = 1.68

Residual diagnostics:
AR 1- 3 F(3, 28) = 0.526 [0.6681]
ARCH 3 F(3, 25) = 0.204 [0.8930]
Normality Chi^2(2)= 3.329 [0.1892]
Xi^2 F(14, 16) = 0.331 [0.9781]
RESET F(1, 30) = 4.248 [0.0480] *

Solved Static Long Run equation

KB6 =	+2.206	-0.1385 EHEL3M
(SE)	(0.233)	(0.0666)

Test for restricting the interest rate coefficient to zero:
WALD test Chi^2(1) = 4.3284 [0.0375] *

During europeriod

Modelling KB6 by OLS
Estimation period: 1999/m1 - 2002/m12

Variable	Coefficient	t-value	PartR^2	Instab
Constant	0.1266	0.401	0.0040	0.47
EHEL3M	0.4828	1.395	0.0464	0.41
EHEL3M_1	-0.0094	-0.016	0.0000	0.38
EHEL3M_2	0.0659	0.110	0.0003	0.35
EHEL3M_3	0.0288	0.048	0.0001	0.33
EHEL3M_4	-0.0165	-0.028	0.0000	0.32
EHEL3M_5	0.3634	0.626	0.0097	0.32
EHEL3M_6	-0.3537	-1.031	0.0259	0.32

R^2 = 0.656 F(7,40) = 10.915 [0.0000] \σ = 0.351 DW = 0.624

Residual diagnostics:
AR 1- 3 F(3, 37) = 12.031 [0.0000] **
ARCH 3 F(3, 34) = 4.163 [0.0129] *

Normality Chi^2(2)= 1.860 [0.3944]
Xi^2 F(14, 25) = 0.808 [0.6542]
Xi*Xj F(35, 4) = 0.267 [0.9878]
RESET F(1, 39) = 0.103 [0.7496]

Solved Static Long Run equation (std. errors in parenthesis)
KB6 = +0.1266 + 0.5613 EHEL3M
(SE) (0.315) (0.083)
Test for restricting the interest rate coefficient to zero:
WALD test Chi^2(1) = 45.826 [0.0000] **

Variables: KB6 = inflation expectations (%), EHEL3M = Helibor/euribor 3 month market interest rate (%)

5.2. Testing for weak rationality of inflation expectations

Lagged values of inflation form the benchmark information set for both expectations and the actual inflation data generating process. From forecast errors between lagged (12 months) expectations and actual inflation we may analyse whether expectations are inefficient and insufficient with respect to what actually happened. This weak rationality of inflation expectations can be tested by means of a simple regression model as suggested originally by Mincer and Zarnowitz (1969) and recently applied by Forsells and Kenny (2002). This regression uses actual inflation as the dependent variable, which is regressed with the expectation series lagged 12 months to the period it intends to forecast

$$\pi_t = \alpha + \beta \pi^e_{t|t-12} + u_t , \qquad (1)$$

where π_t is the observed inflation measure in month t and $\pi^e_{t|t-12}$ is the inflation expectation made 12 months ago for that month. The joint null hypothesis for weak rationality is H_o: $(\alpha, \beta) = (0,1)$. If the null hypothesis is rejected, expectations are not linearly unbiased or weakly rational. The constant measures the unbiasedness in the mean of expectations and β measures the linear efficiency of expectations to forecast variation in inflation. Since the constant and the regression coefficient are in this case negatively correlated, the null hypothesis must be tested as a joint hypothesis e.g. with a F-test. [1]

Table 6 shows the results of the weak rationality tests. It can be seen that inflation expectations have been biased downwards for forecasting the CPI or HICP one year ahead, mostly due to unexpected rises in energy and housing prices in Finland during 2000. As parameter β is often well below unity, we

[1] The correlation between the regression coefficients is in this case negative as the covariance can be represented as $Cov(\alpha, \beta) = -\bar{\pi}^e_t * Var(\beta)$ where $\bar{\pi}$ is the mean of the forecast. The standard deviation of the expectations was .62, while the standard deviation of the actual inflation was .85 between 1996/m10 and 2002/m9, so in this sense the expectations were more conservative than the variance of the actual inflation.

conclude that expectations are somewhat inefficient in measuring variance in the actual CPI or HICP. However, it should be remembered that the variance of inflation expectations should always be smaller than the variance of actual inflation, since the error variance is always non-negative and the following identity would hold true:

$$Var(\pi_t) = Var(\pi^e_{t|t-12}) + Var(u_t), \text{ where } Var(u_t) \geq 0. \tag{2}$$

Consumers' inflation expectations have followed more closely the path of core and median inflations, which can be seen also from the regression results. The F-test does not reject the weak rationality of expectations with respect to these alternative measures of inflation. The same can also be seen clearly from the graphs comparing expectations with these measures (Figure 11).

We also tested the efficiency of inflation expectations by means of the following equation:

$$\pi_t - \pi^e_{t|t-12} = \delta + \gamma\Omega_{t-12} + u_t, \tag{3}$$

where Ω_{t-12} indicates the macroeconomic information available at the time (t-12) the expectations were formed (see Forsells and Kenny, 2002). Therefore, we cannot use any information that has appeared after the time expectations were formed.[1] As already discussed, it would be partly misplaced to test the point-estimate forecast properties of inflation expectations exactly 12 months ahead if we think of expectations as a process. In fact, we are more interested in testing whether a dynamic model specified to forecast inflation would be helped by using inflation expectations in addition to other leading indicators. It is not very surprising that the mean of inflation expectations is rather close to the mean of actual inflation. What seems to be a more prominent feature in expectations is their small variance in comparison to the volatility of actual inflation. This conservatism would be rational if inflation process were to be stationary and the forecast horizon long.

The results from this type of efficiency test can be seen from Table 7, which shows that, basically, consumers' inflation expectations do not use sufficiently all the information contained in the few macroeconomic variables affecting inflation, such as lagged values of earnings or producer and import prices. In the tests we used lags of between 12 and 17 months to account for the available information. With respect to GDP or unemployment rate, the efficiency of inflation expectations seems to hold according to this test. However, consumers' inflation

[1] Of course; we have started to use the lags from 12 although in practice this has not been readily available direct for the current month from statistics on e.g. GDP. With many indicators, publishing takes a few months although the economic activity has been observable to the consumers.

expectations seem to account for all the information related to the real economy, since output gap was not efficiently utilised in the expectations.

Table 6. Weak rationality tests for inflation expectations

Weak rationality regression tests for consumers inflation expectations (lagged 12 months)
with different measures of inflation, estimation period: 1996/m10 - 2002/m12

$$\boxed{\pi_t = \alpha + \beta\pi^e_{t|t-12} + u_t}$$

Null hypothesis: $\alpha = 0$ & $\beta = 1$ (t-values apart from null in parenthesis)

Inflation measure	α	β	F-test for null p-value	R^2	P-values for diagnostic tests DW	AR(5)	Normality
CPIg	0.97 (2.73)	0.44 (3.23)	0.0013**	0.08	0.10	.0000**	0.0000 **
HICPg	0.85 (3.14)	0.54 (3.51)	0.0027**	0.18	0.23	.0000**	0.0936
CPIund	-0.03 (-0.16)	0.93 (0.90)	0.7262	0.64	0.34	.0000**	0.0050 **
CPImed	-0.48 (-2.94)	0.93 (0.83)	0.8634	0.65	0.39	.0000**	0.7781

Variables:
CPIg = CPI inflation, %
HICPg = Harmonised CPI inflation, %
CPIund = Underlying (core) inflation, %
CPImed = Median CPI inflation, %

Table 7. The expectation forecast-error regressed with lagged macroeconomic indicators

Expectations efficiency regressions wrt different explanatory variables
Dependent variable: Inflation forecast error = inflation - infl. expectation (made 12 months ago)
Estimation period: 1996/m10 - 2002/m8

Independent variable (lags 12 - 17 months)	F(1, 63) -test p-values
GDP, %	0.1655
Earnings, %	0.0184 *
Producer prices, %	0.0000 **
Output gap, %	0.0024 **
Import prices of consumer goods, %	0.0003 **
Industrial production, %	0.1275
Unemployment rate, %	0.7133

Wald test for linear restriction: $\beta_i = 0$, i = 1, ...,6

5.3. Error-correction model for inflation

Like Forsells and Kenny (2002), we also tested for cointegration between actual inflation and inflation expectations and checked the cointegration properties of these variables. If forecasts are unbiased and linearly efficient estimators for inflation, it is quite natural to think that they would also be cointegrated. Basing on linear efficiency, the cointegration relationship should be homogenous to degree zero, i.e. the cointegration vector should fulfil restriction $[\beta_1, \beta_2] = [1, -1]$ (see Table 8). In addition in this case, the cointegrated variables should have identical means (unbiasedness), so the cointegration relationship should not even include an unrestricted constant. As we have seen in the performed causality tests, current inflation expectations "Granger cause" future observed inflation, which indicates that the causality runs at least in this direction and future changes in inflation can be seen beforehand in expectations.[1] Even though we think that there is basically more sense to regard inflation as stationary series, let us assume for a while that inflation may contain unit root and thus could be also cointegrated. If cointegration is confirmed, then the estimated error-correction term can also be used to forecast inflation. Here we have used core inflation as the measure of inflation, as it does not contain energy (oil) prices and unprocessed food prices, which are hard to forecast and sensitive to disturbances.

As a preliminary stage for testing cointegration, the order of cointegration must be tested for the variables included, as cointegration might hold true only between integrated variables. Cointegration in this case implies that inflation should be integrated of order one. This would also imply that the CPI is a I(2)-variable. Appendix 2 shows that probably because our sample is relatively short we cannot reject the null hypothesis of a unit root in any of the inflation measures or in the inflation expectations and assessments. However, there are also grounds for arguing that the annual inflation rate should be already stationary as it is the target for the ECB monetary policy and should be kept below 2 per cent within the euro area. Of course, Finnish inflation can divert for a while from the overall target, but not permanently without loss of price competiveness, for example.

Table 8 shows the estimated cointegration tests for the bivariate VAR(4) system. The optimal lag order was tested to be sufficient with 4 lags, and residuals were quite close to white noise.[2] The Johansen cointegration tests performed do

---

[1] It should be emphasised that expectations are not assumed to "cause" inflation in the philosophical sense, but just indicate inflationary pressures. The self-fulfilling nature of expectations is not considered in Granger causality tests. Cointegration was studied using the expectations at the time they were asked about. It should be also remembered that Granger causality tests are valid only for stationary time series.

[2] Only minor autocorrelation was left in the longer lags, e.g. around 12 months, hinting to overdifferencing. The autocorrelation and partial autocorrelation functions were also quite similar for the CPI inflation and for inflation expectations. The diagnostics of the system equations is satisfactory apart from the outliers indicating a level shift in the normality test due to the collapse in inflation expectations after the euro changeover in January 2002.

not reject the possibility of having a cointegration relationship between core inflation and consumers' inflation expectations. Cointegration can be found basing on both maximum eigenvalue and trace tests at the 5 % significance level. In addition, the restriction of homogeneity in the long run coefficients ($\beta_1 + \beta_2 = 0$) cannot be rejected and can, therefore, be imposed on the error-correction model specification. The error-correction term calculated from the restricted cointegration relationship was then used in the EC-model for core inflation with a month's lag, which also showed to be significant. The estimated model implies that few lagged inflation expectations can be useful in forecasting the core inflation. It should also be noted that the simultaneous inflation expectation, which is published about two weeks before actual inflation, can already be used in forecasting. The same applies to the error-correction terms, i.e. here the simple difference between inflation expectations and core inflation can be used successfully in forecasting core inflation.[1] Similarly, it should be noted that the diagnostics of the estimated error-correction model do not show any serious deficiencies apart from the outliers arising from the level shift downwards in inflation expectations starting from January 2002. The results differ considerably from those of Kenny and Forsells (2002), who found that actual inflation is not strongly influenced by expectations. Our finding is quite the opposite, indicating that inflation may even statistically 'adjust' to expectations. Of course our estimation period is short and results could be spurious and only related to the estimated period. To make further interpretations we should also have a mechanism that makes actual inflation adjust towards expectations. However, this result may also be due to the fact that expectations are usually conservative and close to the 2 per cent mean, which is currently also the ECB target for inflation. The error-correction coefficient (-0.08) does not show very rapid convergence to equilibrium.

A growing number of countries have anchored their monetary policy to a strict inflation target. This nominal anchor policy has clearly stabilised monetary conditions and probably also affected consumers' inflation expectations (Figure 12). Monetary policy cannot control future inflation with precision therefore inflation will vary around the targeted level. If a central bank does not succeed in anchoring inflation expectations, it is forced to pursue a more restrictive monetary policy to nail down the accelerating inflation that is linked to self-fulfilling expectations.

Forecasts are usually conditional expectations based on a set of information on the economic mechanism that is thought to generate the observed phenomenon. The future is rarely exactly similar to the past, which leads to forecast errors as we try to imitate the future with the observed past. Surprises will always happen, but the past and economic reasoning about motivations behind observations is the only basis for forecasting.

[1] The same testing procedure was also applied to the ordinary CPI and inflation expectations, where qualitatively similar conclusions could be made.

Hendry and Clements (2001) note that since unpredictable events happen in the future, the future will inherently be more uncertain than the past. However, our knowledge about the past is not perfect even though it has already happened. This argument applies to inflation as well. We do not know all the economic sources that generate inflation even when we have detailed information about the prices of goods and services. We cannot separate the exact demand and supply factors that affect inflation or identify accurately the economic background variables, such as labour costs, raw material or import price effects, demand pressure or other influential factors. There is natural asymmetry when we are considering inflation expectations and inflation assessment.

Table 8. Cointegration tests between core inflation and inflation expectations

Estimation period: 1996/m5 - 2002/m10, VAR(4) -model

		Max eigenvalue test		Trace test	
Ho:rank=p	Eigenvalue	-Tlog(1-\μ)	95%	-T\Sum log(.)	95%
p = 0	0.156	13.23*	11.4	14.39*	12.5
p ≤ 1	0.014	1.16	3.8	1.16	3.8

standardized \beta' eigenvectors [β_1, β_2]

CPIund	KB6
1.0000	-0.9468
-6.6610	1.0000

LR-test for homogeneity restriction [β_1, β_2] = [1, -1]
LR-test, rank=1: $\chi^2(1)$ = 0.50528 [0.4772]

standardized \alpha coefficients

CPIund	-0.13795	-0.00057
KB6	-0.08940	0.00230

System diagnostics: Test stats p-values

CPIund :AR 1- 5 F(5, 65) = 1.7262 [0.1411]
KB6 :AR 1- 5 F(5, 65) = 0.8551 [0.5160]
CPIund :Normality Chi^2(2)= 2.1329 [0.3442]
KB6 :Normality Chi^2(2)= 40.7150 [0.0000] ** (level change in January 2002)
CPIund :ARCH 5 F(5, 60) = 0.6669 [0.6500]
KB6 :ARCH 5 F(5, 60) = 0.7925 [0.5592]

The estimated error-correction model for core inflation using inflation expectations and the error-correction term as explanatory variables:

Modelling dCPIund by OLS
Estimation period: 1996/m7 - 2002/m10

Varable	Coefficient	Std.Error	t-value	PartR^2	Instab	
dKB6	0.13938	0.07887	1.767	0.042	0.07	
dKB6_4	0.15064	0.07723	1.950	0.050	0.22	
CRCPIundvec1_1	-0.07687	0.02936	-2.618	0.086	0.09	ECT-term

dKB6_8	0.20412	0.07746	2.635	0.088	0.22

$R^2 = 0.273$ \σ = 0.155289 DW = 1.98
Instability tests, variance: 0.431 joint: 0.858

Regression diagnostics: Test p-values Test interpretation

Test		p-values	Test interpretation
AR 1- 5 F(5, 67) =	2.2826	[0.0561]	Autocorrelation
ARCH 5 F(5, 62) =	0.7808	[0.5673]	ARCH-heteroscedasticity
Normality $\chi^2(2)$ =	2.5843	[0.2747]	Outliers
Xi^2 F(8,63) =	0.6310	[0.7486]	Heteroscedasticity
Xi*Xj F(14, 57) =	0.5774	[0.8714]	Functional form
RESET F(1, 71) =	1.7972	[0.1843]	Functional form

7. Using distribution moments in forecasting inflation

Before we carry out a statistical analysis of the cross-sectional data we will pay attention to the so-called outliers, as otherwise the results with respect to the mean and other distribution moments would be misleading. A reasonable amount of individual responses were clearly out of any rational range. For the Consumer Survey statistics all observations that are below -15 or exceed +15 are removed at Statistics Finland. This corresponds to about 3 times standard deviation, where the standard deviation is determined from monthly deviations without a significant amount of outliers. These outliers are removed from the data along with the "Don't know" answers.

Examining the monthly distribution parameters reveals that the monthly distributions are commonly clearly asymmetric and gradually changing over time. Skewness and, in particular, kurtosis differ significantly from those of a normal distribution and are time-varying (Figure 13). Thus, the null-hypothesis of normal distribution is rejected in formal tests. It would be interesting to know whether the changes in the distribution moments calculated from the monthly surveys tell us something about future inflation. For instance, it would be plausible to assume that the skewness or kurtosis of inflation expectations could carry information concerning future inflation, e.g. if the kurtosis of inflation expectations increases then inflation expectations are more concentrated to the mean and actual inflation would be slowing down. On the other hand, if skewness increases this is usually related to thicker right hand tail in the distribution of expectations and to a greater number of observed accelerating commodity prices, therefore this may indicate increasing inflation expectations and inflation risk. A related idea was utilised by Suvanto and Hukkinen (2001) in using the skewness measure of the commodity price change distribution for indicating inflation pressures.

The monthly distributional parameters clearly show that the distributions of inflation expectations are asymmetric. The monthly distributions are clearly skewed to the right and are more concentrated than normal distributions. Kurtosis is also the most volatile of the distribution moments.

In Table 9, we test the dynamic relationship between inflation and expectations. The expectations have been shifted backwards to the point when they were made and 6 dynamic lags were then used to calculate the results. We first notice that the mean is significant for the CPI and Harmonised CPI, but not for the core or median inflation, which is not something we anticipated. However, for the median and core inflations, volatility (standard deviation) and skewness got significant coefficients, which did not appear for the CPI or HICP. The kurtosis did not seem to be helpful, either, for the forecasting of inflation measures. Extending the lag structure to 12 lags did not make the other distribution moments significant either, so the conclusion is that distribution moments apart from the mean do not seem to be powerful in forecasting inflation. Establishment of the monetary union and the subsequent convergence in inflation rates must have also influenced the Finnish consumers' expectations on inflation and interest rates. As we noticed above, the diminishing uncertainty concerning inflation should have reduced the variance of past inflation estimates and partly of inflation expectations as well. This development may also reflect growing credibility of the inflation target.

An interesting feature in inflation assessments and expectations is their slight "asymmetry". Already at the beginning August 1996 consumers estimated that consumer prices had fallen from the previous year but, on the other hand, believed the prices would rise in the future. However, the assessments and expectations converge toward the end of 1996. Consumers' valuations of the decline in prices were probably caused by the slight deflation at the end of 1995 and by the fall in food prices that followed EU membership. Thus, we can think that joining the EU distorted the consumers' views on the decline of prices. When people realised that the prices had not fallen by as much as the media told them, for example, they changed their assessments of past price changes to the positive. However, the membership did not affect the expectations as strongly as it did the assessments, because the expectations are forward looking. When the inflation assessments changed to the positive, the standard deviations of the expectations and assessments also converged.

Another interesting feature in the euro changeover was that assessments and expectations diverted so clearly from January 2002 onwards and only partial convergence has appeared so far since (remember Figure 4). One thing that happened in connection with the euro changeover was a large increase in the kurtosis of inflation expectations, although at the same time the mean of expectations collapsed. In fact, if the rise in the assessments were supposed to be based on a one-off upward change in prices due to the change of currency then according to expectations announced in January 2002 inflation assessments January 2003 should have collapsed. However, this did not happen. Therefore, we have a problem in interpreting these responses. Since the changeover consumers have been told widely by the press that inflation did not accelerate significantly during the changeover.

Crevits, Vanhaelen and Vonck (2002) identified two main sources from the euro changeover that affect inflation, namely the costs and benefits related direct to the change of currency, such as cash handling costs, coin automatic costs and the like. The other sources were more psychological, influencing through rounding effects, usually upwards, and through delayed price increases that had been waiting for the right moment. The bias observed in the discrepancy between actual inflation and inflation assessments would suggest that consumers' beliefs may persist in making systematic forecasting errors while observing inflation. This would be quite difficult to interpret. The idea of an adaptive expectations mechanism would impose that consumers learn from their previous forecast errors. Of course, consumers do have serious problems in gathering and learning price information about goods and services, but it would be interesting to see whether a systematic bias could persist for a long time.

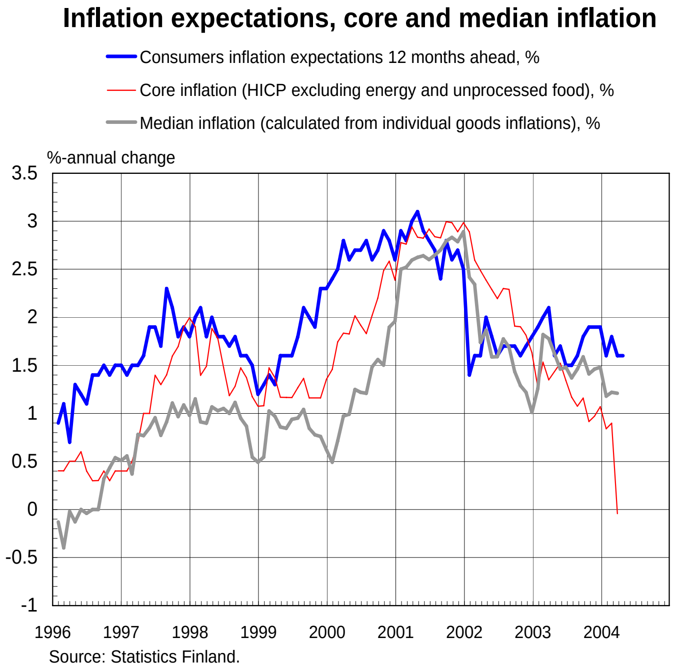


Figure 11.

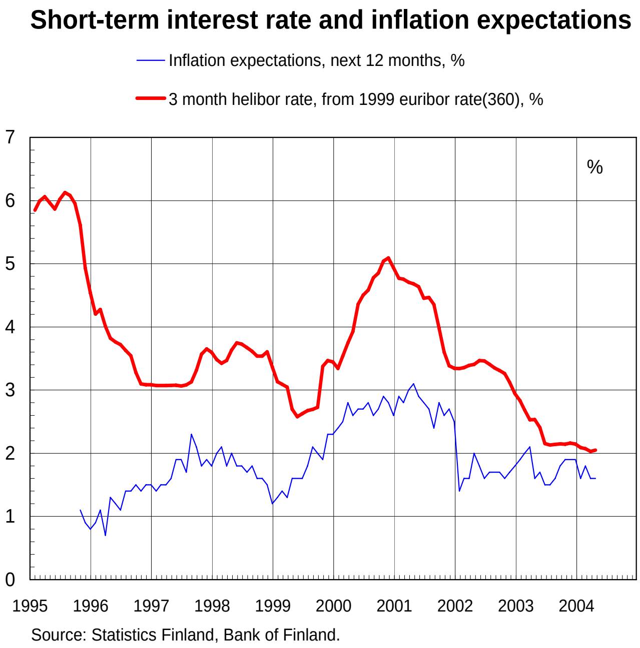


Figure 12.

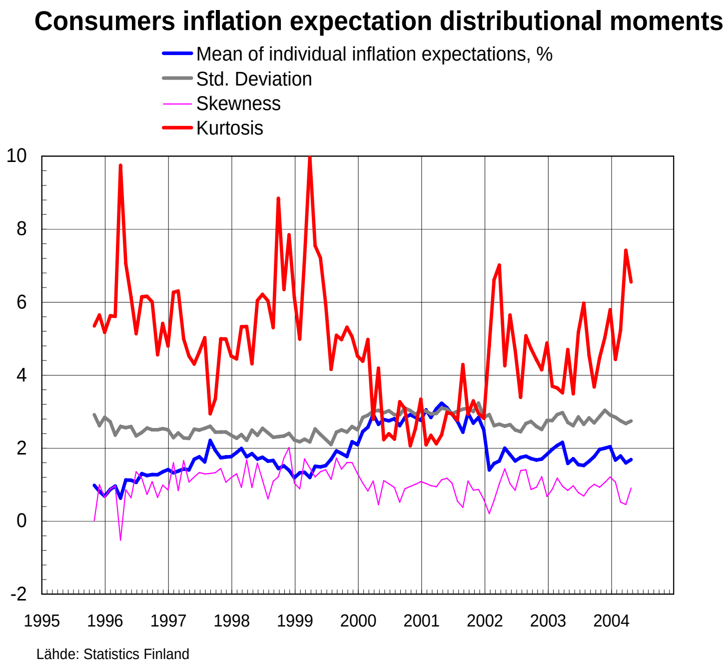


Figure 13.

Table 9. Dynamic regression on inflation expectations moments

Testing predictive power of consumers inflation expectation moments
Estimation period: 1996/m4 - 2002/m12

Dynamic regressions with 6 lags of each expectation moment
Solved static long-run equations (t-values in parenthesis)

Dependent variable	Mean	Std. dev.	Skewness	Kurtosis	R^2	DW	Autocorr. test p-values AR(5)
CPIg	1.38	-0.29	-0.02	-0.05	0.978	0.615	.0000**
	(5.28)	(1.13)	(0.07)	(0.59)			
HICPg	0.96	0.25	-0.53	-0.02	0.979	0.687	.0000**
	(3.51)	(0.93)	(1.85)	(0.27)			
CPIund	0.88	0.04	0.59	-0.15	0.981	0.604	.0000**
	(3.65)	(0.15)	(2.35)	(2.05)			
CPImed	0.89	-0.03	0.17	-0.11	0.959	0.494	.0000**
	(3.04)	(-0.10)	(0.56)	(1.27)			

Variables:
CPIg = CPI inflation, %
HICPg = Harmonised CPI inflation, %
CPIund = Underlying (core) inflation, %
CPImed = Median CPI inflation, %

Mean = 1. moment of consumers' inflation expectations
Stdev = 2. moment of consumers' inflation expectations
Skewness = 3. moment of consumers' inflation expectations
Kurtosis = 4. moment of consumers' inflation expectations

8. Conditioning expectations

Last, we will consider ways for improving or, in fact, distilling the information contained in inflation expectations. Simply by looking at the individual responses to the direct questions concerning inflation percentages, it becomes rapidly clear that a significant number of the responses are either erroneous or nonsense.[1] While recording these obviously erroneous responses as outliers, we could also combine separate answers to improve the quality of the inflation expectations responses. First, inflation expectations below -15 per cent or observations above 15 per cent have been eliminated from the analysis. In addition, it would be useful to filter the answers concerning inflation expectations somehow.

It is probably useful to think that consumers can be separated into two groups basing their inflation expectations on different facts. A significant share of consumers may be backward-looking, basing their expectations purely on the recently observed inflation. The other group may have some rational arguments about how future inflation may evolve. The only problem is how to estimate the sizes of these groups. Without further assumptions we cannot gain any estimate of

[1] Even allowing for the fact that all persons have separate, individual commodity baskets, responses putting the inflation rate at 50 or 100 per cent seem quite absurd.

the sizes, so we may try to approach this question with a few specific assumptions about expectation formation.

The idea was to choose the inflation expectations of those individuals who knew at least approximately the current level of inflation. Basing on the individual questionnaire we selected only those individuals into our sample who knew the current inflation rate correctly within the ± 2 per cent range (assessed inflation not deviating from the CPI inflation more than 2 percent). Inflation expectations were calculated within this group of "informed" consumers. The measures of the unconditional and "conditional" inflation expectations are compared in Figure 14. At first glance, the time series properties of these two inflation expectation measures seem quite similar apart from the mean of expectations, which is clearly lower for those having known the current level of inflation. In fact, it turned out that during the sample period the mean of "conditional" inflation expectations was closer to actual inflation than was the mean the unconditional inflation expectations. The mean of the CPI inflation during 1995/m10 to 2002/m8 was 1.6%, which was exactly the same for "conditional" inflation expectations, but 1.97% for the unconditional expectations.[1] The difference in the frequency distribution can be seen more clearly from the histograms plotted in Figure 15.

This result also held true when lagged inflation expectations were compared with the actual CPI inflation. Even though information is limited because of our short sample, it seems that the quality of inflation expectations can be improved by selecting the answers of "informed consumers". However, other time series analyses, such as accuracy regressions and cross-correlation functions, did not reveal major differences with the earlier results. It should be also noted that the euro changeover was quite similar to all consumers and to the informed consumers, as the drop in inflation expectations was proportional for both groups in January 2002.

[1] The standard deviations were smaller for both unconditional and conditional inflation expectations than the actual CPI inflation.

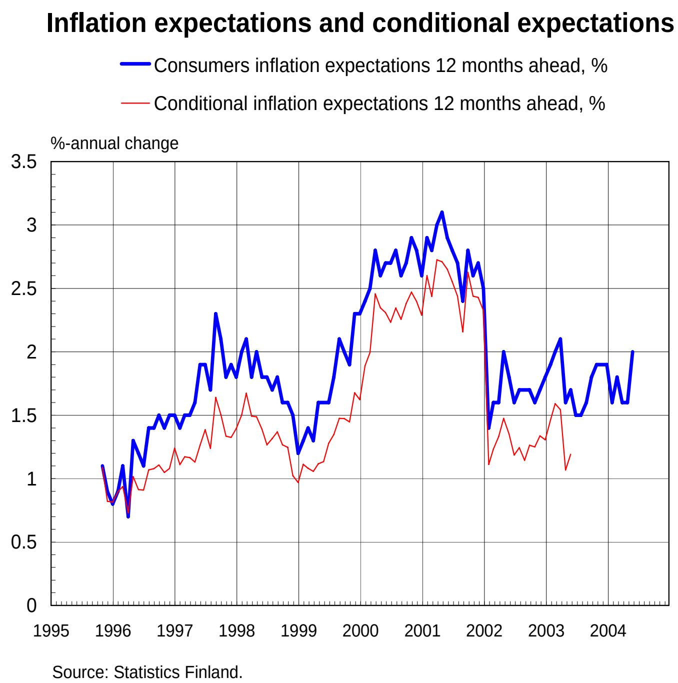


Figure 14.

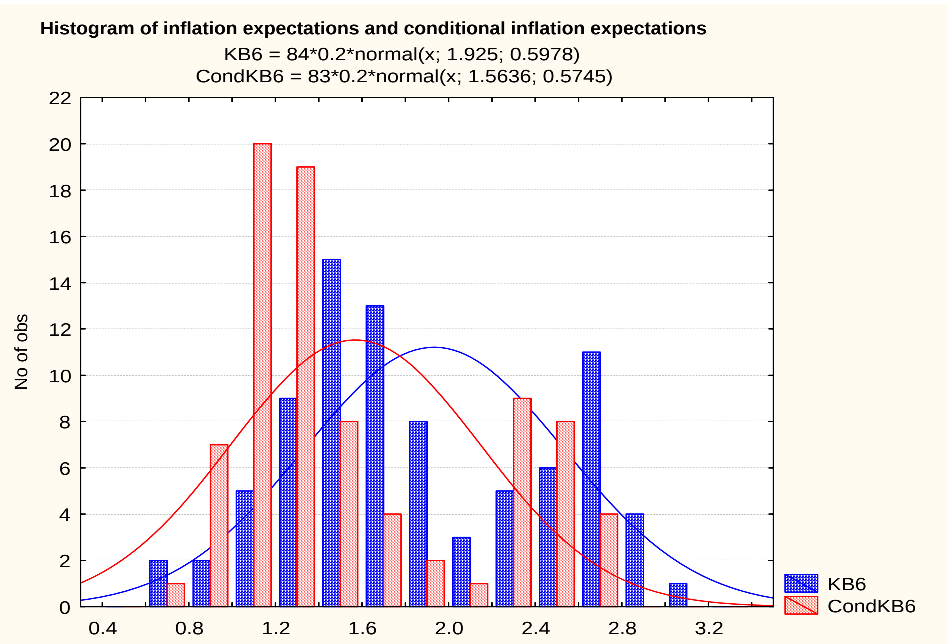


Figure 15.

9. Conclusions

Statistics Finland has asked direct questions about past and future inflation in its Consumer Barometer Survey since October 1995. Basing on the statistical analysis it seems clear that inflation forecasting can be aided by using consumers' per cent inflation expectations, not as direct point estimates but to rather give indication about the overall inflation tendency one year or so ahead. The analysis also showed that assessments about current inflation and inflation expectations cannot be regarded as simple linear combinations of the commodity group prices as used in the CPI weights. We found that certain commodity groups, such as food, health care, housing and restaurant services received over-weighted importance in the formulation of inflation assessments and expectations. Inflation expectations usually measure more closely such inflation measures as core and median inflation that do not contain volatile and difficult components like prices of oil or unprocessed food. Although inflation assessments and expectations are based on very similar background information, there is clear asymmetry between them and inflation expectations have predictive power over inflation as a process.

One very interesting episode has been the euro changeover. It had a quite dramatic effect on both consumers' inflation assessments and expectations. In Finland, and many other euro countries, inflation assessments rose rapidly in January 2002 and at the same time inflation expectations collapsed. Logically, this means that prices were perceived to rise, but this rise was felt to be a once-for-all rise, which would not be repeated within the forthcoming 12 months. There is evidence that services and menu-list prices were raised markedly, but the overall effect from the euro changeover estimated by Statistics Finland was 0.2 percentage points instead of the close on 1 percentage point estimated by consumers. The sensitivity of consumers to certain goods prices may partly explain the reaction, but something also seems to remain unexplained because the high level has prevailed in the assessments although low inflation figures have since then been published for almost a year. However, if consumers had been rational in their assessments and expectations, inflation assessments would have collapsed at the latest in January 2003 after 12 months from the euro changeover. This did not happen, however, and assessments have continued to stay at around the 3 per cent level, while inflation expectations have remained close to the actual inflation of below 2 per cent.

Causality tests and VAR models also prove the usefulness of inflation expectations in forecasting future inflation within a one-year horizon, but it was equally clear that inflation expectations are not sufficient for producing accurate forecasts.

One novelty in this paper was to look at the predictive power of the distribution moments of inflation expectations. Other studies have shown that e.g. skewness in the CPI commodity distribution could help forecasting inflation. However, the results here were not conclusive for any other distribution moments than the mean so as to be of any systematic use for inflation forecasting. There was slight evidence that the skewness of expectations could predict the CPI inflation, but the sign of this coefficient was negative and, therefore, not as expected. Another way for improving the information content of inflation expectations we considered was to condition expectations so that only the expectations of those consumers who knew approximately the current level of inflation would be registered. This fine tuning of inflation expectations brought at least the mean of expectations closer to the levels of the current and future inflation rates (ex post). Otherwise, there was not much difference in the linear efficiency and unbiasedness tests performed for the "conditional expectations". Restricting inflation expectations to those of "informed consumers" only did not help to resolve the mystery of the euro changeover, either.

REFERENCES

BENIGNO P. and LOPEZ-SALIDO J.D. (2002). Inflation Persistence and Optimal Monetary Policy in the Euro Area, ECB, Working Paper No. 178, September 2002.

BERG L. and BERGSTRÖM R. (1996). Consumer Confidence and Consumption in Sweden, Department of Economics, *Uppsala University Working Paper*, 7/1996.

BUNDESBANK (2001). The Information Content of Survey data on Expected Price Developments for Monetary Policy, Deutsche Bundesbank Monthly Report, January 2001, 35-49.

CREVITZ P., VANHAELEN J-J. and VONCK J-P. (2002). The Euro Survey in Belgium: An Efficient Instrument for Monitorin the Preparations for and the Consequences of the Changeover to the Euro, 26^{th} CIRET Conference, Taipei, October 2002.

D' ELIA E. (2003). Inflation perceived by consumers in the euro-zone and proxy exchange rates against euro, Paper presented in Eurostat meeting of the Working Party 'Harmonization of Consumer Price Indices' in Luxembourg, March 2003.

DJERF K. and TAKALA K. (1997). "Macroeconomy and Consumer Sentiment: Performance of the Finnish Consumer Barometer after Ten Years", Bank of Finland Discussion Papers 20/1997, Helsinki. Also published in Oppenländer K.H., Poser G. and Weller S. (eds.) *Selected Papers Submitted to the 23^{rd} CIRET Conference in Helsinki*, Aldershot: Ashgate.

EUROPEAN COMMISSION (1997). European Economy, The Joint Harmonised EU Programme of Business and Consumer Surveys, Reports and Studies No 6, Brussels.

FORSELLS M. and KENNY G. (2002). Survey Expectations, Rationality and the Dynamics of Euro Area Inflation, 26th CIRET Conference, Taipei, October 2002.

FORSELLS M. and KENNY G. (2002). The Rationality of Consumers' Inflation Expectations: Survey Based Evidence for the Euro-Area, European Central Bank, Working Papers No. 163, August 2002.

HAMILTON J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

HENDRY D.F. and CLEMENTS M.P. (2001). Economic Forecasting: Some Lessons from Recent Research, ECB Working Paper Series, WP No. 82, October 2001.

HUKKINEN J. and SUVANTO A. (2001). Price Stability and Changing Prices, Bank of Finland Bulletin, 4/2001.

JOYCE M. (1997). Inflation and inflation uncertainty, Bank of England, Quarterly Bulletin 37.

KANGASSALO P. and HEISKANEN M. (2001). Testing the Mode Effects in the Finnish Consumer Survey, *Proceedings of Statistics Canada Symposium 2001*, Ottawa.

MANKIW G. and REIS R. (2002). What Measure of Inflation should a Central Bank Target?, ECB, Working Paper No. 170, August 2002.

MINCER J. and ZARNOWITZ V. (1969). Evaluation of Economic Forecast, in Economic Forecasts and Expectations ed. by J. Mincer, New York.

PALOVIITA M. (2002). Inflation Dynamics in the Euro Area and the Role of Expectations, Bank of Finland Discussion Papers, 20/2002.

PURSIAINEN J. (1999). Consumers inflation expectations and monetary policy (in Finnish only), Bank of Finland Economics Department Working Papers, 1999.

RICH R. and TRACY J. (2003). Modeling Uncertainty: Predictive Accuracy as a Proxy for Predictive Confidence, Federal Reserve Bank of New York Staff Reports, No. 161, February 2003.

STATISTICS FINLAND (2000-2001). Consumer Survey publications and working papers. (Mostly in Finnish)

SUVANTO A. and HUKKINEN J. (2002). Stable Price Level and Changing Prices, Manuscript, Bank of Finland, October 2002.

VEGA J-L. and WYNNE M.A. (2001). An Evaluation of Some Measures of Core Inflation for the Euro Area, ECB Working Paper No. 53, April 2001.

APPENDIX 1.

Variable	Descriptive Statistics 1996/m1 - 2002/m12								
	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Standard Error	Skewness	Kurtosis
CPIg	84	1.650	1.558	0.32	3.691	0.88	0.10	0.55	-0.66
KB6	84	1.952	1.800	0.70	3.100	0.57	0.06	0.30	-0.88
KB5	84	1.699	1.700	-1.40	3.400	1.11	0.12	-0.74	0.20
HICPg	84	1.796	1.691	0.40	3.445	0.80	0.09	0.36	-0.95
CPIund	84	1.640	1.544	0.30	2.996	0.80	0.09	0.04	-0.94
CPImed	84	1.193	0.995	-0.40	2.891	0.81	0.09	0.54	-0.33

APPENDIX 2.

Unit root tests for inflation and inflation expectations

Estimation period: 1996/m1 - 2002/m9

Null hypothesis: Variable has a unit root

Variable	ADF	P-value	Phillips-Perron	P-value
CPIg	-2.79	0.0616	-1.86	0.3516
HICPg	-2.12	0.2379	-2.07	0.2552
CPImed	-1.65	0.4541	-1.66	0.4456
CPIund	-1.54	0.5067	-1.55	0.5059
KB5	-2.12	0.2363	-1.66	0.4464
KB6	-2.06	0.2614	-1.83	0.3639
Critical values and McKinnon one-sided p-values				
5 %	-2.88		-2.89	
1 %	-3.48		-3.50	

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 365—390

AN ANALYSIS OF INFLATION EXPECTATIONS OF THE TURKISH PRIVATE MANUFACTURING INDUSTRY

Ercan Karadaş, Fethi Öğünç[1]

ABSTRACT

The main purpose of this paper is to make a detailed analysis of the quantified expected inflation series obtained from the Business Tendency Survey conducted by the Central Bank of the Republic of Turkey. Different representations of simple expectation formation mechanisms that appeared in the literature are discussed in the context of Turkey. In selecting the most appropriate mechanism, instead of an *ad hoc* and an arbitrary methodology, nested hypotheses and the sequential testing procedure are applied. A general first order extrapolative mechanism is found as a result of this procedure. But, the selected mechanism is far from being satisfactory. Therefore, a new augmented model is proposed. Finally, taking into account problems regarding the tests of rationality, efficiency and orthogonality tests are carried out. As a result of these tests, rationality assumption cannot be rejected very powerfully. However, this is not stemmed from the accuracy of the expectations, but from the structure of the economy that precludes accurate short-term forecasting.
JEL Classification Numbers: C19, E31, D84

Key words: Quantification; Sequential testing procedure; Formation of inflation expectations; Rationality.

1. Introduction

Although, expectations have begun to play a key role in a wide variety of economic models beginning from the early 1930s, it was not modeled explicitly until the 1950s. Adaptive expectations hypothesis (AEH), introduced by Cagan

[1] Statistics Department, Central Bank of the Republic of Turkey, Istiklal cad. 10 Ulus, 06100 Ankara, Turkey. e-mail: Ercan.Karadas@tcmb.gov.tr, Fethi.Ogunc@tcmb.gov.tr The views expressed in this paper belong to the authors and do not represent the views of the Central Bank of the Republic of Turkey. We would like to thank Prof. Ercan Uygur for his helpful comments and suggestions.

(1956) was the first attempt to form an endogenous expectations formation mechanism. Here, expectations formation mechanism is defined as the specification of a rule by which individuals revise their expectations in the light of new information. In the 1960s and 1970s, AEH played a prominent role in macroeconomics. Then, after the papers of Lucas (1972) and Sargent (1973), rational expectations hypothesis (REH) influenced every branch of macroeconomics deeply and today it is nearly the standard approach to the expectations formation. Since policy implications of macroeconomic models depend on to a large extent assumptions about the expectations formation, it is important to test for the validity of these two hypotheses.

In the literature, quantified expected inflation series have been used extensively[1] following Carlson and Parkin's (1975) seminal work in testing which hypothesis describes the real data best. Following this conventional methodology, this paper aims at a detailed discussion about the validity of simple expectations formation mechanisms and REH utilizing quantified inflation series obtained from the Business Tendency Survey (BTS) conducted by the Central Bank of the Republic of Turkey (CBRT). In the post 2000 period, Turkish economy has experienced two major policy changes. Therefore, the paper also discusses some limitations resulting from these policy changes in testing validity of simple expectations formation mechanisms and REH.

The remainder of this paper is organized as follows: In Section II, data is introduced and methodological problems resulting from conversion of qualitative responses into quantitative measures of inflation expectations are considered. In the following section, simple models of expectations formation are tested using sequential testing procedure, and then a more satisfactory model that is not just related with past history of inflation or inflation expectations is proposed. In Section IV, basic postulates of rational expectationss hypothesis are discussed and some tests of REH are carried out. The last section summarizes the results.

2. Data

Beginning from the last month of 1987, the BTS has been conducted monthly by the CBRT. In BTS, various questions related with orders, stocks, costs, production, selling prices, inflation and interest rates are designed to allow respondents to evaluate past a few months (generally three), and to express their expectations over the next three months. Questions are generally of qualitative type, and respondents are asked to answer in one of "rise", "the same" or "fall" form. Besides qualitative questions, there are five questions asking for an ordering of several factors, and four asking for a quantitative value of credit interest rates and inflation expectations. The sample is chosen on the basis of Istanbul Chamber of Industry's ranking of the biggest 1000 firms. Since respondents are from

[1] For a brief literature review, see Grand and Thomas 1999.

private sector covering all the major manufacturing sub-sectors, in the study the inflation expectations is referred, hereafter, to the inflation expectations of private manufacturing industry.

The BTS consists of three questions on inflation expectations, one for the next three month's inflation expectations in standard qualitative form and the other two for quantitative estimates of the year-end and the next twelve months' inflation expectations respectively. Although, working with direct inflation expectations is very advantageous, as we shall see, we are limited with insufficient number of observations in using these three series, since the first one was added to the survey in May 1997, and the others were in January 1999. Therefore, we are only left with the question asking average selling price expectations of each respondent's product basket over the next three months[1] as an estimate of his inflation expectation.

Sampling frequency (monthly) of the question is different from the forecast horizon (three months), therefore using monthly data causes serial correlation in the error terms. Formally, let ${}_{t-3}w_t$ be the forecast error made at the end of $(t-3)$ for the end of next three months (t). Under the null hypothesis that the survey respondents' forecasts are unbiased and efficient, ${}_{t-3}w_t$ can be decomposed into unobserved monthly components:

$$ {}_{t-3}w_t = {}_{t-3}w_t^m + {}_{t-3}w_{t-1}^m + {}_{t-3}w_{t-2}^m \qquad (1) $$

where ${}_{t-3}w_{t-i}^m$ is the unobserved forecast error made at the end of the month $(t-3)$ for the month $(t-i)$, $(i = 0,1,2)$. Under the null hypothesis stated above, only ${}_{t-3}w_{t-2}^m$ is likely to be a white noise process, and the remaining two terms in (1) cause serial correlation.

In order to solve the overlapping intervals problem, we have ignored overlapping months from the sample, similar to Bryan and Gavin (1986), so that consistency has maintained between the sample frequency and the forecast horizon[2].

Although selected quantification method may have important effects on the results concerning expectations formation mechanisms and rationality, we have not expanded this discussion here since it has already been discussed in detail in Oral (2002), Ogunc et. al (2004), and Uygur (1989). At this stage, we use the

---

[1] Original form of the question is: (Trend of the next 3 months) Average price for the new orders received from the domestic market (excluding seasonal variations).

[2] Hansen (1982) proposes another procedure to avoid the problem in cases for which the forecast horizon is longer than the sample frequency. The methodology involves using OLS but making appropriate modifications in the estimation of the asymptotic covariance matrix. Though this methodology is superior in terms of gaining degree of freedom, we have preferred to drop the overlapping months from the sample, which is intuitively simpler.

Carlson and Parkin's (CP) methodology in order to obtain a time series of the expected inflation rates (P_t^e). In this methodology, it is assumed that firms do not expect any change in their product prices if their actual price changes lie within a certain interval ($-\delta$, δ), which is called as indifference interval. Then, the conversion is based on the subjective probability approach that can be summarized in the following equations.

$$P_t^e = \hat{\delta}\left(\frac{a_t + b_t}{b_t - a_t}\right) \tag{2}$$

where $\Phi(a_t) = 1 - A_t$, $\Phi(b_t) = B_t$ and $\hat{\delta} = \sum_{t=1}^{T} P_t \Big/ \sum_{t=1}^{T}\left(\frac{a_t + b_t}{b_t - a_t}\right)$

The percentage of the firms that expect a rise and a fall in their prices for the next three months is denoted by A_t and B_t respectively[1]. $\Phi(\cdot)$ refers to the cumulative distribution function for the selected form of distribution of expected inflation across the population. Finally, a_t and b_t can be derived from the inverse cumulative distribution function of the selected distribution. We have used three different statistical distributions: Uniform, Normal and Logistic. Though all distributions have given good results, the uniform distribution has yielded the best result in the sense that contemporaneous correlation with the actual inflation rates is the highest. Therefore, in the rest of the study quantified expectations series obtained by this method is employed.

Additionally, we have relaxed the CP's restrictive assumption that indifference interval is symmetric and constant over the whole period, so that the upper and lower boundaries of the indifference interval are allowed to vary over time ($\delta_{m,t}$, $\delta_{p,t}$). Asymmetric and time varying estimate of the indifference interval is achieved via adaptive Kalman filter (see Appendix B.1)[2]. As a result, two expectations series are derived, one from the original CP methodology with constant and symmetric indifference interval ($P_t^{e(c)}$) and the other from the extension of it that permits time varying and asymmetric indifference intervals ($P_t^{e(v)}$) (see Table A.1 in the Appendix A.1). In the rest of the study, in order to see how sensitive the results concerning expectations formation and rationality are to the selected quantification method, we employed both $P_t^{e(c)}$ and $P_t^{e(v)}$

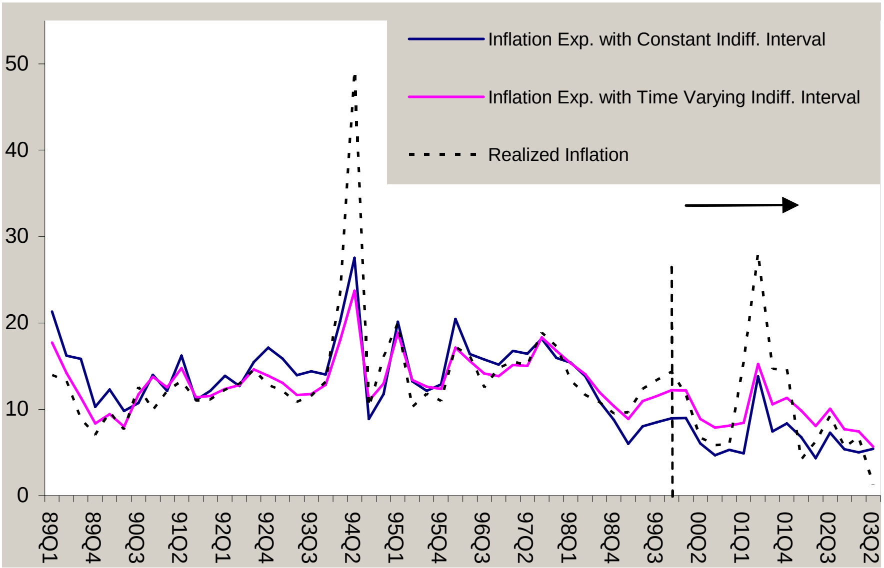
[1] In this formula, it is assumed that the upper and lower boundaries of the indifference interval have the same absolute value. For the asymmetric indifference interval case, the formula can be rewritten as $P_t^e = \hat{\delta}_m\left(\frac{a_t}{a_t - b_t}\right) + \hat{\delta}_p\left(\frac{-b_t}{a_t - b_t}\right)$

[2] See Seitz (1988), Smith and McAleer (1995) and Nardo (2003), for the use of asymmetric and time varying estimate of the indifference intervals.

series, but we have only reported test results obtained from $P_t^{e(c)}$ unless results did not indicate a significant difference.

We have omitted a few observations from beginning period of the survey in order to prevent any bias that may come from misunderstanding of the question or the forecast horizon by the respondents, late coming and similar issues. We have also excluded the period after the disinflation program of the year 2000. In order to explain the reasoning behind this decision, we need to make an early discussion of the logic behind the simple expectations formation mechanisms and REH. Firstly, when we consider the definition of expectations formation mechanism given in the introduction, it is clear that agents need more or less a stable economic environment in order to be able to form a specification of a rule by which they can revise their expectations in the light of the new information. However, the period we have excluded contains two-sub periods that represent different major policy changes. The first sub-period excluded is characterized by crawling peg exchange rate regime ends at February 2001, and the second sub-period -after February 2001- covers the floating exchange rate and implicit inflation-targeting regime. Although, econometrically it is hard to describe expectations formation mechanisms of these two sub-periods due to limited number of observations, we expect that the rule agents used in their expectations formation differentiates significantly. Therefore, taking the period 1989:Q4-2003:Q2 as a whole will probably lead to get a quite blur picture of the expectations formation. This may also be problematic in terms of rationality tests. In the test of rationality, unless you have a good mathematical representation of the true inflation generating process, which means that agents could learn the structure of the model properly, the power of your rationality test is low. It should be emphasized also that the REH assumes that agents converge to the true model by time with no systematic errors. But for the reasons mentioned above, it is difficult to give a good mathematical representation of the inflation generating process for the post 2000 period, so a probable test of the rationality may not be very meaningful for this period. As a result, we have worked with the period 1989:Q4-1999:Q4. In this period, both $P_t^{e(c)}$ and $P_t^{e(v)}$ are stationary at one percent of significance level according to ADF test results (Appendix C.1).

Figure I: Inflation Expectations and Realized Inflation



When we look at the expected and realized inflation rates (Figure I), they move generally together. However, as seen from the graph, $P_t^{e(c)}$ series are mostly over the realized values. A possible explanation can be the following. A significant underestimation of inflation in the second quarter of 1994[1] may result in higher estimations in other periods since the scaling parameter is determined with the assumption of equal change in inflation expectations and actual inflation over the sample period. Less likely, the respondents may be generally pessimistic about short-term inflation expectations.

3. Sequential Testing of General Models of Expectations Formation

Simple models of expectations formation can be broadly classified as "extrapolative" and "adaptive" expectations. The first one is based on the idea that variations in expected inflation rate are simply explained by the past rates of inflation. On the other hand, under the latter, expectations are assumed to follow an error-learning process, which means that they adjust in proportion to the recorded errors. Although, these models are not very realistic in the sense that they do not take into account any variables other than past history of inflation and inflation expectations, they still need to be investigated in order to see whether

[1] In the second quarter of the 1994, Turkey experienced a deep financial crisis after which year-end wholesale price index rose by 149 percent, and GNP declined by 6 percent.

people think in a simple manner as assumed by the models. Therefore, in this part of the paper we will concentrate on various representations of simple expectations formation processes, and try to see how well they perform empirically.

In the literature different *ad hoc* criteria, such as goodness of fit, economic sensibility or generality of the model are taken into the consideration in deciding on which model is better among the alternatives. However, as Mills (1981) states "*ad hoc* and arbitrary methodology for model selection and hypothesis testing is inadequate for obtaining the most appropriate mechanism generating the derived inflation series". Following his suggestion, we applied the sequential testing procedure that is originally proposed by Mizon (1977).

This procedure starts with a number of different models of extrapolative and adaptive mechanisms as shown in Table A.2 (see Appendix A.1). In the table, from F(1) to F(8) and F(9) to F(18) are different extrapolative and adaptive mechanisms, ordered from simple to general, which can be nested in F(8) and F(16) respectively. In the table, the most general expectations formation mechanism is F(19) within which, all the other can be nested. Since there is no unique ordering from F(1) to F(19), a complete list of nests is given in Table A.3 (see Appendix A.1). The procedure starts with the most general mechanism (unrestricted) of each nest. It is tested against the second most general (restricted) one by putting necessary restrictions that equate two mechanisms to each other. If the restrictions are accepted jointly, it means that the most general mechanism can be reduced into the second general one without loosing much information. Then this procedure continues until restrictions put on a more general one are rejected, and the mechanism that cannot be reduced into a less general one is selected from each nest. As a test statistics the usual large sample likelihood ratio (LR) test, $T\ln(SSR_{rest.}/SSR_{unrest.})$, is used. This test statistics is asymptotically distributed as χ^2 with r degrees of freedom where r is the number of restrictions imposed. Level of significance for each test is 0.025, and this implies that overall significance level for each sequence is below 0.10, since there are at most three tests in each sequence.

Results of the sequential testing procedure can be found in Table A.3. Among the extrapolative nests (F8) is selected from Nest1 (N1), N2, and N3, and (F4) from N4 and N5. On the other hand, (F16) is selected in all adaptive nests except for N8 and N9. Among these nests, (F12) and (F11) are selected mechanisms. Therefore, we are left with five mechanisms, namely (F4) and (F8) from the extrapolative nests, (F11), (F12) and (F16) from the adaptive nests. However, (F8) can be reduced into (F4) in N4 and N5, therefore we choose (F4) among the extrapolative mechanisms. Similarly, (F16) can be reduced into (F12) and (F12) can be reduced into (F11) in N9. As a result, two non-nested mechanisms, (F4) and (F11), remained. Estimation results of these mechanisms are:

$$P_{t+1}^{e(c)} = 0.20 - 0.25P_t - 0.26P_{t-1} + 0.74\hat{w}_{t-1} \quad \text{(F4)} \tag{3}$$

(7.01) (-2.69) (-2.91) (5.76)

R^2_{adj} =0.23, SBC=-3.58, WH=0.27, LM(1)=0.57, LM(2)=0.47, LM(3)=0.65, LM(4)=0.76, ARCH LM(1)=0.08
ARCH LM(2)=0.12, ARCH LM(3)=0.21, ARCH LM(4)=0.36

$$P_{t+1}^{e(c)} = 0.08 + 0.65P_t^e - 0.21P_t \quad \text{(F11)} \tag{4}$$

(3.41) (2.85) (-1.51)

R^2_{adj} =0.15, SBC=-3.54, WH=0.15, LM(1)=0.88, LM(2)=0.84, LM(3)=0.55, LM(4)=0.68, ARCH LM(1)=0.13, ARCH LM(2)=0.25, ARCH LM(3)=0.21, ARCH LM(4)=0.33

Note: t-statistics in parenthesis and p-values for the diagnostic tests.

At this point it should be noted that coefficient of the realized inflation in (F11) is insignificant. If we drop this variable from the equation, then coefficient of the expected inflation will be 0.39. Now, since (F4) and (F11) are non-nested, an additional decision criterion is needed. Here, we have used both J-test proposed by Davidson and MacKinnon (1993) and SBC values. According to J-test, we reject both specifications, against the alternatives, suggesting that another mechanism is needed. According to SBC values, we have chosen (F4), general first order extrapolative mechanism, as a result of the sequential testing procedure.

This result might be justified in a number of ways. Firstly, in this kind of qualitative surveys respondent never knows his prediction error, since he just says a direction for inflation, but not a value that can enable him to compare his prediction with a realized value later. This may partly explains why AEH failed here. Secondly, in an environment that is highly volatile, it may be more practical to say just past inflation rates as inflation expectations, instead of computing prediction error, and trying to correct himself in the next period.

When we look at (F4), all the coefficients are significant, but those of the lagged inflation rates enter the equation with an unexpected sign, which is difficult to explain. Constant term is highly significant and indicating inflation expectations start with 20 percent, and then adjusted downward in proportion to the first and second lagged values of inflation. Note that there is also an adjustment of 74 percent for the error in predicting the previous quarterly change.

The well-known criticism that these simple expectations formation mechanisms do not take into account any information other than past history of inflation can be restated here. In order to overcome this shortcoming, we move on to an augmented model that tries to explain inflation expectations at a more satisfactorily level by employing some other economic variables.

Our augmented model of expectations formation is based on a few quite simple principles. Firstly, we do not think that it is very realistic to include variables that computationally burdening or costly available to the respondents. Secondly, we have returned to the original question and then tried to determine which factors can influence the respondent's perception about next three months' selling prices. The possible candidates are demand and cost conditions plus an uncertainty parameter. Thirdly, and more importantly, based on the estimation results, it is observed that in shaping their expectations, people do not go far away in the time span. Therefore, in defining the lag structure, we have preferred the most recent available variables, which probably affect the respondents at the date of survey filling. Finally, here we have tried to investigate which factors might contribute to explain the inflation expectations, but not necessarily the inflation itself; therefore the proposed augmented model may or may not be successful in explaining the actual inflation.

During the sequential testing procedure we had observed that the mechanisms containing prediction errors do not explain expectations satisfactorily, so that we did not insist on an adaptive component, but we kept an extrapolative component in a slightly different manner. As an extrapolative component, last month's inflation rate that we are sure all the respondents are aware of is taken. Since our data points are December, March, June and September, we pick up the inflation rates of November, February, May and August in every year. Exchange rates are also included in the model in order to capture its influence on costs through imported raw materials, and also it may be used as a policy variable that measures uncertainty. Real Sector Confidence Index[1] is also used as an alternative policy variable, but it did not perform well, probably due to some opposite movements with inflation. As a demand variable private capacity utilization rate and respondents' evaluation of last three months' stocks of produced goods are taken. Finally, as a forward-looking component average cost expectations of respondents over the next three months is included in the model. Estimation result is:

$$P_{t+1}^{e(c)} = -0.12 + 0.72PMINF + 0.61DLR + 0.29PCUR - 0.22BTS20 + 0.37BTS21 \quad (5)$$
$$(-2.04) \quad (4.20) \qquad (7.86) \qquad (5.43) \qquad (-3.44) \qquad (5.41)$$

R_{adj}^2=0.83, SBC=-4.97, WH=0.78, LM(1)=0.42, LM(2)=0.18, LM(3)=0.31, LM(4)=0.47, ARCH LM(1)=0.65, ARCH LM(2)=0.77, ARCH LM(3)=0.73, ARCH LM(4)=0.47

Note: t-statistics in parenthesis and p-values for the diagnostic tests.

PMINF: Private manufacturing sector inflation rate in the months November, February, May and August. (SIS).

DLR: Percentage changes in average dollar exchange rates of December, March, June and September in regards to previous month's average dollar exchange rates. (CBRT)

[1] It is constructed by making use of the responses given to the selected questions of the BTS.

PCUR: Logarithmic private sector capacity utilization rate divided by 100, quarterly. (SIS)

BTS20: Respondents' evaluations of produced goods stocks of the last three months -diffusion index- divided by 100, in the months November, February, May and August, taken from 20th question of the BTS. (CBRT)

BTS21: Respondents' average cost expectations over the next three months -diffusion index- divided by 100, in the months November, February, May and August, taken from 21st question of the BTS. (CBRT)

All the variables are significant with expected signs, and diagnostics of the regression are fairly good. Expectations are very sensitive to the last month's inflation rate and exchange rate changes as expected. Also, production capacity and stock levels of the last three months are important determinants of price expectations of the next three months. An increase in production capacity or a decrease in stock levels in the last three months are perceived as a sign of increase in demand, so they enter the regression with positive and negative signs respectively.

The role of public sector in the formation of expectations has been tested by considering public manufacturing sector prices[1], total expenditures and some budget deficit definitions but found to be insignificant. Some cost variables such as energy prices, wage payments and interest rates have also been found to be insignificant[2], therefore not presented in the equation. Among them insignificance of wage payments may be attributed to some factors such as small share of wage payments in total costs, lack of union activities during the period[3] or to the flexibility of wage rates during economic crises[4].

The policy implications of these results, if correct, are mixed. Firstly, as the proposed model shows, the respondents do not take into account the past history of the variables; instead they give more importance to the recent data. It is encouraging in the sense that inflation expectations may respond policy changes rather quickly. But the model also points out that the dollar exchange rate is a highly significant component of the expectations formation, and the volatility in this item may disturb the expectations. This result is not an unexpected one so long as dollar exchange rate preserves its importance both through the imported raw materials and as a store of value in the studied sample period. Also note that to reduce firms' inflation expectations, it is necessary to reduce the actual rate of inflation and to give more weight to the factors that affect the firms' average cost

[1] This sector is mainly composed of administered prices and it is thought that monitoring the behavior of this sector may carry important information for expectations since public manufacturing goods are direct inputs to private manufacturing sector and revenue that is generated by increasing these prices is intensively used as a source of financing the budget deficits.

[2] Wages and energy prices are also found to be insignificant in Uygur (1989).

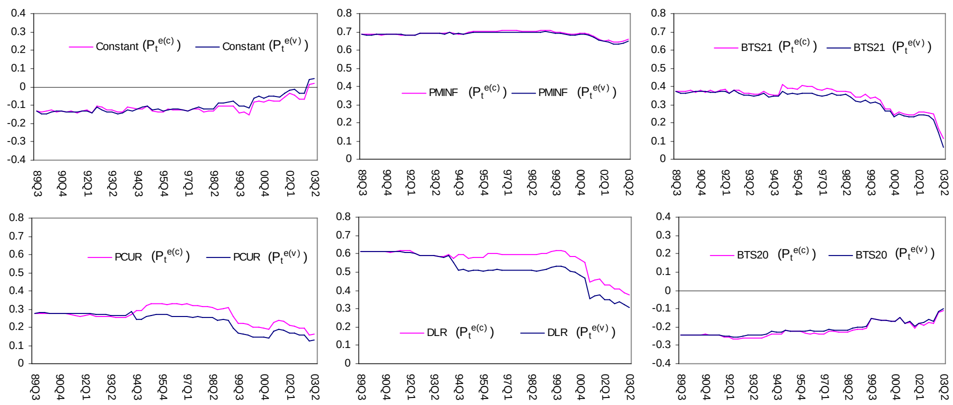
[3] Lim and Papi (1997).

[4] Yeldan (2001).

expectations since the behavior of these variables apparently have a significant effect on expectations.

As mentioned in the first part, Turkish economy underwent a major structural change after the stabilization and disinflation program. As a last exercise we have tried to see whether this structural change can be traced from the estimated time varying coefficients of augmented model via the adaptive Kalman Filter algorithm (see Appendix B.2) and the results are presented in Figure II. Time varying coefficients suggest that realized inflation rates preserve its significant role on expectations after the stabilization program, whereas respondents give less weight, in absolute value, to changes in dollar exchange rate, cost expectations, stock evaluations and capacity utilization rates.

Figure II: Time Varying Parameter Results of the Augmented Model



4. Tests of Rationality

Rationality hypothesis in Muthian sense has three implications (Muth, 1961). Firstly, economic system generally does not waste scarce information set. Secondly, the way expectations are formed depend specifically on the structure of the relevant system describing the economy, and lastly, a public prediction will have no substantial effect on the operation of the economic system. Based on these three assertions REH can be defined as the true mathematical expectation implied by the model, conditional on the information set assumed to be available when expectations are formed. Here, the true mathematical expectation stands for the expectation that can be derived by writing the correct structural model of the economy. Furthermore, the information set is assumed to have three components: "knowledge of the structure of the model; knowledge of government policies in operation; and knowledge of the past values of economic variables" (Begg, 1982, p.72).

REH defined as above has some statistical properties that utilized in empirical testing. These are (Evans and Gulemani, 1984):

1. Unbiasedness: Under REH, expectations should not systematically over or under estimate the realized inflation rates. Formally,

$$P_t = P_t^e + u_t \tag{6}$$

where u_t is uncorrelated with P_t^e and $u_{t-i}(i=1,2,...)$. Regressing the equation

$$P_t = \alpha + \beta P_t^e + u_t \tag{7}$$

and testing $H_0: (\alpha, \beta) = (0,1)$ is the conventional test of unbiasedness.

2. No serial correlation: Let e_t denote forecast error made in time t, if there is a pattern of systematic forecast error, then e_t's are serially correlated. This hypothesis can be tested by making use of the following regression

$$e_t = \sum_{i=1}^{k} \beta_i e_{t-i} + u_t \tag{8}$$

and testing $H_0: \beta_i = 0 \; (i = 1,...,k)$ for a selected k.

3. Efficiency: REH assumes people use, at least, the information regarding the past history of inflation. To be precise the forecast error should not be improved by utilizing P_{t-i}, $(i=1,2,...)$. This can be tested by estimating

$$e_t = \sum_{i=1}^{k} \beta_i P_{t-i} + u_t \tag{9}$$

and testing the null hypothesis $H_0: \beta_i = 0$, $(i = 1,...,k)$ for a selected k.

4. Orthogonality: Let I_{t-1} consists of all the costlessly available information that helps to explain inflation at the end of period $t-1$. If the forecast error could be improved by exploiting I_{t-1}, then the forecasts are not rational. Therefore, a test of orthogonality looks for a lack of serial correlation between e_t and a sub-sample of I_{t-1}. Let S_t be a vector of information variables from I_t, and β is a vector of coefficients, then regressing

$$e_t = S_{t-1}^{'} \beta + u_t \tag{10}$$

and testing $H_0: \beta = 0$ is the test of orthogonality.

In the tests of rationality, we have to bear in mind that P_t^e series is not the direct measure, it is derived from the qualitative surveys and is likely to subject to some measurement error. In order to illustrate what kind of errors we are faced with, consider the following representation[1]:

$$P_t^e = \varphi_1 + \varphi_2 P_t^* + \varepsilon_t \tag{11}$$

here P_t^* stands for the true measure of inflation expectations which is in fact not available, and ε_t represents the random component of the measurement errors, whereas the parameters φ_1 and φ_2 correspond to the systematic components. It is assumed that the random component has a zero mean and constant variance. If the scaling problem is serious, then we should have φ_1 and φ_2 values that are significantly different from zero and one respectively.

The unbiasedness and no serial correlation tests can be applied only when the expectations are measured without systematic or random errors. However, as Pesaran (1985, p.950) stated “the conversion of qualitative responses into quantitative measurements, no matter how ingeniously carried out, will not be free from error” owing to incorrect scaling (systematic error) and general uncertainty (random error), so there is no point to apply these tests. Expectation measures obtained according to CP’s methodology prevents the test of the unbiasedness itself because of its manner of construction. The scaling parameter, which is changeable according to the sample size and the distribution chosen, may bring about diverse results for the unbiasedness tests. Also, Pesaran emphasized that existence of the random measurement errors brings in the moving-average errors, which in turn precludes a precise test of no serial correlation. However, these do not preclude examining the series in terms of rationality with the remaining tests provided that the measurement error is white noise. As a result, we can apply only the efficiency and the orthogonality tests.

If we turn back to the efficiency test defined in equation (9) and as well consider the true expectation measure, specifically the measurement error problem, then we have the following regression for the efficiency test:

$$e_t = P_t - P_t^* = c_0 + \sum_{i=1}^{k} \beta_i P_{t-i} + u_t \tag{12}$$

However, in the existence of the measurement error, our inflation expectation measure differentiates from the true one by $P_t^* = \varphi_2^{-1}(P_t^e - \varphi_1 - \varepsilon_t)$ hypothetically. Hence we are faced with the following regression:

$$P_t - \varphi_2^{-1} P_t^e = c_0^{'} + \sum_{i=1}^{k} \beta_i P_{t-i} + \delta_t \tag{13}$$

where $c_0^{'} = c_0 - (\varphi_1 / \varphi_2)$ and $\delta_t = u_t - (\varepsilon_t / \varphi_2)$. In this case, according to the degree of the measurement error, it is likely to expect $c_0^{'}$ to be different from

[1] See Pesaran (1985) for the further details concerning the measurement error problem.

zero, and δ_t to be serially correlated, and these do not necessarily cause to reject REH. In the applications, however, it is seen that there is no need to consider various specifications for the autocorrelation. Moreover, as in the Pesaran's work, three different values of φ_2 (0.9, 1.0 and 1.1) are assumed since it is not possible to estimate this parameter precisely. We also apply the same logic to the orthogonality test later.

Table I gives empirical results of the efficiency test for four different lag structures. It shows that REH cannot be rejected as a result of efficiency test, namely forecast errors could not be improved by past data of realized inflation rates which means that the basic condition for the rationality, fully utilization of the most costlessly available information that is past inflation rates, has been satisfied.

Table I: Results of the Efficiency Test

Sample: 1989:04-1999:04						
	$\varphi_2 = 0.9$		$\varphi_2 = 1$		$\varphi_2 = 1.1$	
	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value
P_{t-1}	1.743	0.194	1.942	0.164	2.108	0.147
P_{t-1}, P_{t-2}	1.973	0.373	2.118	0.347	2.234	0.327
$P_{t-1}, P_{t-2}, P_{t-3}$	2.312	0.510	2.276	0.517	2.267	0.519
$P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}$	3.758	0.440	3.682	0.451	3.611	0.461

Now, we turn to the most comprehensive test of rationality that is the orthogonality test. Here, crucial point is selection of independent variables and the number of lags that these variables enter into the equation. We begin with a large number of variables; say information set I that is potentially related with inflation. Possible candidates are lags of inflation rates (q, m), nominal wages (q), exchange rates (q, m), energy prices (q, m), public sector manufacturing prices (q, m), money supply (q) and deposit interest rates (q), where (q) indicates quarterly data, and (q, m) both quarterly and monthly data.

In the literature, usually, quarterly data with appropriate lag numbers have taken into consideration as independent variable whenever dependent variable is quarterly. If we use the conventional procedure and take one lag of each quarterly variable in the information set I, except money supply and deposit interest rates where two lags are allowed in order to capture publication delays, then rationality assumption is rejected (not shown). However, this method may be misleading for two reasons. Firstly, we do not know a priori that information set should consist of only quarterly variables. There may be some variables with non-quarterly frequency that perform better in explaining inflation. Especially, if the pass-through from changes in cost items to inflation is immediate, then the effect of the contemporaneous changes in these items may weight more on inflation expectations. If this is the case, orthogonality test should be carried out with more recent information set, specifically monthly data may also be allowed. Secondly, in rationality tests independent variables are regressed on the forecast error, therefore there is a possibility of identifying some variables in explaining the forecast error significantly but fail to explain realized inflation rates. This problem may arise from some variables that are spuriously correlated with inflation expectations but actually are not predictors of inflation (Madsen, 1996).

In order to overcome these shortcomings of the orthogonality test, we have first tried to determine which variables in the information set I are successful in explaining the inflation. Once we form such information set, we expect that a rational respondent should be aware of these variables and form his/her inflation expectation accordingly. It is found that dollar exchange rate (DLR), one lag of inflation (PMINF) and private sector capacity utilization rate (PCUR) that used in augmented expectations formation mechanism are significant and satisfactorily explain the inflation process. Estimation result is:

$$P_t = 0.11 + 1.19PMINF + 0.64DLR + 0.18PCUR + 0.22D94Q2 \qquad (14)$$
$$\begin{matrix}(4.41) & (4.25) & (4.26) & (2.55) & (6.19)\end{matrix}$$

R^2_{adj} =0.85, SBC=-4.23, WH=0.35, LM(1)=0.74, LM(2)=0.72, LM(3)=0.88, LM(4)=0.94, ARCH LM(1)=0.57, ARCH LM(2)=0.86, ARCH LM(3)=0.93, ARCH LM(4)=0.59

Note: t-statistics in parenthesis and p-values for the diagnostic tests

Description and justification of independent variables used in equation (14) has already been discussed in Section III, except the dummy variable $D94Q2$ that put into the equation in order to capture effect of the financial crises occurred in the second quarter of 1994. Now, we are sure that PMINF, DLR and PCUR explain the inflationary process satisfactorily and available to the respondents when they form their expectations. At this stage it is more sensible to investigate whether these variables are fully exploited or not. However, it should be kept in mind that the results of the orthogonality tests are conditional on the variables appeared in the inflation equation.

Table II shows results of the orthogonality test under three different values of φ_2. Null hypothesis that the forecast error cannot be improved by a better utilization of PMINF, DLR and PCUR is rejected at different levels of significance depending on value φ_2 takes[1]. If we examine each variable separately, it is seen that the forecast error could be improved only by PCUR.

Table II: Result of the Orthogonality Test

Sample: 1989:04-1999:04					
$\varphi_2 = 0.9$		$\varphi_2 = 1$		$\varphi_2 = 1.1$	
Chi-square	P-value	Chi-square	P-value	Chi-square	P-value
9.428	0.024	7.795	0.050	6.698	0.082

From the efficiency and orthogonality tests we can conclude that although respondents fully utilize past data on inflation rates, the forecast errors could be improved partially with better understanding of PCUR-inflation relation, and in this sense inflation expectations are not rational though for φ_2=1.1 we have not rejected rationality very powerfully.

Since rejection of rationality stemmed from just underutilization of one variable, and it is not very powerful for each value of φ_2, we have begun to suspect whether underutilization of the relevant information, PCUR, common in the whole period or it is peculiar to some very specific time period. If we examine Figure I a little bit more closely, it is seen that in 1999, $P_t^{e(c)}$ measure decreases more sharply as compared to the realizations, which results in large forecast errors. The variable PCUR may owe its significance to co-movements with the forecast error in this period. In order to test this argument and to see whether our results regarding the rationality robust to time period selected or not, we have excluded the last year that inflation expectations fall very sharply, and then reapplied efficiency and orthogonality tests. The results are presented in Table III.

Results of efficiency and orthogonality tests indicate that we cannot reject the rationality for the period 1989:4-1998:4, indicating respondents utilize available information set I efficiently. The sample period selected has a crucial effect on the outcome, and it suggests that PCUR is not neglected in the whole period. Underutilization of the information regarding PCUR in just a few periods should not create a systematic bias in terms of the rationality hypothesis.

[1] Actually, this result is valid for the $P_t^{e(c)}$ series. When $P_t^{e(v)}$ is used, test results conclude that expectations are significantly rational for all values of φ_2.

Table III: Results of the Revisited Efficiency and Orthogonality Tests

Sample: 1989:04-1998:04						
Efficiency Test	$\varphi_2 = 0.9$		$\varphi_2 = 1$		$\varphi_2 = 1.1$	
	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value
P_{t-1}	3.037	0.081	2.982	0.084	2.896	0.089
P_{t-1}, P_{t-2}	3.927	0.140	3.637	0.162	3.357	0.187
$P_{t-1}, P_{t-2}, P_{t-3}$	3.871	0.276	3.542	0.315	3.258	0.354
$P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}$	4.905	0.297	4.531	0.339	4.197	0.380
Orthogonality Test						
	2.665	0.446	2.849	0.415	3.875	0.275

Results of efficiency and orthogonality tests indicate that we cannot reject the rationality for the period 1989:4-1998:4, indicating respondents utilize available information set I efficiently. The sample period selected has a crucial effect on the outcome, and it suggests that PCUR is not neglected in the whole period. Underutilization of the information regarding PCUR in just a few periods should not create a systematic bias in terms of the rationality hypothesis.

However, these results should be taken with some care for a number of reasons. As mentioned above, the rationality assumption demands a more or less stable economic environment in order to allow economic agents to learn true inflation generating mechanism. But, the period we have studied on witnessed a major domestic and two external (East Asian and Russian Crises) financial crises that make it difficult to assume respondents had enough time to learn the true generating mechanism. If this argument is accepted, it is expected to create a bias in the direction of rejecting rationality and, as we have not rejected the rationality it may be seen irrelevant here. But, it may also imply the reverse. If the lack of economic stability prevents construction of a strong econometric model of inflation, then it may also weaken the test, since in such a situation it is enough to give expectations as good as the econometric model to pass the rationality tests. Namely, when the economy could not be expressed in a structurally consistent way that enables forecasters to give accurate inflation expectations at time t-1 for t, it gets more difficult to blame them of not incorporating some relevant information in their inflation expectations that would lead to rejection of the rationality. The same thing may also arise when pass-through is too speedy from changes in demand or cost conditions to inflation or price adjustments. Since, in this study we have taken private manufacturing sector inflation rates as realized inflation series, changes in demand and especially in cost conditions that are strictly related with changes in exchange rates and pricing policy of public sector, are probably reflected to manufacturing prices within the same month or one month later. If this is the case, both expectations and econometric estimates formed with all available information at time t-1 carry out an uncertain part due to changes in cost conditions in three months between t-1 and t. This also may explain why energy prices that heavily controlled by the public sector did not enter into the equations (5) and (14).

5. Conclusion

In this paper, we have attempted to analyze inflation expectations of private manufacturing sector. Firstly, overlapping nature of the data resulting from the inconsistency between frequency of data and forecast horizon is discussed and, this problem is solved by picking up non-overlapping data points. Then, by utilizing sequential testing procedure, it is studied to what extent the simple models of expectations formation explains the expectation measures obtained from the BTS. Resultant mechanism is a general first order extrapolative. But it is found that simple expectations formation hypotheses considered in the sequential testing procedure are too simple to describe the formation procedure of producers' price formation sufficiently. Inadequacy of the simple models led us to propose a new model of expectations formation that incorporates economic variables other than past history of the inflation and expectations. In this model, last month's inflation rate, changes in average dollar exchange rates of the last month, quarterly private sector capacity utilization rate and respondents' evaluation of produced goods stocks of the last three months and average cost expectations of the next three months have been found to be significant.

In the last part, rationality has been discussed rather comprehensively. From that discussion, it can be stated the conventional procedure that use the same frequency for both dependent and independent variables may not be appropriate in our case and it is important to assure significance of independent variables used in orthogonality test as explanatory variables of inflation. From the tests, results on rationality seem heavily depending on the time period selected, but for the period 1989:04-1998:04 expectations are rational in the sense that there is no underutilization of relevant available information. However, results suggest that failing to reject the rationality may not come from the accuracy of the expectations, but of structure of the economy that preclude accurate short term forecasting for time t with available information set at t-1.

APPENDIX:

A.1. Tables

Table A.1: The Actual and Expected Quarterly Inflation Series

	Realized Inflation	Inflation Exp. (Constant Indiff. Interval)	Inflation Exp. (Time Varying Indiff. Interval)		Realized Inflation	Inflation Exp. (Constant Indiff. Interval)	Inflation Exp. (Time Varying Indiff. Interval)
89:Q1	13.95	21.32	17.74	Q2	16.13	16.41	15.57
Q2	13.24	16.23	14.21	Q3	12.51	15.77	14.13
Q3	8.90	15.83	11.37	Q4	14.76	15.17	13.86
Q4	7.04	10.29	8.38	97:Q1	15.52	16.77	15.12
90:Q1	9.65	12.27	9.44	Q2	15.10	16.44	15.04
Q2	7.66	9.80	7.99	Q3	18.88	18.18	18.33
Q3	12.59	10.74	11.72	Q4	17.28	15.96	16.77
Q4	9.93	13.98	13.78	98:Q1	13.36	15.41	15.32
91:Q1	12.11	12.14	12.54	Q2	11.65	13.84	14.08
Q2	13.25	16.21	14.76	Q3	10.82	10.88	11.97
Q3	11.05	11.08	11.40	Q4	9.52	8.78	10.34
Q4	11.15	12.14	11.56	99:Q1	9.67	6.01	8.87
92:Q1	12.24	13.88	12.35	Q2	12.37	8.03	10.97
Q2	12.94	12.69	12.82	Q3	13.38	8.51	11.56
Q3	14.63	15.48	14.64	Q4	14.31	8.96	12.22
Q4	12.81	17.15	13.89	00:Q1	11.71	9.01	12.19
93:Q1	12.09	15.86	13.06	Q2	6.69	6.02	8.89
Q2	10.83	13.94	11.65	Q3	5.85	4.66	7.88
Q3	11.59	14.39	11.78	Q4	5.93	5.28	8.11
Q4	13.17	14.01	12.85	01:Q1	15.52	4.88	8.44
94:Q1	23.43	20.16	18.01	Q2	28.08	13.82	15.25
Q2	49.21	27.53	23.72	Q3	14.71	7.43	10.57
Q3	10.29	8.89	10.98	Q4	14.62	8.37	11.34
Q4	16.04	11.76	12.99	02:Q1	4.19	6.78	9.85
95:Q1	19.98	20.12	18.87	Q2	6.21	4.33	8.05
Q2	10.27	13.24	13.40	Q3	9.26	7.29	10.07
Q3	11.81	12.15	12.60	Q4	5.61	5.37	7.71
Q4	10.97	12.87	12.36	03:Q1	6.83	5.00	7.43
96:Q1	17.21	20.45	17.15	Q2	1.21	5.42	5.68

Table A.2: Expectations Formation Mechanisms[1]

Extrapolative

Simple first order:

(F1): $P_{t+1}^{e} = P_t + \theta(P_t - P_{t-1}) + w_t$, (F2): (F1) with $w_t = \rho w_{t-1} + u_t$

General first order:

(F3): $P_{t+1}^{e} = \beta_1 P_t + \beta_2 P_{t-1} + w_t$, (F4): (F3) with $w_t = \rho w_{t-1} + u_t$

Simple second order:

(F5): $P_{t+1}^{e} = P_t + \theta_1(P_t - P_{t-1}) + \theta_2(P_{t-1} - P_{t-2}) + w_t$, (F6): F(5) with $w_t = \rho w_{t-1} + u_t$

Adaptive

Simple first order:

(F9): $P_{t+1}^{e} = P_t^{e} + \lambda(P_t - P_t^{e}) + w_t$, (F10): (F9) with $w_t = \rho w_{t-1} + u_t$

General first order:

(F11): $P_{t+1}^{e} = \alpha_1 P_t^{e} + \beta_1 P_t + w_t$, (F12): (F11) with $w_t = \rho w_{t-1} + u_t$

Simple second order:

(F13): $P_{t+1}^{e} = P_t^{e} + \lambda_1(P_t - P_t^{e}) + \lambda_2(P_{t-1} - P_{t-1}^{e}) + w_t$, (F14): (F13) with $w_t = \rho w_{t-1} + u_t$

General second order:

(F15): $P_{t+1}^{e} = \alpha_1 P_t^{e} + \alpha_2 P_{t-1}^{e} + \beta_1 P_t + \beta_2 P_{t-1} + w_t$, (F16): (F15) with $w_t = \rho w_{t-1} + u_t$

Maintained

(F19): $P_{t+1}^{e} = \gamma_1 P_t^{e} + \gamma_2 P_{t-1}^{e} + \gamma_3 P_{t-2}^{e} + \delta_1 P_t + \delta_2 P_{t-1} + \delta_3 P_{t-2} + \delta_4 P_{t-3} + w_t$

Note: P_{t+1}^{e} is the expectation formed at t for t+1. Constant term is omitted in the table for simplicity.

[1] Mills (1981,p.157), Table I

Table A.3: Sequential Testing Results of the Ordered Nests

Extrapolative Nests:								
								Model selected
N1	Model	**(F19)**	**(F8)**	**(F6)**	**(F5)**	**(F1)**		
	Test Stat.	Maintained	1.01 (3)	29.39 (1)	-	-		**(F8)**
N2	Model	**(F19)**	**(F8)**	**(F7)**	**(F5)**	**(F1)**		
	Test Stat.	Maintained	1.01 (3)	10.77 (1)	-	-		**(F8)**
N3	Model	**(F19)**	**(F8)**	**(F7)**	**(F3)**	**(F1)**		
	Test Stat.	Maintained	1.01 (3)	10.77 (1)	-	-		**(F8)**
N4	Model	**(F19)**	**(F8)**	**(F4)**	**(F3)**	**(F1)**		
	Test Stat.	Maintained	1.01 (3)	0.59 (1)	12.98 (1)	-		**(F4)**
N5	Model	**(F19)**	**(F8)**	**(F4)**	**(F2)**	**(F1)**		
	Test Stat.	Maintained	1.01 (3)	0.59 (1)	55.48 (1)	-		**(F4)**
Adaptive Nests:								
N6	Model	**(F19)**	**(F16)**	**(F14)**	**(F13)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	14.38 (2)	-	-		**(F16)**
N7	Model	**(F19)**	**(F16)**	**(F14)**	**(F10)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	14.38 (2)	-	-		**(F16)**
N8	Model	**(F19)**	**(F16)**	**(F12)**	**(F10)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	5.49 (2)	9.39 (1)	-		**(F12)**
N9	Model	**(F19)**	**(F16)**	**(F12)**	**(F11)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	5.49 (2)	0.01 (1)	12.36 (1)		**(F11)**
N10	Model	**(F19)**	**(F16)**	**(F15)**	**(F12)**	**(F10)**	**(F9)**	
	Test Stat.	Maintained	1.32 (2)	5.37 (1)	-	-	-	**(F16)**
N11	Model	**(F19)**	**(F16)**	**(F15)**	**(F13)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	5.37 (1)	-	-		**(F16)**
N12	Model	**(F19)**	**(F16)**	**(F15)**	**(F11)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	5.37 (1)	-	-		**(F16)**
N13	Model	**(F19)**	**(F16)**	**(F18)**	**(F17)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	17.44 (1)	-	-		**(F16)**
N14	Model	**(F19)**	**(F16)**	**(F15)**	**(F17)**	**(F9)**		
	Test Stat.	Maintained	1.32 (2)	5.37 (1)	-	-		**(F16)**
N15	Model	**(F19)**	**(F16)**	**(F18)**	**(F14)**	**(F10)**	**(F9)**	

	Test Stat.	Maintained	1.32 (2)	17.44 (1)	-	-	-	**(F16)**
N16	Model	**(F19)**	**(F16)**	**(F18)**	**(F14)**	**(F13)**	**(F9)**	
	Test Stat.	Maintained	1.32 (2)	17.44 (1)	-	-	-	**(F16)**
N17	Model	**(F19)**	**(F16)**	**(F18)**	**(F17)**	**(F13)**	**(F9)**	
	Test Stat.	Maintained	1.32 (2)	17.44 (1)	-	-	-	**(F16)**
N18	Model	**(F19)**	**(F16)**	**(F15)**	**(F17)**	**(F13)**	**(F9)**	
	Test Stat.	Maintained	1.32 (2)	5.37 (1)	-	-	-	**(F16)**

$\chi^2_{1,0.025}=5.02, \chi^2_{2,0.025}=7.38, \chi^2_{3,0.025}=9.35$

Note: Number of restrictions in parenthesis

B.1. State Space Representation for Time Varying Indifference Intervals

The vector of observed variable (inflation) is denoted as **y**, while the vector of unobserved state variables (time varying indifference interval bounds) is denoted by β. Then the measurement equation where the evolution of the observed variables is described as a function of the unobserved state variables and transition equation are:

$$y_t = H_t \beta_t + v_t \tag{B.1.1}$$

$$\beta_t = \Phi_t \beta_{t-1} + G_t w_t \tag{B.1.2}$$

where the **v** and **w** denotes vectors of normally distributed iid shocks, which are assumed to be uncorrelated with each other and also with the initial state vector β_0.

For $\beta_t = (\delta_{mt}, \delta_{pt})'$ we can write the state space representation as:

$$[P_t] = \begin{bmatrix} \dfrac{a_t}{a_t - b_t} & \dfrac{-b_t}{a_t - b_t} \end{bmatrix} \begin{bmatrix} \delta_{m,t} \\ \delta_{p,t} \end{bmatrix} + v_t \tag{B.1.3}$$

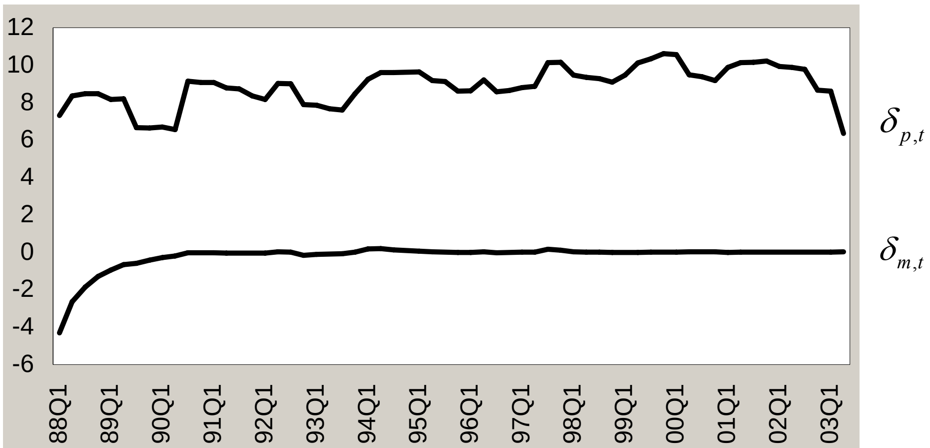
$$\begin{bmatrix} \delta_{m,t} \\ \delta_{p,t} \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \begin{bmatrix} \delta_{m,t-1} \\ \delta_{p,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon^m \\ \varepsilon^p \end{bmatrix} \tag{B.1.4}$$

By using the adaptive Kalman filter[1], our aim is to obtain the adaptive estimate of v_t and in this way to include the transitory shocks into the model by allowing them vary over time.

[1] See Harvey (1990) and Ozbek(2000) for the technical details of Kalman filter and its adaptive version.

In the literature, Φ_t is assumed to be identity, which means threshold parameters, deviates from its previous value according to the white noise shocks. However, as stated by Nardo (2003), there are no economic or psychological reasons to suppose that individuals have an indifference interval following a random walk. For that reason, in this model time varying lower and upper threshold parameters are assumed to be shaped with respect to different values of the ρ_1 and ρ_2 parameters.

Figure: Estimated Lower and Upper Threshold Parameters



As it can be seen from the Figure, lower threshold parameter takes values close to zero or with zero included. This can be interpreted as, for participants there is no need to have a fall in their prices to declare "down" option for their price expectations instead no change case is enough for them. This result can be perceived as meaningful especially considering that inflation was in high levels along the sample period. On the other hand, it is seen that upper bound has a changeable structure but its average is close to the values obtained from the constant indifference interval case under different distribution assumptions.

B.2. State Space Representation for Augmented Model of Expectations Formation

Assuming parameter vector evolves according to a random walk process, and following (5), (B.1.1) and (B.1.2), the state space representation can be written as:

$$\left[P_{t+1}^{e}\right]=\left[1 \quad PCUR_t \quad PMINF_t \quad DLR_t \quad BTS21_t \quad BTS20_t\right]\begin{bmatrix}\theta_{1,t}\\ \theta_{2,t}\\ \theta_{3,t}\\ \theta_{4,t}\\ \theta_{5,t}\\ \theta_{6,t}\end{bmatrix}+e_t \tag{B.2.1}$$

$$\begin{bmatrix}\theta_{1,t}\\ \theta_{2,t}\\ \theta_{3,t}\\ \theta_{4,t}\\ \theta_{5,t}\\ \theta_{6,t}\end{bmatrix}=\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}\begin{bmatrix}\theta_{1,t-1}\\ \theta_{2,t-1}\\ \theta_{3,t-1}\\ \theta_{4,t-1}\\ \theta_{5,t-1}\\ \theta_{6,t-1}\end{bmatrix}+\begin{bmatrix}\varepsilon_t^{\theta_1}\\ \varepsilon_t^{\theta_2}\\ \varepsilon_t^{\theta_3}\\ \varepsilon_t^{\theta_4}\\ \varepsilon_t^{\theta_5}\\ \varepsilon_t^{\theta_6}\end{bmatrix} \tag{B.2.2}$$

where P_{t+1}^{e} is the observed expectations on the system and θ_t is the state vector of six parameters. The observation error e_t and the state disturbance ε_t^{θ} are mutually uncorrelated, normally distributed, random vectors of zero mean with covariance matrices R_t and Q_t respectively. By using the initial guess of the parameter vector and its covariance matrix and then applying the adaptive Kalman filter, time varying estimates of the parameters in the model are obtained and reported in the Figure III in the text.

C.1. Augmented Dickey-Fuller Unit Root Results (1989Q4-1999Q4)

	Augmented Dickey-Fuller test statistic		
Variables	t-statistic	Prob.*	LM(4)**
$P_t^{e(c)}$	-4.061	0.003	1.26 (0.30)
$P_t^{e(v)}$	-3.901	0.005	1.32 (0.28)
PMINF	-3.902	0.005	0.49 (0.74)
DLR	-7.824	0.000	0.24 (0.91)
PCUR	-3.420	0.016	0.40 (0.81)
BTS20	-2.814	0.065	0.15 (0.96)
BTS21	-3.327	0.020	1.19 (0.33)

		With Constant
Test critical values:	1% level	-3.600987
	5% level	-2.935001
	10% level	-2.605836

* The critical values for the ADF test are based on MacKinnon (1996).

** LM(4) represents Breusch-Godfrey serial correlation LM test F statistic for 4 lags. The figure in parenthesis is its p-value.

REFERENCES

BEGG, K. H.D. 1982. *The Rational Expectations Revolution in Macroeconomics*. The Johns Hopkins University Press, Baltimore, Maryland.

BRYAN, M.F. and W.T. GAVIN, 1986. Models of Inflation Expectations Formation: A Comparison of Household and Economist Forecasts: A Comment, *Journal of Money, Credit, and Banking*, Vol.18, iss.4, 539—44.

CAGAN, P. 1956. The Monetary Dynamics of Hyperinflation in M. Friedman (ed.), *Studies in the Quantity Theory of Money*, University of Chicago Press.

CARLSON, J.A. and M. Parkin, 1975. Inflation Expectations, *Economica*, Vol.42, 123—38.

DAVIDSON, R. and J. G. MACKINNON, 1993. *Estimation and Inference in Econometrics*. Oxford University Press.

EVANS, G. and R. GULAMANI, 1984. Tests for Rationality of the Carlson-Parkin Inflation Expectations Data, *Oxford Bulletin of Economics and Statistics*, Vol.46, No.1.

GRANT, A.P. and L.B. THOMAS, 1999. Inflationary Expectations and Rationality Revisited, *Economics Letter*, Vol.62, 331—38.

HANSEN, L.P. 1982. Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029—54.

HARVEY, A.C. 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

LIM, C.H. and L. PAPI, 1997. An Econometric Analysis of the Determinants of Inflation in Turkey, *IMF Working Paper*, WP/97/170.

LUCAS, R.E. 1972. Expectations and the Neutrality of Money, *Journal of Economic Theory*, 4, 103—124.

MACKINNON, J. G. 1996. Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, 11, 601—618.

MADSEN, J.B. 1996. Formation of Inflation Expectations: From the Simple to the Rational Expectations Hypothesis, *Applied Economics*, Vol.28, 1331—37.

MILLS, T.C. 1981. Modelling the Formation of Australian Inflation Expectations, *Australian Economic Papers*, Vol.21, 150—60.

MIZON, G.E. 1977. Inferential Procedures in Nonlinear Models: An Application in a U.K. Industrial Cross Section Study of Factor Substitution and Returns to Scale, *Econometrica*, Vol.45, 1221—42.

MUTH, J.F. 1961. Rational Expectations and the Theory of Price Movements, *Econometrica,* Vol.29, 315—35.

NARDO, M. 2003. The Quantification of Qualitative Survey Data: A Critical Assessment, *Journal of Economic Surveys*, Vol.17, No.5, 645—68.

ORAL, E. 2002. Inflation Expectations on the Basis of Qualitative Surveys, *Presented at the 26th CIRET Conference*, Taipei, October.

OGUNC, F., HAMSICI, T., ORAL, E and E. KARADAS, 2004. Quantification of Inflation Expectations, Unpublished Manuscript.

OZBEK, L. 2000. Uyarlı Kalman Filtresi, *Gazi Ünv. Fen Bilimleri Ens. Dergisi*, Cilt 13, No:2, 369—80.

PESARAN, M.H. 1985. Formation of Inflation Expectations in British Manufacturing Industries, *The Economic Journal*, Vol.95, 948—75.

SARGENT, T.J. 1973. Rational Expectations, the Real Rate of Interest and the Natural Rate of Unemployment, *Brookings Papers on Economic Activity*, 2, 429—472.

SEITZ, H. 1988. The Estimation of Inflation Forecasts From Business Survey Data, *Applied Economics*, 20, 427—38.

SMITH, J. and M. McALEER 1995. Alternative Procedures for Converting Qualitative Response Data to Quantitative Expectations, *Journal of Applied Econometrics*, Vol. 10, No. 2, 165—85.

UYGUR, E. 1989. Inflation Expectations of the Turkish Manufacturing Firms, *Central Bank of the Republic of Turkey Discussion Paper.*

YELDAN, A.E. 2001. *Küreselleşme Sürecinde Türkiye Ekonomisi: Bölüşüm, Birikim ve Büyüme*. İletişim Yayınları, İstanbul.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 391—425

OCCUPATIONAL SEGREGATION IN THE RUSSIAN LABOUR MARKET

Anastasia Klimova[1]

ABSTRACT

This paper examines gender inequalities in the labour market in Russia by examining gender occupational segregation. First, a statistical overview of the trends in the Russian labour market outcomes by gender during the Soviet period, perestroika and the period of transition is presented. This is followed by a review of current economic theories of occupational segregation and discrimination in the labour market, including explanations based on tastes, occupational exclusion and statistical discrimination. Following this, different measures of occupational segregation are evaluated and a brief methodological discussion of the advantages and disadvantages of each is set out. This paper relies on an analytical framework identifying segregated, that is male- or female-dominated, and integrated occupations. Evidence from a nationally representative longitudinal survey of Russian citizens (RLMS) is used to analyse the trends, extent and causes of gender-based occupational segregation from 1994 to 2001 during the transition to a market economy.

1. Intraduction

An American folk saying, " A woman's place is in the home…", illustrates vividly the traditional views concerning the role of women in the Russian labour market. In 1992, the then Minister of Labour was quoted as saying that "…wherever there is a man unemployed, we should not worry about female unemployment." A number of recent studies have revealed that the ideology has changed only slightly since the period of transition begun in 1991, and that the status of women in the Russian economy remains a major concern (Gerry *et al*, 2001).

One of the most important issues concerning the position of women in the Russian labour market is occupational segregation and women's concentration in low-paying occupations (Ogloblin, 1999). The answer to the question of whether

[1] Department of Economics and Business, University of Sydney, Australia.

or not occupational segregation in Russia increased between 1994 and 2001 is examined in this paper.

Despite the importance of occupational segregation as a barrier to improve equity and efficiency in the Russian labour market, there has been relatively little research done on this topic, especially since the period of transition begun in 1991. In large part, this is the result of data limitations. This gap in the literature has begun to be filled by important work by Ogloblin (1999) and Katz (2001).

The study of occupational segregation is important since the distribution of occupations has implications for the distribution of economic and social rewards such as income, material assets, class standing and prestige. It is also important because segregation can be a major source of labour market rigidity.

In this paper, quantitive evidence of changes in the gender differences in occupational distribution in Russia between 1994 and 2001 is presented. Specifically, the focus is on:

(1) extent women and men were employed in different occupations,
(2) trends in the distribution of occupations and trends in occupational segregation,
(3) empirical analysis of the degree to which women have entered traditional “male”
occupations and vice versa.

This paper’s contribution stems from its use of the country-wide dataset called the Russian Longitudinal Monitoring Survey (RLMS). This dataset contains detailed information on the labour market activity, including occupation of almost 4,000 individuals for the period between 1994 and 2001.

The structure of the paper is as follows. The next section documents the role of women and the process of occupational change in the Russian labour market during the Soviet period, perestroika (1985-1991) and the period of transition beginning 1991. Next, a brief summary of economic theories of occupational segregation and discrimination is presented. In the following section several measures of occupational segregation, along with a brief methodological discussion of the advantages and disadvantages of each are descibed. This paper relies on an analytical framework which distinguishes segregated, that is male- or female-dominated, and integrated occupations. The conclusion makes up the final section.

2. Background

It is generally accepted that inequality between men and women stems from attitudes, prejudices and assumptions concerning the different roles assigned to men and women in society. The role of women in the Soviet labour market was shaped by a distinctive set of ideological assumptions and historical developments that influenced Soviet policy prior to the collapse of the centrally planned economy in 1991 (Moghadam, 1993).

In Soviet Russia, women's full integration into the "social sphere of production" was an ideological principle, which concurred with the demand of the extensively growing economy for labour resources. Unemployment was not officially recognized, and being without a job brought criminal charges and imprisonment. Policies predicated on this principle resulted in a female labour force participation rate of almost 90%, exceptionally high by international standards and near that of male participation rate (Ogloblin, 1999; Brainerd, 2000). However, according to Moghadam (1993), the massive scale of female participation in Soviet economic life did not eradicate the differences between male and female workers in the distribution of income, skills, status and power.

Although the principle of equal pay for equal work regardless of gender was institutionalized through the Soviet Constitutions and Labour Codes, the publication in 1989 of the first Soviet figures on distribution of wages by age and gender confirmed that average full-time female earnings were only 65-75% those of males (Moghadam, 1993). This is the legacy of the Soviet period, where the centralized pay-scale policies gave preference to traditionally "male" sectors such as the heavy and extraction industries, construction and installation, and transport. Sectors with lesser national economic significance – the light and food industries, education, and health care, where female employment predominated – were markedly lower paid (Rzhanitsyna, 2001).

Another essential feature of the Soviet-type occupational segregation by gender was that the patriarchal stereotypes were institutionalized and sustained through official attitudes and policies. In spite of proclaimed gender equality, Soviet labour legislation always regarded women as a "specific labour force" because of their maternity and childcare function. Therefore, although women fully participated in working life, men occupied all levels of decision-making and top-level occupations (Ogloblin, 1999).

Despite the intellectual and political liberation initiated by Gorbachev's reforms during the perestroika period (1987-1991), the patriarchal order of society and official attitude toward women as a specific part of the labour force remained largely intact. Fundamental change in gender roles and in gender division of labour was never seriously proposed (Pasadskaya, 1994). Nonetheless, reform measures did relax the rigidity of the centralized wage system and weakened the ideological determinants of the gender pay differences and the first steps were made toward wage-setting conditions in which economic forces are an important factor determining the earnings gap between "male" and "female" jobs.

The transition from a centrally planned to a market economy in Russia that started in the early 1990s has led to profound changes in the country's labour market. On the vestiges of the highly centralized and rigid Soviet system of employment and wages, a rather decentralized and flexible structure has emerged. Russia's transformation into market economy has given rise to dramatic increases in wage inequality, a decline in female participation rates and the emergence of peculiar and widespread labour market institutions such as wage arrears and payment in-kind (Gerry, Kim *et al*, 2001).

The situation for many women has worsened over the transition period. The breakdown of the social infrastructure, the virtual absence of legal protection or collective agreements in the new private sector and the decline of wages in the public sector and many former state enterprises had a detrimental effect on women's employment and income. Discriminatory tendencies towards women increased, while opportunities of finding a new work place, career development, raising qualification and retraining reduced (Meshcherkina, 1999).

Significant structural changes have adversely affected employment sectors and job types where women have tended to predominate. These include sectors such as light industry, education, public health, and culture. Katz (2001) notes that the average decline in women's employment of 3 percentage points from 1990 to 1998 was unevenly distributed over sectors. According to the official Goskomstat statistics, during the reform years the structure of both male and female employment was essentially changed. Female labour was gradually pushed out of the high-paid industries. The share of women in the total number of employed in high-paid industries, such as finance and credit, state administration, communication, has decreased between 1992 and 1999 (Goskomstat, 1999). There is a tight correlation between rates of growth in salaries in a profession and the exclusion of women from that profession.

In the Soviet Union and in Russia today women often surpass men in their education. However, a high level of women's education was not accompanied by an equally high social status. Among managers, women were four times scarcer than men. The analysis of employment of female civil servants in the Executive Federal bodies also demonstrates that, although women constitute 55.9% of all officers, they make only 2.4% of the leading positions (Meshcherkina, 1999). A decreasing share of women among managerial staff forms the idea of the so-called "glass ceiling" in women's professional careers. Consequently, segregation of women to low-paid industries and to low status occupations leads to male and female labour being not equally renumerated.

There are different economic theories trying to explain the reasons of the existence and persistence of occupational segregation. Some of the theories are presented below.

3. Theoretical Approaches

Most of the research literature dealing with occupational segregation by gender is not concerned with occupational segregation per se, but with the effect it has on female-male pay differentials. For this reason, many theories and explanations treat the determinants of occupational segregation by gender and male-female pay inequality as if these phenomena were the same. This is unfortunate, since female-male pay differentials have many sources and occupational segregation by gender is only one of them. Furthermore, gender occupational segregation is an important topic in its own right (Anker, 2001).

In order to deal with the problems of occupational gender segregation it is necessary to understand the principal explanations for the existence and persistence of occupational segregation by gender. There are important policy implications to be derived from an understanding of why these differences exist. To the extent that they arise from unequal opportunities caused by unfair hiring or promotional practices, the economy has failed to make appropriate use of human resources and has created inefficiencies. In this case there is justification for intervention to facilitate greater gender equality in the labour market. On the other hand, to the extent that gender differences in occupational structure and earnings arise from differences in individual productivity or choice, despite equal labour market opportunity, interventions to change either employment or earnings patterns would lead to distortions in resource allocation and creation of inefficiencies (Filer, 1989).

Occupational segregation may be due to either supply- or demand-side factors or a combination of both. On the supply side, workers seek the best-paying jobs after taking into consideration their personal endowments (e.g. education and experience), constraints (e.g. childcare responsibilities) and preferences (e.g. pleasant work environment). The simplest way to explain occupational segregation would be to ascribe it to inherent differences in the preferences of men and women. An alternative explanation of occupational segregation, also starting an assumption that men and women are biologically and/or psychologically different, asserts that men and women have different innate abilities. This implies that each gender brings different endowments of human capital to the market. With this approach occupations are separated by the requirement of different skills (Jonung, 1998).

Supply-side effects can also be due to what has been labelled “societal discrimination” which occurs when women are socialized to enter traditionally female pursuits and/or face barriers to obtaining education and pre-job training in traditional male fields (Blau *et al*, 1998). Preferences and “natural talents” are also socially constructed through children’s upbringing and through a continuing lifelong process of adjusting to gender roles. This is what is termed pre-market discrimination, societal discrimination or gender role socialization (Jonung, 1998).

Among labour supply theories, the human capital explanation has received the most attention. Human capital explanation holds that women who adhere to traditional roles in the family will anticipate a shorter and more discontinuous pattern of labour market experience than men. They will have less incentive to invest in work-related education and on-the-job training and will thus select occupations where such human capital investments are less important. Gender differentials in human capital investments lead, in turn, to occupational segregation. In this view, female jobs pay less than male jobs because of their lower skills (Blau *et al*, 1998).

According to the "new home economics", the family is viewed as a production unit, trying to maximize its utility by producing final commodities for the family. The person who can receive the highest market wage will spend more time in market work, while the person most productive in household work will take the major responsibility in this area. If the same person is most productive in both areas, the division of labour will be determined according to comparative advantage. If women enter the labour market it should be for occupations with requirements compatible with the demands of work at home. However, the theory of human capital can not explain the occupational segregation by gender between occupations with similar amounts and type of human capital such as within the higher professions, within blue-collar work, between blue-collar work and service work (Jonung, 1998).

On the demand side, employers are concerned with two main factors in their hiring decision: productivity and labour cost. Occupational segregation could result from a genuine evaluation of productivity such as a better education profile. However, women are often considered to be higher-cost workers because of a number of supposedly higher, indirect labour costs associated with women workers.

To explain the persistence of occupational segregation by gender, neo-classical economists have developed a compensating differentials model which states that women prefer occupations with good working conditions as they wish to avoid unpleasant and dangerous working conditions and/or to have jobs with good fringe benefits. In these circumstances, the lower monetary rewards in typical "female" occupations are said to be partly explained by some "pay" being taken in non-wage forms (Anker, 2001).

Modern theorising about labour market discrimination began with Becker's "taste" theory of discrimination. Theories of discrimination assume that the supply and demand observed in the market and the resulting occupational and wage structure are affected by discrimination. Because of their prejudice, employers are said to sustain disutility, or cost when they hire someone from the group discriminated against. According to this theory, employers behave rationally when they hire fewer people from that group, since they can thus avoid such a "cost" (Anker, 2001). One implication of this is that in a competitive labour market discrimination by employers should eventually disappear, because it is economically inefficient. Second, even if some employers are predisposed to discriminate, the great overlap in the skills and preferences of individual men and women should make it likely that both genders would be substantially represented in every occupation (Anker,2001).

According to the institutional theory statistical discrimination is employed at the entry position The starting point of institutional theories is the assumption that institutions, such as unions and large enterprises, play an important role in determining who is hired, fired and promoted, and how much they are paid. These theories view occupational segregation as the result of employment adjustment to given wage relations. The lack of competition within the firms leaves room for discrimination (Jonung, 1998).

Statistical discrimination theory seeks to explain observed outcomes in the labour market on the assumption that there are differences, on average, in the productivity, skills and experience of distinct groups of workers. Further, there are high search and information costs associated with recruitment and promotion decisions. The statistical discrimination argument posits that employers, faced with imperfect information, use group-level characteristics (workers` gender) as an indicator of stability, productivity, and specialized aptitudes and as an inexpensive screening mechanism to fill jobs (Wells, 1998).

In conclusion, the consideration of each of the economics’ explanations of occupational segregation by gender and gender discrimination in the labour market is important for understanding possible reasons, existence and persistence of this phenomenon, because the explanations are not mutually exclusive and are not always easy to disentangle from each another. In the next section, the measurement of occupational segregation is discussed.

4. Measuring Occupational Segregation

There are a number of different measures which give a useful indication of the nature and extend of segregation across occupations by gender. Most research on occupational gender segregation has employed single number indexes which have the advantage of simplicity, as they condense into one number all variation in the distribution of jobs between men and women.

The Index of Dissimilarity (ID) is the index most frequently used and most frequently criticized. When applied to the labour market, this index is interpreted as measuring the percentage of the female (male) workforce required to shift between occupation categories to ensure that the distribution of females (males) is the same as that of males (females). The value assumed by the ID varies from a minimum value of 0, when there are no differences among male and female occupational distributions – complete integration — to a maximum of 1 or 100% — complete segregation.

ID can be written as:

$$ID = \frac{1}{2}\sum\left|\frac{F_i}{F} - \frac{M_i}{M}\right|, \qquad (1)$$

where F_i and M_i are the number of females and males, respectively, in the ith occupation and F and M are the total number of females and males, respectively, in the workforce (Karmel and MacLachlan, 1988).

There are several reasons for the prominence of the ID. First, the ID allows comparison of periods with different female labour force participation rates, since it is invariable to transformations on the sex ratio. Second, the ID has a rather simple and intuitive interpretation. Third, even if ID shows weakness in some respects, there is no alternative measure which is not weak in other respects. However, ID values are sensitive to changes in the overall gender composition of the labour force and changes in the occupation structure of the labour force (Oliveira, 2001; Kalter, 2000).

The Size-Standardised Dissimilarity index (D_s) is the absolute measure of segregation that controls for the effect of the occupational structure, using all occupations as if they were of the same size, computed over a fixed number of comparable occupational categories (Oliveira 2001).

The D_s is given by equation:

$$D_s = \sum_{j=1}^{J} \left\| \left[\frac{\left(\frac{F_j}{T_j}\right)}{\sum_{j=1}^{J}\left(\frac{F_j}{T_j}\right)} \right] - \left[\frac{\left(\frac{M_j}{T_j}\right)}{\sum_{j=1}^{J}\left(\frac{M_j}{T_j}\right)} \right] \right\| \cdot 100 \cdot \frac{1}{2}, \quad (2)$$

where $T_j = M_j + F_j$ total number of men and women in occupation j; the numerators $\left(\frac{F_j}{T_j}\right)$ and $\left(\frac{M_j}{T_j}\right)$ index the female and male proportions in occupation j; the denominators adjust such values on the proportions in the other occupations (Oliveira, 2001).

The Ds eliminates the impact of varying category sizes in a very radical way, and it is open to question whether this is a desirable procedure. While the D_s simply eliminates the impact of structural changes, the problem of the D_s is that it is a correct measure of both kinds of changes together (Kalter, 2000).

Karmel and MacLachlan (1998) propose a new index called the Karmel and MacLachlan index (KM), which is closely related to the ID that takes into account the relative size of the female and male workforce. This index measures the proportion of persons (as opposed to females or males) required to change occupations in order that the distribution of the female workforce is the same as that of males while keeping the occupational structure constant (Karmel and MacLachlan 1988). The important characteristic of the KM index is that if any redistribution of persons across occupations is required in order to eliminate segregation it will be associated with a constant occupational structure and any bias arising from differences in the size of the groups of interest is also eliminated (Lambert, Petridis *et al*, 1996).

The KM index is defined as:

$$KM = \frac{1}{T}\sum\left|(1-a)M_i - aF_i\right|, \tag{3}$$

where T represents total employment, F_i and M_i are defined as female and male employment in the i^{th} occupation, respectively; and a is the proportion of males in the overall workforce (Karmel and MacLachlan, 1988).

The advantage of the KM index is that a change in the index over time can be decomposed into a margin-free composition and mix effects, where the latter can be broken up into occupation, gender, and interaction effects (Watts, 2001).

Before decomposition of the KM three inter-related indices will be calculated. First, totals of the matrix at time 1 are transformed to those of the matrix at time 2:

$$\begin{aligned} &KM(a) = \left(\frac{1}{T_2}\right)\sum_{i=1}^{n}\left|(1-\bar{a})M_{i1} - \bar{a}F_{i1}\right|\left(\frac{T_{i2}}{T_{i1}}\right), where \\ &\bar{a} = \sum_{i=1}^{n} M_{i1}\left(\frac{T_{i2}}{T_{i1}}\right)/T_2 \end{aligned} \tag{4}$$

The subscripts *1, 2* refer to time periods 1 and 2. The index *KM(a)* is computed by proportionately adjusting the number of males and females in each occupation by the change in the employment level in that occupation from period 1 to period 2. The resulting male share of total employment is denoted by $\bar{a}$. Therefore, the initial gender composition of each occupation is maintained but the share of total employment in each occupation is adjusted to that prevailing in period 2 (Karmel and MacLachlan, 1988).

Second, similarly, the gender totals of the matrix at time 1 are transformed to those at time 2:

$$KM(b) = \left(\frac{1}{T_2}\right)\sum_{i=1}^{n}\left|(1-a_2)\left(\frac{M_2}{M_1}\right)M_{i1} - a_2\left(\frac{F_2}{F_1}\right)F_{i1}\right| = Z_{21}KM_1 \tag{5}$$

$$where \; Z_{21} = \frac{a_2(1-a_2)}{a_1(1-a_1)}$$

The index *KM(b)* is calculated by adjusting the numbers of females (males) in each occupation by the rise in total female (male) employment between the two time periods. Hence, the overall gender composition of employment corresponds to period 2, however, the size of occupations will be different from that of period 2 (Karmel and MacLachlan, 1988).

Third, the matrix at time 1 is transformed so that the occupation and gender totals are equal to those of the matrix at time 2. This can not be done analytically but must be done iteratively by alternately transforming the matrix to have occupation and gender totals equal to those of the matrix at time 2. *KM(c)* is obtained by successive transformations of the original distribution by the occupation and gender calculations detailed above. Therefore, after each iteration total employment corresponding to period 2 is gained but, after the odd iterations, individual occupation totals are realized whereas, after the even iterations, the period 2 gender totals are attained (Karmel and MacLachlan, 1988). The distribution is said to converge when the proportional error is less than 0.025% with respect to either the gender totals after the occupational transformation or the occupational totals after the gender transformation (Watts, 1998).

To decompose the change in the index we use KM_1 and KM_2 which denote the first and the second years value indices respectively, and the *KM(a)*, *KM(b)*, and *KM(c)* as the values corresponding to the three transformations described above.

The percentage (forward) composition effect, the impact of changes in the gender structure of occupations retaining the overall gender and occupational structures constant, is written as $100*(KM_2 - KM(c))/((KM_1 + KM_2)/2$. The forward percentage mix effect, the impact of changes in the occupational structure and the overall gender distribution, is formulated as $100*(KM(c) - KM_1)/((KM_1 + KM_2)/2$. This effect is subdivided into the occupation, gender and interaction effects. The percentage occupation effect, the impact of changes in the overall occupational structure, is written as $100*(KM(a) - KM_1)/((KM_1 + KM_2)/2$. The percentage gender effect, the effect of changes in the overall gender composition of the labour force, is written as $100*(KM(b) - KM_1)/((KM_1 + KM_2)/2$. Finally, the forward residual interaction effect is measured as $100*\left[(KM(c) - KM_1) - (KM(a) - KM_1) - (KM(b) - KM_1)\right]/((KM_1 + KM_2)/2$. Changes in the occupational structure of employment and the growth rates of males and females employment are interrelated; therefore, this effect is useful to define (Karmel and MacLachlan, 1988).

This decomposition procedure highlights the major difference between the KM and the ID. By definition, the gender component of the mix effect for the ID can be driven only by changes in the occupational structure (this is, the gender effect is zero), while the mix effect for the KM is affected by both changes to the occupational structure and changes to the overall gender mix (Karmel and MacLachlan, 1988).

An alternative index of gender job segregation is a new technique of Marginal Matching (MM). The MM was developed to measure changes in occupational segregation over time, ensuring sex composition and (gendered) occupation invariance. The mathematical expression of the MM is identical to that of the ID. The difference between ID and MM lies in the way "male" and "female" occupations are defined. MM defines "female" occupations as those that contain the same number of workers (male and female) as there are employed women in the labour force. This calculation is done by ordering occupations according to the female concentration in the occupations (MRSD, 2000). Occupations are ordered according to declining female share of employment, so that the most female dominated occupation appears first. The cumulative distribution of the employed labour force along this ordering starting at the "female" end of the occupational ordering and move along the cumulative distribution is calculated until the cumulative number of workers equals the number of women in employment. The level of female concentration at this point is the dividing point between "male" and "female" occupations. By doing so, marginal total of workers in "male" occupation and marginal total of workers in "female" occupation are respectively "matched" to number of male workers in total and number of female workers in total. Under marginal matching the cutting point between the two groups of occupations is varied to match the changing gender composition of employment over time, so that the corresponding two by two segregation table remains symmetric (Watts, 1995).

The marginal matching can be written as:

$$MM = \frac{F'_f}{F} - \frac{M'_f}{M}, \tag{6}$$

M'_f is the number of males in the "female" occupations, F'_f is the number of females in the "male" occupations, F and M are total female and male employment, respectively (Watts, 1995).

One unappealing aspect of the marginal matching approach is employment data by gender and occupation is suppressed by the construction of the two by two table (Watts, 1995).

The Gini Index (GI) and the associated Lorenz Curve are used extensively in the inequality literature to present estimates of the degree of inequality in some metric such as income. Gini index is another index which is gender symmetric. In recent gender segregation studies, the Loprenz Curve and the GI have been adapted to measure the degree of segregation. The Gini coefficient gives a single value measure of the area between the line of no segregation and the segregation curve (Harrison, 2002).

One of the equations for analysing the Gini index can be written as

$$G = 1 - \sum_{i=1}^{n} \frac{F_i}{F} \left(\frac{M_i}{M} + 2 \sum_{j=i+1}^{n} \frac{M_j}{M} \right), \tag{7}$$

where there are *n* occupations, *i* denotes the *i*th occupation and *j* denotes as occupation included in the cumulative total (*i* sums across all occupations and j sums across occupations (i+1) to the end). The Gini index is computed on data that arranged such that $\frac{F_1}{M_1} \le \frac{F_2}{M_2} \le ... \le \frac{F_n}{M_n}$ (Hutchens, 2000). The Gini index ranges between 0, where there is no concentration (perfect equality), and 1 where there is total concentration (perfect inequality).

The Gini index is not characterised by occupational invariance, which is the requirement that the measure of segregation should be invariant to changes in the relative size of occupations, if the gender composition of these occupations remains constant. Although it has a tendency to overweight the extreme component values (occupations with high proportions of one sex) it is otherwise very suitable (Blackburn *et al*, 2001)

5. Data

For this study data from the Russian Longitudinal Monitoring Survey (RLMS), a household based survey encompassing the Russian Federation, is utilised. The RLMS surveys were funded by the World Bank and by the United States Agency for International Development (US AID) as a panel survey in order to reveal the effects of economic transformation on the welfare of households and individuals.

The RLMS was conducted in two phases, with two separate panels. The first phase (rounds I—IV) covers the period from 1992 to 1994, and the second (rounds V—X) 1994 to 2001. A limitation of the RLMS is that data from rounds I—IV cannot be merged with rounds V—X data as different samples were used for the first and the second phases of the survey. Data from the second phase of the RLMS, used in this study, reflects the economy after the initial stage of the radical reforms, when price liberalization and privatization have been virtually completed.

Although the RLMS embraces a sample of randomly selected households from across Russia, population weights provided with the surveys are used in the empirical work so the results are nationally representative. There are around 10,000 observations in each round, of which around 4,000 individuals are employed. The study focuses on working-age individuals, women aged 25—55 years old and men aged 25—60.

Several studies have analysed recent labour market dynamics and gender wage differentials in Russia. Among them works by Glinskaya and Mroz (1996), Newell and Reilly (1996), Foley (1997), Ogloblin (1999), Gerry *et al* (2001), Mroz *et al* (2003).

6. Estimation of Russian Labour Market Characteristics and Occupational Distribution

Labour market characteristics

This section presents the standard labour force characteristics by gender in Russia between 1994 and 2001. The analysis shows that in the period between 1994 and 2001, the total rate of employment decreased by 2.5 percentage points from 75.2% in 1994 to 72.7% in 2001 (Table1). This change, however, masks a large drop in the employment rate in 1998. In 1998, the employment rate decreased to 69.3% from 73.1% in 1996, followed by an increase to 71.9% in 2000. The main possible explanation of a large drop in the employment in Russia in 1998 is a financial crisis in August 1998. The interviews of the RLMS survey were always undertaken in the last quarter of each year. Thus, the 1998-year survey was influenced by the financial crisis on the Russian labour market. The crisis began when the Russian government devalued its currency, the rouble, and declared a moratorium on its debt servicing on August 17th 1998. The country's payment system ceased to operate and a number of large banks were unable to meet their commitments to depositors. The effects on the Russian economy were immediate and devastating. The inflation touched 70.0%; unemployment increased to a high of 13.7%; real wages fell by 30% and GDP plunged by almost 5% for the year (Gerry & A Li, 2002: 2). The net effect on the labour market was profound. Employment and labour force participation declined and unemployment increased substantially during 1998.

Traditionally, the female employment rate is less than that of males. The smallest difference in employment between men and women was in 1998, when the employment rate was 67.7% for females and 70.7% for males. However, by 2001 the gap had risen to over 6.0% in 2001.

Between 1994 and 2001, the male employment rate decreased by 1.7 percentage points, from 77.7% in 1994 to 76.0% in 2001. In comparison, the employment rate for females dropped by 3.2 percentage points between 1994 and 2001. By 2001, the male employment rate stood at 76.0%, its highest level since 1996. Similarly, in 1998, female employment fell to its lowest level of 67.7% over the period. Although it rose slightly by 2001, in contrast to the male employment rate it remained below all levels reported in the RLMS for 1996 and earlier.

Table 1. Economically active population (Weighted), 1994—2001

	1994		1995		1996		1998		2000		2001	
	% of gender	% of total	% of gender	% of total	% of gender	% of total	% of gender	% of total	% of gender	% of total	% of gender	% of total
Employment rate, total	75.2		75.8		73.1		69.3		71.9		72.7	
Male	77.7	53.8	78.3	54.1	75.3	53.8	70.7	52.6	75.0	52.8	76.0	52.8
Female	72.5	46.2	73.1	45.9	70.8	46.2	67.7	47.4	68.8	47.2	69.3	47.2
Percentage difference	5.2		5.2		4.5		3.0		6.2		6.7	
Unemployment rate, total	7.8		7.6		9.4		11.3		8.7		7.5	
Male	6.8	43.6	6.7	43.5	8.8	46.3	11.2	50.4	8.6	50.0	7.6	49.7
Female	8.8	56.4	8.7	56.5	10.2	53.7	11.3	49.6	8.7	50.0	7.4	50.3
Percentage difference	-2.0		-2.0		-1.4		-0.1		-0.1		0.2	
Labour force participation, total	81.5		82.1		81.0		78.0		78.7		78.8	
Male	83.5	53.2	84.0	53.4	82.7	52.8	79.7	52.5	82.1	52.7	82.3	52.7
Female	79.5	46.8	80.2	46.6	79.3	47.2	76.3	47.5	75.3	47.3	75.3	47.3
Percentage difference	4.0		3.8		3.4		3.4		6.8		7.0	

Between 1994 and 1998, unemployment in the Russian Federation increased sharply. Over this period, the unemployment rate increased from 7.8% in 1994 to 11.3% in 1998. A characteristic feature of the unemployment rate after 1998 is that it decreased from its highest level of 11.3% in 1998 to 7.6% in 2001. The 2001 rate of unemployment represented its lowest level since the end of 1994. Over the period between 1994 and 2001, male unemployment increased from 6.8% in 1994 to 7.6% in 2001, while female unemployment declined from 8.8% in 1994 to 7.4% in 2001. Between 1994 and 1996, women had a higher unemployment rate than men. However, since 1998 female unemployment rate was almost equal to that of males. By 2001, men had higher unemployment rates than women.

The female participation rate conforms to the historical pattern; it is very high and also close to the male participation rate. According to Table 1, there was a decline in the labour force participation rate from 81.5% to 78.0% between 1994 and 1998. Although it increased in subsequent years, reaching 78.8% in 2001, it remained below the levels reported in the RLMS for 1996 and earlier. Between 1994 and 2001 the labour force participation rates for males are constantly higher than those for females. The gap between male and female labour force participation rates were approximately 3.5 percentage points between 1994 and 1998, before rising to almost 7 percentage points by 2001. This increase was due both to a rise in the labour force participation for men and a decline for women since 1998.

In conclusion, according to the analysis here, the labour force participation of women in the Russian labour market decreased between 1994 and 2001. Despite the fact that female unemployment declined from 8.8% in 1994 to 7.4% in 2001, at the same time, female employment fell from 72.5% to 69.3% over that period. This is an indication of female leaving employment not into unemployment but out of the labour force. The figures of the female labour force participation rate confirm it. The labour force participation rate for women dropped from 79.5% in 1994 to 75.3% in 2001. The question here is how much of this fall is associated with discouraged worker effect and how much with a females' choice not to work. The interpretation of this change in the position of women in the Russian labour market will be different in a situation when females choose to leave the Russian labour market comparing to a situation when the lack of employment opportunities and other barriers for women are present in the Russian labour market. According to Gvozdeva (2004), in the 1990s, young women, ages 20 to 29, left the labour force in larger numbers than young men because of the lack of suitable jobs. Many women moved from active participation in the labour market to the unpaid work of the households after losing hope of finding a job.

Occupational distribution

To characterize the distribution of the female employment the RLMS makes use of the ISCO-88 occupational ranking (Geneva: International Labour Organization, 1990). In this analysis, one-digit level occupational classification (ten occupational categories) was utilized. It is important to stress that this classification does not constitute a valid ordinal scale. For example, craft workers may well be more skilled, educated and highly paid than some clerks.

At the beginning, a descriptive analysis of women's and men's occupational distribution among ten main occupations is discussed (Table 2). In Russia, women and men have radically different patterns of employment by occupation. The RLMS data shows that between 1994 and 2001 almost half of the female workforce dominated the professional and technicians/associate professional occupations. In comparison, approximately two-thirds of the male labour force were clustered in craft, plant and machine operators occupational categories.

In 1994, more than one quarter of all employed women worked as professionals. However, the share of this occupational category in women's employment fell by 3.3 percentage points to 22.0% in 2001. Women's relatively high share in professional jobs is due to their predominance in teaching, nursing, social science, archivists and librarian jobs. The share of men employed in the professional occupational category is lower than that of women, decreasing from 14.0% of total male employment in 1994 to 8.4% in 2001. A high but declining share of all working women was also employed as technicians/associate professionals (22.0% in 1994 and 18.5% in 2001). The share of men in this group among all male labour force is on average 6.0%.

Females are also engaged frequently in service/shop and market sales occupations. The proportion of females employed in service/shop and market sales among all working women increased from 9.3% in 1994 to 14.0% in 2001. In comparison, the proportion of all males engaged in this occupational category remained relatively steady at approximately 3.5% between 1994 and 2001. It should be noted, however, that the category "sales worker" is very heterogeneous, ranging from store managers and reasonably well-paid employees to low-income pavement traders. Analysis and interpretation of figures concerning sales occupations must be treated with caution.

Table 2. Proportion within gender

Major occupational group	1994		1995		1996		1998		2000		2001	
	male	female	male	female	male	female	male	female	male	female	male	female
Legislators/ senior officials/ managers	2.8	0.9	5.2	2.9	1.7	0.9	4.3	2.9	8.0	5.2	8.7	6.7
Professionals	14.0	25.3	8.8	19.8	11.9	25.2	11.1	22.1	8.0	21.6	8.4	22.4
Technicians/ associate professionals	5.0	22.4	6.1	23.1	6.7	22.9	7.0	24.7	6.1	21.4	6.8	18.5
Clerks	0.8	11.5	1.4	11.7	1.2	12.4	1.2	10.1	0.8	8.6	1.2	9.3
Service/shop and market sales workers	3.2	9.3	3.8	10.7	3.9	10.1	3.6	10.3	3.8	14.1	3.7	14.0
Skilled agricultural/ fishery workers	1.1	0.1	0.7	0.0	1.2	0.3	0.9	0.1	1.1	0.1	1.1	0.1
Craft/related trade workers	29.7	6.9	27.6	5.1	29.2	6.2	23.8	5.1	23.2	4.3	22.5	4.2
Plant/machine operators and assemblers	30.1	6.3	29.8	6.6	32.0	7.8	28.5	7.4	30.0	6.5	27.6	7.4
Unskilled workers	7.4	13.5	7.3	14.9	9.7	13.9	10.2	11.6	8.3	11.6	8.7	10.5
Armed forces	1.2	0.1	2.2	0.4	2.6	0.4	1.8	0.3	1.9	0.3	1.2	0.2

In contrast to females a high proportion of males were employed in two major occupational categories of craft/related trade occupations and plant/machine operation/assembling occupations. In 1994 and 1996, approximately 29.0% of males and 6.0% of females were engaged in craft/related trade occupational category. In subsequent years, the proportion of both males and females in this occupational group decreased to around 23.0% for males and 4.0% for females. Males were also most commonly employed as plant/machine operators and assemblers. The proportion of both males and females who worked in this occupational category was relatively stable during all years (about 30.0% for males and 7.0% for females).

The RLMS data show that a proportion of females employed as legislators/senior officials/managers among all working females was larger in 2001 compared to 1994. It accounted for almost 1.0% in 1994 increasing to almost 6.7% of all working females in 2001. The share of males employed in this major occupational category increased from 2.8% of all working males in 1994 to 8.7% in 2001. However, between 1994 and 2001, an increase in the share of the legislators/senior officials/managers occupational category among all major occupational categories was not gradual. There were significant fluctuations of the share of this major occupational category for both males and females over this period. An increase in the share of the occupational category of legislators, senior officials and managers from 1.9% in 1994 to 3.6% in 1998 was mainly due to an increase in the share of general managers within it. However, further increase among employed in this occupational group in 2000 and 2001 was due to a rise in employment of corporate managers.

To complement this analysis, one can also look at the representation ratios (RRs). Coefficients of female representation are calculated for each of the major occupational groups by dividing the female share of employment in the occupation by the female share of total employment. Coefficients above one indicate female over-representation in the occupation; coefficients below one indicate female under-representation. The measure facilitates comparison across occupational groups which vary substantially in size, and across years. In addition, the ratios enable us to look at specific occupations and therefore provide greater insight into the qualitative aspects of segregation. Coefficients of female representation of 10 major occupational groups are shown in Table 3.

During all years, the occupational category with a substantial over-representation of women was clerks including secretaries, cashiers and tellers, material recording and transport clerks (RR of 1.9). The other occupational groups with female over-representation were technicians/associate professions (RR of 1.6), service/shop and market sales workers occupations (RR of 1.5), professional occupations (RR of 1.4) and unskilled occupations (RR of 1.2). Conversely, females were under-represented in armed forces (RR of 0.2), skilled agricultural workers (RR of 0.2), craft/related trade workers (RR of 0.3), and plant/machine operators and assemblers (RR of 0.3-0.4).

Between 1994 and 2001, female representation ratios declined slightly for some occupations such as technicians/associate professions (1.7 in 1994, 1.5 in 2001), clerks (2.0 in 1994, 1.8 in 2001) and elementary occupations (1.3 in 1994, 1.1 in 2001). Over the same period, female representation ratios increased in professional occupations (1.3 in 1994, 1.5 in 2001). Female representation in the legislators/senior officials/managers occupational category improved between 1994 and 2001. The RR for females in this major occupational category increased from 0.5 in 1994 to 0.9 in 2001. Unfortunately, due to a small number of interviewees who were engaged in legislative and managerial jobs among those who participated in the survey, it is hard to analyze this occupational category of a more detailed level.

Table 3. Representation ratios of females (1-digit level occupational classification)

Occupational categories	1994	1995	1996	1998	2000	2001
Legislators, senior officials and managers	0.5	0.7	0.7	0.8	0.8	0.9
Professionals	1.3	1.4	1.4	1.4	1.5	1.5
Technicians and associate professionals	1.7	1.7	1.6	1.6	1.6	1.5
Clerks	2.0	1.9	2.0	1.9	1.9	1.8
Service workers and shop and market sales workers	1.5	1.5	1.5	1.5	1.6	1.6
Skilled agricultural and fishery workers	0.2	0.0	0.3	0.2	0.2	0.2
Craft and related trade workers	0.4	0.3	0.3	0.3	0.3	0.3
Plant and machine operators and assemblers	0.3	0.3	0.4	0.4	0.3	0.4
Elementary occupations	1.3	1.4	1.2	1.1	1.2	1.1
Armed forces	0.1	0.3	0.2	0.3	0.2	0.3

To provide a deeper and structural description of the occupational segregation, the gender distribution within occupations was analysed. This analytical approach is based on men’s and women’s percentages within occupations (Table 4). Deciding which occupations are gender-dominated generally requires a decision to be made as to where the dividing lines should be drawn. The analysis presented here used segregation classification in which occupations are classified into five categories according to whether they are highly female- or male-dominated occupations (90-100% of the employed labour force are female or male); female or male-dominated occupations (60-90% of the employed labour force are female or male); or whether there are equal numbers of women and men in the occupation (40-60% of the employed labour force are of one gender, i.e., the share of women and men are roughly equal).

According to the analysis here, between 1994 and 2001, four out of ten major occupational categories were female-dominated. Those occupational categories include clerks, technicians/associate professionals, services/shop and market sales and professionals major occupational categories.

The clerks occupational category was the only highly female-dominated occupational category among 10 major occupational groups. Between 1994 and 2001, the share of women among clerks was approximately 90.0%. The largest share of females among clerks worked as customer service clerks (97.0% — 99.0%). Females also comprised between 85.0% and 92.0% of office clerks.

Between 1994 and 2001, females increased their representation among professionals from 61% to 71% and services and sales workers from 71% to 77%, while the share of women among technicians and associate professionals decreased from 80% to 71%. Highly female-dominated occupations within professional occupational category were nursing and midwifery, primary and pre-primary, and special education teaching professionals, as well as archivists and librarians. Women also dominated health, secondary education teaching, business and social science professional occupations, while among physical, mathematical and engineering science professional occupations males and females were distributed equally with a prevalence of males.

Next major occupational group with a significant share of female employment between 1994 and 2001 was technicians/associate professionals occupational category. Between 1994 and 2001, life science/health and teaching associate professionals occupations were highly female-dominated. Males dominated only among ship and aircraft controllers and technicians. At the same time, the proportion of women was declining within physical and engineering science technicians, safety and quality inspectors and finance and sales associate professionals.

Further, females dominated among service workers and shop/market sales workers. The share of women in the total number of employed in services major occupational category grew from approximately 70.0% between 1994 and 1998 to more than 77.0% in 2000 and 2001. Between 1994 and 2001, females dominated among models/salespersons/demonstrators. Personal and protective services sub-major occupational group which includes travel attendants, housekeeping/restaurant services workers, personal care workers and protective service workers, could be referred as integrated between 1994 and 1998. However, the share of females in it increased from 50.0% in 1998 to 65.0% in 2001. Females dominated in personal/protective services sub-major occupational category mainly because of the female prevalence among travel attendants, housekeeping and restaurant service, and other personal care and service workers. Protective service occupations were male dominated between 1994 and 1996 (the share of males accounted for more than 80.0%), becoming highly male occupation between 1998 and 2001 (the share of males increased to more that 90.0%).

In comparison, some types of work remained almost exclusively male including skilled agricultural/fishery, craft/related trades, plant/machine operation occupations, and armed forces. Among skilled agricultural and fishery workers women's share of employment remained almost unchanged at 9%. Within craft occupations the proportion of women decreased from 17% in 1994 to 14% in 2001. A slight increase in the share if women within plant and machine operation occupations from 15% in 1994 to 19% in 2001 can be explained by the negative growth of this occupational category during this period and by the decline of the number of men employed in it.

Because large occupational categories can mask underlying differences in employment by gender, it is important to examine data for more detailed occupations. For example, within plant and machine operation occupations men were concentrated in metal- and mineral-products machine operation occupations (80%), while women were in textile-, fur-, leather-products and rubber- and plastic-products machine operation occupations (95%). Among craft workers machinery mechanics and fitters were exclusively men (98%), while printing (on average 72%), food processing (on average 74%), and textile, garment workers (92%) were women.

The results of the analysis show that, in Russia between 1994 and 2001 the labour market was clearly gender segregated. However, occupational segregation (both the structure and concentration of labour force) diminished in some and increased in other occupations over this period. Between 1994 and 2001, more females became employed in former male-dominated occupations For example, the share of females increased in plant/machine operations occupational categories such as stationary-plant operation occupations from 34.0% in 1994 to almost 40.0% in 2001. However, it was rare for female-dominated occupations to become more male-dominated. The only exception to this was increased share of males among technical/associate professionals, mainly in finance and public trade. Finally, within many occupations the gender structure remained the same over the period. For example, occupations such as professional, clerks and service/shop and market sales occupations remained female-dominated, while agricultural/fishery and craft/related trades occupations and army remained male-dominated.

Table 4. Proportion within occupation

	1994		1995		1996		1998		2000		2001
Major occupational group	male	female	male	female	male	female	male	female	male	female	male
Legislators/senior officials / managers	77.9	22.1	67.9	32.1	68.9	31.1	61.9	38.1	62.7	37.3	59.2
Professionals	39.0	61.0	34.0	66.0	34.5	65.5	35.3	64.7	28.7	71.3	29.4
Technicians/ associate professionals	20.7	79.3	23.6	76.4	24.7	75.3	23.5	76.5	23.6	76.4	29.1
Clerks	7.0	93.0	12.1	87.9	9.6	90.4	11.3	88.7	9.0	91.0	13.0
Service/shop and market sales workers	28.9	71.1	29.5	70.5	30.1	69.9	27.6	72.4	22.6	77.4	22.9
Skilled agricultural/ fishery workers	92.3	7.7	100.0	0.0	84.6	15.4	88.2	11.8	90.5	9.5	90.9
Craft/related trade workers	83.3	16.7	86.3	13.7	84.0	16.0	83.6	16.4	85.4	14.6	85.8
Plant/ machine operators and assemblers	84.8	15.2	84.0	16.0	82.1	17.9	80.7	19.3	83.3	16.7	80.7
Unskilled workers	38.9	61.1	36.3	63.7	43.8	56.2	49.0	51.0	43.8	56.2	47.8
Armed forces	96.3	3.7	86.3	13.7	88.5	11.5	85.7	14.3	88.6	11.4	88.0
Total	53.7	46.3	53.8	46.2	52.8	47.2	52.1	47.9	52.2	47.8	52.7

Analysis of gender distribution between and within occupational categories has advantages compared to single number indexes. It allows identifying which individual occupations are male, female or integrated as well as the possible changes in gender-ratio of occupations. Moreover, one can describe what kind of work males and females concretely do in the labour market, how the workforce actually concentrates in the segregated and integrated occupations. In addition, the results express both the dimension of structure of occupational segregation (division of occupations by gender) and the dimension of occupational concentration of the labour force (division of labour force within segregated occupations). Nevertheless, to estimate the change in the level of occupational segregation and its trend between 1994 and 2001 the indices of occupational segregation need to be used.

7. Estimations of the Indices of Segregation

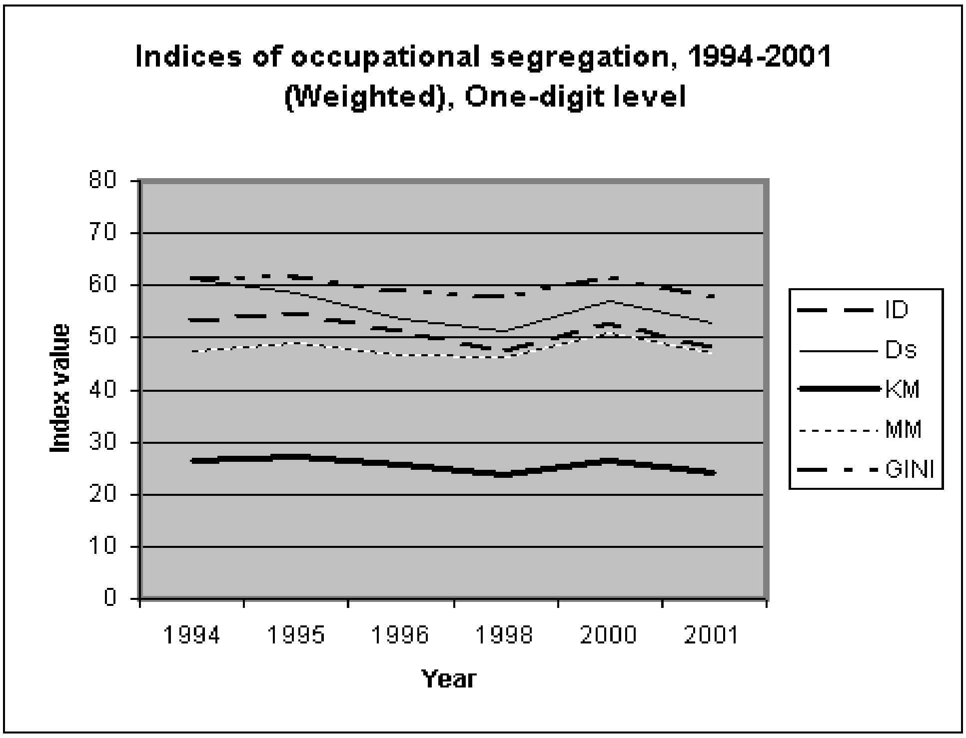
Five summary measures of segregation - the index of dissimilarity (ID), the size-standardized index (Ds), the Karmel and MacLachlan index (KM), the Marginal Matching (MM) and the Gini index (GI) calculated from the weighted RLMS data and based on the ten occupational categories (one-digit level) are presented in Table 5 and plotted in Figure 1.

Table 5. Indices of occupational segregation (Weighted), 1994-2001.
One-digit level

Index	1994	1995	1996	1998	2000	2001
ID	53.3	54.8	51.2	47.4	53.0	48.1
Ds	61.2	58.7	53.6	51.1	56.9	52.9
KM	26.5	27.3	25.5	23.7	26.5	24.0
MM	47.4	49.1	46.5	46.4	50.9	47.0
GINI	61.3	61.5	59.0	57.6	61.7	57.9

After 1995 all indices demonstrated similar trends in their changes. From 1995 to 1998 gender segregation was declining. The ID and Ds declined by about 7.5 percentage points, the KM, MM and Gini fell by about 3 to 4 percentage points. However, an increase in gender segregation was witnessed during 2000. The ID and the Ds increased by almost 5.5 percentage points, the MM and the Gini by more than 4 and the KM by almost 3 percentage points. In 2001, gender segregation continued to decline. The biggest change of 4.9 percentage points was revealed by the ID. The Ds, MM and Gini dropped by roughly 4 percentage points, while the KM by 2.5 percentage points.

Figure 1



The results received demonstrate the relation between the KM and the ID. In this analysis, the KM has values that equal half of the ID each year between 1994 and 2001. The KM represents a weighted version of the ID. The difference in size between the KM and the ID depends on the share of the females in the total labour force. Thus, the situation in which the KM equals half of the ID can be explained by the fact that the share of females in the total labour force equals that of males that is approximately 50%. That explains why the KM index has the smallest values of all.

Overall, the ID declined from 53.3% in 1994 to 48.1% in 2001 and the KM dropped from 26.5% to 24% during those years. The MM values were slightly lower than those of the ID and fell from 47.4% in 1994 to 47% in 2001, while the Ds and Gini values were somewhat the highest of all. The Ds declined from 61.2% to 52.9% and the Gini declined from 61.3% to 57.9% between 1994 and 2001. This difference in the results can be attributed to the difference in the definition of the indices explained earlier. The estimations show that in spite of the evidence of substantial segregation, integrative shifts occurred during 1994 and 2001.

In order to adjusticate between sources of change in levels of gender segregation over time, decomposition the KM is performed. The results of the decomposition of the change in the KM during the period from 1994 to 2001 are shown in Table 6.

Table 6. Decomposition of changes in the KM
(KM=26.5 in 1994 and KM=24 in 2001)
One-digit level

Cause of change	Change	Percent
Total	-9.9	100%
Composition effect	-11.1	112%
Mix effect	1.2	-12%
***gender**	0.4	-4%
***occupation**	0.2	-2%
***interaction**	0.7	-7%

The estimates of the KM (Table 5) reveal that the number of people who needed to relocate jobs to achieve zero segregation would have been 24% in 2001, as compared to 26.5% in 1994. A decline in the index value indicates that a smaller proportion of workers would need to change occupations in order to equate the female labour force with that of male, and hence a lower level of segregation. The computation of the decomposition of the KM reveals the factors which caused the decline in the occupational gender segregation for the 10 major occupations between 1994 and 2001.

At the one-digit level, the decomposition procedure calculated a total decline in segregation of almost 10%. The decomposition indicates that the decline in the level of gender segregation in Russia between 1994 and 2001 can be attributed to changes in the sex ratio of individual occupational categories, which is known as the composition effect. The composition effect for KM is large and negative, more than offsetting the tiny positive mix effect. It shows that the individual occupations became more integrated over time and that segregation across 10 occupational categories decreased by approximately -11%. This suggests that females had greater work opportunities and they could compete with males more equally. Its interpretation is that, after taking into account changes in the overall occupation and gender mix, about 11 percent fewer persons would have had to relocate jobs in 2001 than in 1994 in order to have equal female and male distribution across the overall occupational structure.

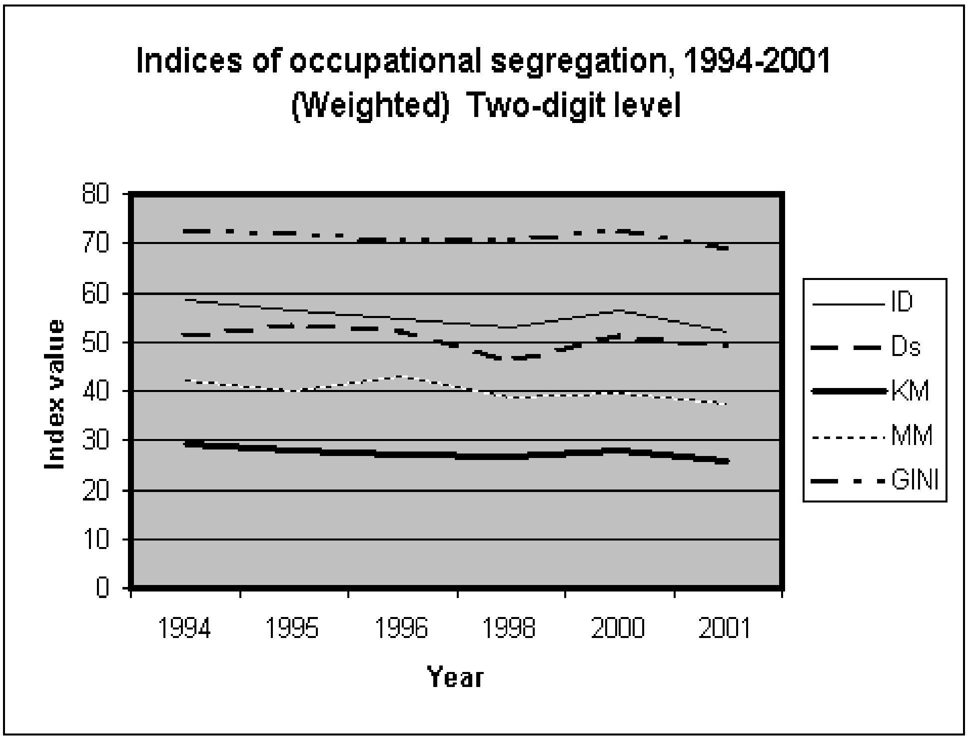
Changes in the occupational mix accounted for only 1.2% of the increase in occupational segregation. Gender effect comprised 0.4% of the mix effect. This indicates that the growth of the female employment was small. According to the Table 1, the share of female employment increased from 46.2% in 1994 to 47.2% in 2001. The changing distribution of employment across the whole occupational structure (occupation effect) had minimal impact (0.2%) as well. Change of the overall occupational structure is a long-run process which takes time to take place and to be captured by the occupation component of the mix effect (Watts and Rich, 1992).

Table 7 and Figure 2 illustrate the results for the same segregation indices using a two-digit level disaggregation of occupation codes. The indices' estimates demonstrate that regardless of the index used occupational segregation by gender fell between 1994 and 2001.

Table 7. Indices of occupational segregation (Weighted), 1994-2001
Two-digit level

Index	1994	1995	1996	1998	2000	2001
ID	58.4	56.4	54.5	53.0	56.2	51.9
Ds	51.4	53.3	51.8	46.0	51.0	49.0
KM	29.0	28.1	27.1	26.5	28.1	25.9
MM	41.9	39.9	42.9	38.8	39.4	37.4
GINI	72.4	72.0	70.7	70.4	72.5	68.9

Figure 2



All indices, other than the Ds, exhibit similar trends. Their estimates reveal that occupational gender segregation declined gradually from 1994 to 1998, then increased in 2000. In 2001 occupational segregation continued to decline and reached its lowest point over the whole period. Contrary to other indices, the Ds increased in 1995. This increase was followed by the decline in the index from 1996 to 1998 when its value was the lowest over the whole period. In 2000 the Ds rose, but similar to other four indices it declined in 2001.

The results for the measurement of gender occupational segregation using the KM indicate that over the period between 1994 and 2001 the level of segregation declined from 29% to 26%. The index values are higher for the two-digit level estimates than those for the one-digit estimates. Index calculated using two-digit level occupation codes indicated that the number of people who needed to change occupations to achieve zero segregation would have been 26% in 2001 as compared to 29% in 1994. In comparison, the results for the one-digit level occupation codes revealed that the number would have been only 24% in 2001, as compared to 26.5% in 1994. The index values are different, with those for the two-digit level occupational codes being somewhat higher, but the percentage change in the index between 1994 and 2001 was almost the same (around 10%) in both cases.

To reveal the factors that caused the change in the occupational segregation for the two-digit level occupational categories the decomposition of the KM is performed (Table 8).

Table 8. Decomposition of changes in the KM
(KM=29 in 1994 and KM=25.9 in 2001)
Two-digit level

Cause of change	Change	Percent
Total	-11.4	100%
Composition effect	-16.0	141%
Mix effect	4.6	-41%
***gender**	0.4	-3%
***occupation**	3.4	-30%
***interaction**	0.9	-8%

The total change in occupational segregation consisted of a composition effect of -16% and a mix effect of 4.6% (gender, occupation and gender/occupation effects). The occupation effect indicates that the decline in the level of occupational segregation was due to the change in the gender composition of individual occupations. Moreover, shifts in the gender composition of employment within particular occupations were the only cause of the decline in the occupational segregation. The interpretation of the composition effect for the two-digit level occupational coding is that almost 16% fewer people had to change occupations in 2001, compared with 1994, to equalize the distribution of males and females across the whole occupational structure. The composition effect for the two-digit level occupational categories was 5 percentage points higher than that for the one-digit level occupational categories. This suggests that the two-digit level occupational categories were more gender segregated as compared to the one-digit level occupational categories.

The positive mix effect (4.6%) is dominated by the occupation effect (3.4%) which demonstrates a larger change in the distribution of employment across two digit-level occupational categories than that across one-digit level occupational categories. Obviously, the gender effect (0.4%) for the two-digit level occupational distribution was very small and equalled the gender effect for the one-digit level occupational distribution. The small effect of the change in the rates of female and male employment was due to the small increase in the share of females in the economy.

8. Conclusion

With the number of jobs decreasing and tensions in the labour market rising, the idea of drawing women back into the family is gaining popularity in Russia. To a certain extent, this is the result of a natural desire to overcome the involuntary over-employment of women that was characteristic of the Soviet period. However, pushing the image of a woman-housewife has the hidden purpose of decreasing the labour supply by removing a large group of potential job-seekers from the labour market. Crowding out women is fraught with a host of negative consequences such as a denial of opportunities for self-fulfilment, decreasing per-capita income and correspondingly a potential impoverishment of families, and in the long run to a loss of a considerable part of the qualification potential of the population. In this way, promoting women's adjustment to the new employment situation and increasing their competitiveness in the labour market is a task of vital importance for Russia.

The experience of western economies has proved beyond any reasonable doubt that women can have a positive effect on economic growth, social harmony and political stability within society. Thus, the process of economic transformation in Russia should not result in excluding women from the sphere of economic activities pushing them to a low-paying jobs but should envisage their adjustment to the new conditions in the labour market. Women of the Russian Federation should be given every opportunity to realize their potential and make a valuable contribution to the prosperity of their families and their country.

REFERENCES

ANKER, R. (2001), 'Differences in occupational earnings by sex', in M. Loutfi (ed), *Women, gender and work: what is equality and how do we get there?*, International Labour Office, Geneva.

ASHWIN, S. (2001), 'Utverzhdenie muzhskoi identichnosti na rynke truda sovremennoi Rossii', *Rubezh (almanakh socialnykh issledovanii)* #16—17.

BENJAMIN, D., GUNDERSON, M. & RIDDELL, W.G. (1998), *Labour market economics*, McGraw-Hill, Ryerson, Toronto.

BERGMANN, B.R. (1971), 'The effect of white incomes on discrimination in employment', *Journal of Political Economy* 20(2), pp. 294—313.

BLAU, F.D. & JUSENIUS, C.L. (1976), 'Economists' approaches to sex segregation in the labour market: an appraisal', in Blaxall, M. & Reagan B. (eds), *Women and the Workplace*, University of Chicago Press, Chicago.

BLAU, F.D., SIMPSON, P., & ANDERSON, D. (1998), 'Continuing progress? Trends in occupational segregation in the United States over 1970s and 1980s', National Bureau of Economic Research, Cambridge, Working Paper #6716.

BOUTENKO, I.A. & RAZLOGOV, K.E. (1997), *Osnovnye tendentsii sotsialnogo razvitiia Rossii,1960-1995, (Recent social trends in Russia, 1960—1995)*, McGill-Queen's University Press, London.

BRAINERD, E. (1998), 'Winners and losers in Russia's economic transition', *American Economic Review* 88 (5), pp.1094—116.

BRAINERD, E. (2000), 'Women in transition: Changes in gender wage differentials in Eastern Europe and the former Soviet Union', *Industrial and Labor Relations Review* 54(1), pp.138—62.

BUCKLEY, M. (1989), *Women and ideology in the Soviet Union*, University of Michigan Press, Ann Arbor.

BUCKLEY, M. (1992), *Perestroika and Soviet women*, Cambridge University Press, Cambridge.

BUCKLEY, M. (1997), *Post-Soviet women: from the Baltic to Central Asia*, Cambridge University Press, Cambridge.

CLARKE, S. (1996), *Conflict and change in the Russian industrial enterprise*, Brookfield, VT, USA: E. Elgar, Cheltenham, UK.

CLARKE, S. (1996), *Labour relations in transition: wages, employment and industrial conflict in Russia*, Brookfield, VT, USA: E. Elgar, Cheltenham, UK.

CLARKE, S. (1998), *Structural adjustment without mass unemployment?: lessons from Russia*, Northampton, MA, USA: E. Elgar, Cheltenham, UK.

CLARKE, S. (1999), *The formation of a labour market in Russia*, E. Elgar, Cheltenham, UK.

DEUTSCH, R., MORRISON, A., PIRAS, C., & NOPO, H. (2001), 'Working within confines: occupational segregation by gender for three Latin American countries', Sustainable Development Department, Inter-American Development bank, Washington D.C., Technical Paper SOC-126.

DOLADO, J.J., FELGUEROSO, F., & JIMENO, J.F. (2003), 'Where do women work?: analysing patterns in occupational segregation by gender', *Annales d`Economie et de Statistique*, (Special Issue on Discrimination and Unequal Outcomes).

DUFWENBERG, M. & MUREN, A. (2002), 'Discrimination by gender and social distance', Department of Economics, Stockholm University, Working Paper.

DUSTMANN, C. & KUNZE, A. (2001), 'Male-female wage growth, occupational segregation, and technological change', Department of Economics, University College London, Working Paper.

FILER, R.K. (1989), 'Occupational Segregation, compensating differentials, and comparable worth', pp.153-176 in R.T. Michael, H.I. Hartmann, & B. O`Farrwll (eds), *Pay Equity: Empirical inquiries*, The National Academies Press, Washington, D.C.

FOLEY, M. (1997), 'Labor market dynamics in Russia', Economic Growth Center, Yale University, New Haven, Discussion Paper #780.

GERBER, T.P. (2002), 'Structural change and post-socialist stratification: Labor market transitions in contemporary Russia', *American Sociological Review* 67 (5), pp. 629—59.

GERRY, C.J, KIM, B., & A LI, C. (2001), 'The gender wage gap and wage arrears in Russia: Evidence from the RLMS', Department of Economics, University of Essex, Working Paper.

GLINSKAYA, E. & MROZ, T. (1996), *The gender gap in Russia from 1992 to 1995*, Mimeo, Carolina Population center, The University of North Caroline.

GOSKOMSTAT (1999), Rossiiskii statisticheskiy ezhegodnik, Moscow.

HAAVIO-MANNILA, E. & ROTKIRCH, A.(1996), *Women's voices in Russia today*, Aldershot, Hants, England.

HAKIM, C. (1996), *Key issues in women's work: female heterogeneity and the polarisation of women's employment*, Atlantic Highlands, NJ: Athlone, London.

HAKIM, C. (1998), *Social change and innovation in the labour market: evidence from the census SARs on occupational segregation and labour mobility, part-time work and student jobs, homework and self-employment*, Oxford University Press, Oxford.

HANSEN, J. & WAHLBERG, R. (2000), 'Occupational gender composition and wages in Sweden', IZA, Bonn, Germany, Discussion Paper #217

HARBRIDGE, R.& THICKETT G. (2003), 'Gender and enterprise bargaining in New Zealand: Revisiting the equity issue', *New Zealand Journal of Industrial Relations* 28(1), pp. 75—89.

HARRISON, J. (2002), 'Gender segregation in employment in Australia: a workplace perspective', Department of Economics, Murdoch University, Working Paper #186.

HELLERSTEIN, J.K, NEUMARK, D., & TROSKE, K.R. (1997), 'Market forces and sex discrimination', National Bureau of Economic Research, Cambridge, Working Paper #6321

ILO (1987), Resolutions adopted by the international conference of labour statisticians: resolution concerning the revision of the International Standard classification of occupations, adopted by the Fourteenth International Conference of Labour Statisticians (October-November 1987).

ILO (2003), International Standard Classification of Occupations ISCO-88.

JACOBSEN, J.P. (1997), 'Workforce sex segregation in developing countries: general patterns and statistical relationships for developing countries in the Western Hemisphere', in J.M Rives. & M. Yousefi (eds), *Economic dimensions of gender inequality: a global perspective*, Conn.: Praeger, Westport.

JOHNSON, G.E. & STAFFORD F.P. (1998), 'Alternative approaches to occupational exclusion', in I. Persson & C. Jonung (eds), *Women's work and wages: a selection of papers from the 15th Arne Ryde Symposium on Economics of Gender and Family*, New York: Routledge, London.

JONUNG, C. (1998), 'Occupational segregation by sex and change over time', in I. Persson & C. Jonung (eds), *Women's work and wages: a selection of papers from the 15th Arne Ryde Symposium on Economics of Gender and Family*, New York: Routledge, London.

JURAJDA, S. & PLANOVSKY, J. (2000), 'The gender wage gap and workplace segregation in late transition economies', Vienna, SOCO Project Paper #76.

JURAJDA, S. (2001), 'Gender wage gap and segregation in late transition', CEPR, IZA, WD, Working Paper #182.

JURAJDA, S. (2003), 'Gender wage gap and segregation in enterprise and the public sector in late transition countries', *Journal of Comparative Economics* 31(2), pp.199—222.

KALTER, F. (2000), 'Measuring segregation and controlling for independent variables', MZES, Mannheim, Working Paper #19.

KAPELIOUCHNIKOV, R. (2000), 'Russia's labour market: adjustment without restructuring', in R. Weichhardt (ed), *Economic developments and reforms in cooperation partner countries: the link between economics, security and stability*, Economic Directorate, Brussels, NATO.

KARMEL, T. & MACLACHLAN M. (1988), 'Occupational Sex Segregation-Increasing or Decreasing?', *The Economic Record* 64(186), pp. 187-95.

KATZ, K. (2001), *Labour in Transition: women and men in Taganrog, Russia*, Department of Economics, Stockholm University.

KAUFMAN, B. E. (1994), *Economics of Labour Markets*, Dryden Press, Fort Worth.

KIDD, M. P. (1993), 'Sex Discrimination and Occupational Segregation in the Australian Labour Market', *The Economic Record* 69(204), pp. 44—55.

KOLEHMAINEN-LINDEN, S. (1997), 'Occupation and career opportunities of women in female-dominated occupations', Ministry of labour, Helsinki, Labour Policy Studies #173.

KOZINA, I. (2002), 'Professional segregation: gender stereotypes on the labour market', *Sociological Magazine* 3, pp.126—136.

LAMBERT, S., PETRIDIS, R., & GALEA, J. (1996), *Occupational segregation and female employment over the business cycle*, Murdoch University, Perth.

LEA, S. (1997), 'Logistic regression and discriminant analysis', University of Exeter, Department of Psychology, Washington Singer Laboratories, UK, web-page.

LEHMANN, H. & WADSWORTH, J. (2001), 'Wage Arrears and the Distribution of Earnings in Russia', CERT, School of Management, Heriot-Watt University, Edinburgh, Discussion paper #02/02.

MENG, X. & MILLER P. (1993), *Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector*, Department of Economics, University of Western Australia, Nedlands, W.A.

MENG, X. (1998), ‘The economic position of women in Asia’, IIAS/IISG, Amsterdam, CLARA Working Paper #4.

MENG, X. (2001), ‘The informal sector and rural-urban migration - a Chinese case study’, *Asian Economic Journal* 15 (1), pp.71—89.

MESHCHERKINA, E. (1999), ‘Women in the Labour Market in Russia’, *Feminist Studies* (1999), pp.253—70.

MILLER, P.W. (1987), ‘The wage effect of the occupational segregation of women in Britain’, *The Economic Journal* 97(388), pp.885—96.

MINCER, J & POLACHEK, S. (1974), ‘Family investment in human capital: earnings of women’, *Journal of Political Economy* 82(2), pp.76—108.

MOGHADAM, V.M. (1993), *Democratic reform and the position of women in transitional economies*, Oxford University Press, Oxford.

MROZ, T., HENDERSON, L., BONTCH-OSMOLOVSKII M., & POPKIN B. (2003), ‘Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992—2002’, Report submitted to the U.S. Agency for International Development, Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina.

MANPOWER Research and Statistics Department (MRSD), (2000), ‘Occupational segregation: a gender perspective’, Paper #1/00, Ministry of MANPOWER, Singapore.

NEWELL, A. & REILLY, B. (1996), ‘The Gender Wage Gap in Russia’, *Labour Economics* 3, pp. 337—56.

NEUMARK, D. (1993), ‘Sex discrimination and women’s labour market interruptions’, National Bureau of Economic Research, Inc, NBER Working Paper #4260.

OGLOBLIN, C.G. (1999), ‘The gender earnings differential in the Russian transition economy’, *Industrial and Labor Relations Review* 52(4), pp.602—28.

OLIVEIRA, A.M.H.C. (2001), ‘Occupational gender segregation and effects on wages in Brazil’, Anais da XXIV IUSSP General Conference, Salvador, Working Paper.

PALAZ, S. (1999), ‘Determining the presence and extent of occupational gender segregation in the Turkish labour market’, METU, Ankara, Working Paper.

PILKINGTON, H. (1996), *Gender, generation and identity in contemporary Russia*, Routledge, New York.

POLACHEK, S. (1981), ‘Occupational self-selection: a human capital approach to sex differences in occupational structure’, *Review of Economics and Statistics* 63(1), pp. 60—69.

POSADSKAYA, A. (1994), *Women in Russia: a new era in Russian feminism*, Verso, New York.

ROSENBERG, C. (1989), *Women and perestroika*, Bookmarks, Chicago.

ROSSIISKII Statisticheskii Ezhegodnik (Russian Statistical Yearbook) (1999), Goskomstat of Russia, Moscow.

RZHANITSYNA, L. (1993), *Rabotaiushchie zhenshchiny v usloviakh perekhoda Rossii k rynku*, RAN, Moscow, Russia.

SANDUL, I. (2002), ‘Trying to break through the glass ceiling’, *The Russian Journal* #6.

SELIM, R. & ILKKARACAN, I. (2002), ‘Gender inequalities in the labor market in Turkey: differentials in wages, industrial and occupational distribution of men and women’, Istanbul Technical University, Istanbul, Working Paper.

SIGETO, T. (1998), ‘Dynamics of occupational segregation and the sexual division of labor’, SSM Research Series, Department of Sociology, University of Tokyo.

SPRIGGS, W.E. & WILLIAMS, R.M. (1996), ‘A Logit Decomposition Analysis of Occupational Segregation: Results for the 1970s and 1980s’, *The Review of Economics and Statistics* 78(2), pp. 348—55

SUNG, Y., ZHANG, J., & CHAN, C. (2000), ‘Gender wage differentials and occupational segregation in Hong Kong, 1981—1996’, *Pacific Economic Review* 6(3), pp 345—59.

WATTS, M. (1995), *The use and abuse of measures of occupational gender segregation: a review*, Department of Economics, University of Newcastle, Callaghan, NSW.

WATTS, M.J. (2001). ‘The Evolving Pattern of Occupational Segregation by Race and Gender of enlisted Personnel in the United States Armed Forces, 1984-1998’, *The Electronic Journal of the Center for Military and Strategic Studies*, Winter 2000/Spring 2001.

WEEDEN, K.A. (2002), ‘A Profiles of change: sex segregation in the United States, 1910-2000’, in M. Charles & D.B. Grusky, *Occupational Ghettos: The worldwide segregation of men and women*, Stanford University Press, Palo Alto, CA.

WEEDEN, K.A. & SORENSEN J.B. (2003), ‘A Framework for analysing industrial and occupational sex segregation in the United States’, in M.

Charles & D.B. Grusky, *Occupational Ghettos: The worldwide segregation of men and women*, Stanford University Press, Palo Alto, CA.

WELLS, T. (1998), 'Changes in occupational sex segregation during the 1980s and 1990s', Center for Demography and Ecology University of Wisconsin-Madison, CDE Working paper #98—14.

WHITFIELD & ROSS R. (1996), *The Australian Labour Market*, HarperEducational Publishers, Pymble, NSW, Australia.

WOOTTON, B.H. (1997), 'Gender differences in occupational employment', *Monthly Labor Review* (April 1997), pp.15—24.

ZVEGLICH, J.E. & RODGERS Y. (1999), 'Occupational segregation and the gender earnings gap', Asian Development Bank, Manila, Philippines, Working Paper #21.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 427—443

ALTERNATIVE SAMPLING DESIGNS SOME APPLICATIONS OF QUALITATIVE DATA IN SURVEY SAMPLING

Barbara Kowalczyk[1]

ABSTRACT

The objective of this paper is to present a possibility of application of some alternative sampling designs, which are based on ranked sets in market and consumer surveys. Ranking can be done both on the basis of the study or auxiliary variable. The author introduces and analyzes an example of sales estimation in pharmacies. Data for the survey are collected directly from the pharmacies' computers. Additional information used in the survey to implement the ranking is qualitative data about sales from questionnaires. The author also analyzes a survey aimed at estimation of the number of pharmacies in Polish towns of 10000-100000 inhabitants. An auxiliary variable used to implement the ranking is the number of inhabitants in the town.

Key words: survey sampling, market and consumer surveys, qualitative and quantitative data, order statistics, ranked sets.

1. Introduction

The need for statistical information seems endless in modern society. One of the most important manners of providing statistical data is survey sampling, that is, a partial investigation of the finite population. A sample survey costs much less than a complete enumeration, is less time consuming, and may even be more accurate than the complete enumeration (in large censuses there is usually a large number of non-sampling errors). Apart from all that, a partial investigation of the finite population is in many cases the only possible way of investigation.

In conducting a survey, our main purpose is to provide the most accurate assessments of population parameters under restrictions imposed on the cost, duration and organization of the survey. So the search of new methods that lead to

[1] Barbara Kowalczyk, Warsaw School of Economics, Institute of Econometrics, email: bkowal@sgh.waw.pl

the increase of the efficiency of the estimation is one of the most important goals of survey sampling theory. In general the increase of the efficiency can be achieved in two ways: by using composite estimators and composite sampling schemes. The latter one will be a matter of discussion.

In many environmental, social and economic studies the exact measurement of an element is very difficult (in terms of money, time, labor and organization). But there are real life situations where the variable of interest, although not easily measurable, can be relatively easily ranked (order) at no cost or very little cost. In other words, in many situations there is enough information to do ranking without quantifying the units. The ranking can be done on the basis of visual inspection, prior information, earlier sampling episodes or other rough methods not requiring actual measurement. If there is a related variable which is readily observable and can be easily ranked, and which we feel is correlated with the variable of interest, the ranking can also be done on the basis of this concomitant variable.

The standard example illustrating the matter under discussion is the following. If interest lies in estimating the mean height of trees, then measuring the height of the sampled trees could pose a problem, but it would be relatively easy to rank small sets of trees on the basis of visual inspection of their heights.

Similar examples can be found in economic studies. When we are inquiring a company about its sales or income, they are not very keen to cooperate. So to measure sales is rather difficult. But to conduct the ranking is often relatively simple. For instance, if sales of hairdressers' shops are considered, we can reasonably order the shops on the basis of the following statement. The better location the shop has in a city or town and the more employees engages the higher should be its sales.

However, in the above example, as in most other situations, the ranking without exact measurement may not be done perfectly, of course. This will also be a matter of discussion and as we will see in later sections errors in ranking should not discourage us.

So, when elements of a given set can be easily drawn and ranked with reasonable success by judgment, the sampling schemes based on ranked samples have promise of applications instead of standard simple random sampling. The ranked set sampling procedure is conducted by selecting n random samples of size n each, and ranking each element within each set with respect to the variable of interest. Then the actual measurement is taken from the element with the smallest rank from the first sample. From the second sample an actual measurement is taken from the element with the second smallest rank, and the procedure is continued until the element with the largest rank is chosen for actual measurement from the nth sample. There is a total of n measured elements, one from each ordered sample.

It is known that the estimator of the population mean or total using ranked set sampling RSS is more efficient than the one obtained by using standard simple random sampling SRS, i. e. estimator based on RSS gives more precise assessments. However, the precision of the estimator depends on target population, ranking accuracy and the set size.

For sets containing large number of units, ranking the *i*th lowest unit (other than the lowest and the largest) may sometimes pose a problem. In such situations we can use a sampling scheme, analogous to ranked set sampling, which is based only on the lowest and largest order statistics, called extreme ranked set sampling ERSS. The ERSS procedure involves randomly drawing *n* sets of *n* units (*n* is an even number). From the first set of *n* units the lowest ranked unit is measured. From the second set of *n* units the largest ranked unit is measured. From the third set of *n* units the lowest ranked unit is measured, and so on. From the last set the largest ranked unit is measured. ERSS procedure is definitely easier for implementation than the one based on all order statistics, i.e. RSS. But estimator of the population mean or total using extreme ranked set sampling ERSS is in many practical cases less efficient than the one obtained by using RSS. The problem will be discussed widely in later sections.

The ranked set method was first suggested by McIntyre (1952) as a way to estimate mean pasture and forage yields. McIntyre did not supply any mathematical theory to support his suggestion. Takasi and Wakimoto (1968) independently arrived at the same results and supplied the necessary statistical theory. McIntyre (1952) and Takahasi and Wakimoto (1968) assumed perfect ranking of the elements. Dell and Clutter (1972) considered the case in which the ranking may not be perfect i.e., may be done with errors. Stokes (1977) considered the situation when an auxiliary variable may be used to acquire a ranked set sample of the variable of interest. Stokes (1980) also proposed an estimator of the population variance based on a ranked set sample and (1986) gave a useful review of some theoretical aspects of RSS. Mutlak and McDonald (1990a and 1990b) developed the RSS theory when the experimental units are selected with size-biased probability of selection. Samawi et al., (1996) investigated the use of extreme ranked set samples ERSS for estimating the population mean. Yu and Lam (1997) proposed regression-type RSS estimators of the population mean.

Extensive research has been done on the problem. As far as its applications are concerned very much is known about such applications in environmental, biological and agricultural studies [see e.g. Halls and Dell (1966), Yanagawa and Chen (1980), Martin et al. (1980), Cobby et al. (1985), Jonson et al. (1993), Patil and Taillie (1993), Yu and Lam (1997)]. Still very little is known about applications in the economic field.

The major objective of this paper is to present a practical application of the ranked set sampling procedure in market and consumer surveys. To reach the aim, in Section 2 the author introduces in details the methods of ranked set sampling and extreme ranked set sampling. In Section 3 the author presents statistical properties of an estimator of the population mean obtained by using RSS and ERSS. Examples of application of sampling schemes based on ranked sets are given in Section 4. The author describes and analyses an example of sales estimation in pharmacies. It has to be mentioned that the method of data collection in this survey is exact (without non-sampling errors) and is done by a connection with pharmacies' computers (not by questionnaires). An additional information used in the survey is qualitative (grouped) data from questionnaires, which are known to be not exact because of non-sampling errors (e. g. intentional incorrect answers about sales, poor memory etc.). Despite of these errors the use of auxiliary information in form of qualitative data and implementation of RSS leads to the increase of the efficiency of the estimation and gives more precise assessments than the widely known and used simple random sampling. Also another example of a successful application of RSS method is presented. The author analyses the survey aimed at estimation of the number of pharmacies in Polish towns of 10000 – 100000 inhabitants. The variable of interest is the number of pharmacies and the auxiliary variable used to conduct the ranking is the number of inhabitants in the town. In this case RSS also proved to be superior to standard simple random sampling and gave more precise assessments.

2. The Method of Sampling

2.1. Ranked Set Sampling

The first step of ranked set sampling RSS procedure is to draw n random samples with n elements in each sample and rank them (without actual measuring) with respect to the variable of interest X. From the first sample we choose the element with the smallest rank. From the second sample we choose the element with the second smallest rank. We continue the procedure until the element with the largest rank from the nth sample is chosen. This procedure yields a total number of n elements chosen to be measured, one from each sample.

Let X_{11}, X_{12}, ..., X_{1n}, X_{21}, X_{22}, ..., X_{2n}, ..., X_{n1}, X_{n2}, ..., X_{nn} be independent random variables all having the same cumulative distribution function $F(x)$. The values X_{ij} for the randomly drawn units can be arranged as in the following diagram.

Set

$$\begin{array}{ll}
1 & X_{11} \;\; X_{12} \;\; X_{13} \ldots \; X_{1n} \\
2 & X_{21} \;\; X_{22} \;\; X_{23} \ldots X_{2n} \\
3 & X_{31} \;\; X_{32} \;\; X_{33} \ldots \; X_{3n} \\
 & \ldots \\
n & X_{n1} \;\; X_{n2} \;\; X_{n3} \ldots X_{nn}
\end{array}$$

Let $X_{i(1)}$, $X_{i(2)}$, ..., $X_{i(n)}$ denote the corresponding order statistics of X_{i1}, X_{i2}, ..., X_{in}. After ranking the units appear as shown below:

Set

$$
\begin{array}{ll}
1 & (X_{1(1)})\ X_{1(2)}\ \ X_{1(3)}\ \ldots\ X_{1(n)} \\
2 & X_{2(1)}\ (X_{2(2)})\ X_{2(3)}\ \ldots X_{2(n)} \\
3 & X_{3(1)}\ \ X_{3(2)}\ \ (X_{3(3)})\ldots\ X_{3(n)} \\
 & \ldots \\
n & X_{n(1)}\ \ X_{n(2)}\ \ X_{n(3)}\ \ldots (X_{n(n)})
\end{array}
$$

Then $X_{1(1)}$, $X_{2(2)}$, ..., $X_{n(n)}$ is the ranked set sample. The quantified units are taken in parenthesis.
The mean of the ranked set sample is denoted by $\overline{X}_{RSS}$, where

$$\overline{X}_{RSS} = \frac{1}{n}\sum_{i=1}^{n} X_{i(i)} \ . \tag{1}$$

It has to be emphasised that Xi(i), i=1,2,…,n are independent as contrasted with the usual order statistics which are correlated.

2.2. Extreme Ranked Set Sampling

The extreme ranked set sampling ERSS procedure involves randomly drawing *n* sets of *n* units (*n* is an even number). From the first set of *n* units the lowest ranked unit is measured. From the second set of *n* units the largest ranked unit is measured. From the third set of *n* units the lowest ranked unit is measured, and so on. From the last set the largest ranked unit is measured.

Using notation as in the previous section, the extreme ranked set sample is $X_{1(1)}$, $X_{2(n)}$, $X_{3(1)}$,..., $X_{n(n)}$. The quantified units can be also presented as below:

Set

$$
\begin{array}{ll}
1 & X_{1(1)}\ \ *\ \ \ *\ \ \ldots\ * \\
2 & \ \ \ \ *\ \ \ *\ \ \ *\ \ \ldots X_{2(n)} \\
3 & \ \ \ \ X_{3(1)}\ \ *\ \ \ *\ \ \ldots\ * \\
 & \ldots \\
n & \ \ \ \ *\ \ \ *\ \ \ *\ \ \ldots X_{n(n)}
\end{array}
$$

The mean of the extreme ranked set sample is denoted by $\overline{X}_{ERSS}$, where

$$\overline{X}_{ERSS} = \frac{1}{n}\left(X_{1(1)} + X_{2(n)} + X_{3(1)} + X_{4(n)} + \ldots + X_{n-1(1)} + X_{n(n)}\right). \tag{2}$$

3. Properties of Estimators Based on Ranked Sets

3.1. Estimator of the Population Mean Under Ranked Set Sampling

3.1.1. Perfect ranking

Let us assume for the moment that the ranking is perfect, i.e. it may be done on the basis of visual inspection or some prior information in such a way that there are no ranking errors. In this section we give the properties of the mean under ranked set sampling and compare it with the usual mean under simple random sampling. $\overline{X}_{RSS}$ has the following properties. Proofs can be found in Takasi and Wakimoto (1968).

1. The mean of the ranked set sample $\overline{X}_{RSS}$ is an unbiased estimator of the population mean μ.
2. The mean of the ranked set sample $\overline{X}_{RSS}$ is more efficient than the usual sample mean $\overline{X}$ under simple random sampling. In other words, when both estimators are constructed on the basis of the same number *n* of actual measurements, $\overline{X}_{RSS}$ has smaller variance than $\overline{X}$.
3. The variance of the estimator $\overline{X}_{RSS}$ is given by a formula:

$$Var(\overline{X}_{RSS}) = \frac{1}{n}\left(\sigma_X^2 - \frac{1}{n}\sum_{i=1}^{n}\left(E(X_{i:n}) - \mu_X\right)^2 \right), \tag{3}$$

   where σ^2 is the population variance and $X_{i:n}$ are usual order statistics of *n* quantifications.

4. The efficiency of the estimator $\overline{X}_{RSS}$ compared with the usual sample mean $\overline{X}$ based on simple random sampling is given by

$$\frac{Var(\overline{X}_{RSS})}{Var(\overline{X})} = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E(X_{i:n}) - \mu_X}{\sigma_X} \right)^2 \right) \tag{4}$$

3.1.2. Errors in Ranking

Accurate ranking is the most difficult part of the implementation of ranked set sampling procedure. The case in which ranking may not be perfect seems typical for many applications of the presented method. This situation was studied theoretically by Dell and Clutter (1972).

They considered the case in which the ranking might be done with errors, i. e. the quantified element from the *i*th sample may not be the *i*th order statistic, but rather the *i*th “judgment order statistic”. If we denote the “judgment order statistic” and the “true order statistic” by $X_{[i:n]}$ and $X_{(i:n)}$ respectively, then Dell and Clutter (1972) used the model

$$X_{[i:n]} = X_{(i:n)} + \varepsilon_i \tag{5}$$

in order to study the impact of ranking errors. They assumed that $X_{(i:n)}$ and ε_i are independent and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Dell and Clutter conducted many simulations and showed (simulation results) that the estimator based on a general presented above model for ranked set sampling with errors in ranking is an unbiased estimator of the population mean despite the presence of the ranking errors and that is more efficient than the usual sample mean $\overline{X}$ under simple random sampling. Analogous theoretical model when the underlying distribution of X is normal was considered by David and Levine (1972).

3.1.3. Ranking by an Auxiliary Variable

In many practical situations ranking by visual inspection or prior information is rather difficult or even impossible. So the ranking may be accomplished by means of some auxiliary (concomitant) variable *Y* that is relatively easily measured and is correlated with the variable of interest *X*.

To carry out the ranking *n* samples are drawn, each having *n* units. From the first sample of size *n*, the *X* associated with the smallest *Y* is measured. From the second sample of size *n* the *X* associated with the second smallest *Y* is measured. We continue this way until the *X* associated with the largest *Y* from the *n*th sample is chosen for measurement.

Obviously, the benefit derived by using this method should depend on how many information *Y* yields about *X*. If *X* and *Y* are independent, the estimator $\overline{X}_{RSS(a)}$ obtained on the basis of an auxiliary variable should be equivalent in precision to the usual sample mean $\overline{X}$ under simple random sampling. If the correlation between the two variables is 1 or -1 then the estimator should be equivalent in precision to $\overline{X}_{RSS}$, i. e. to the estimator which is calculated from a ranked set sample of actual order statistics of *X*.

Stokes (1977) showed the following. Let the regression of *X* and *Y* is linear

$$E(X \mid Y) = \mu_X + \frac{\rho_{XY}\sigma_X}{\sigma_Y}\left(Y - \mu_Y\right), \tag{6}$$

$$Var(X \mid Y) = \sigma_X^2\left(1 - \rho_{XY}^2\right) \tag{7}$$

and the standardized variates $(X - \mu_X)/\sigma_X$ and $(X - \mu_Y)/\sigma_Y$ have the same cumulative distribution function. Then the estimator $\overline{X}_{RSS(a)}$ obtained on the basis of an auxiliary variable is more efficient then the one $\overline{X}$ of standard simple random sampling and the relative efficiency is given by

$$\frac{Var(\overline{X}_{RSS(a)})}{Var(\overline{X})} = \left(1 - \frac{\rho_{XY}^2}{n} \sum_{i=1}^{n} \left(\frac{E(Y_{i:n}) - \mu_Y}{\sigma_Y} \right)^2 \right) \tag{8}$$

3.1. Estimator of the Population Mean Under Extreme Ranked Set Sampling

In practice, ranking a sample of large number of units and observing the *i*th ranked unit (other than extremes) may be a difficult task. Therefore, Samawi et al. (1996) introduced extreme ranked set sampling procedure ERSS based on extreme ranked units only. The method of selection was described in the previous section. Now we give statistical properties of the mean estimator $\overline{X}_{ERSS}$ based on extreme ranked set samples.

1. The mean of the extreme ranked set sample $\overline{X}_{ERSS}$ is an unbiased estimator of the population mean μ_X if the underlying distribution of X is symmetric.
2. The mean of the extreme ranked set sample $\overline{X}_{ERSS}$ is more efficient than the usual sample mean $\overline{X}$ in the case of uniform distribution.
   The two above theorems are theoretical results. Samawi et al. (1996) also gave some simulation results, which are the following.
3. The mean of the extreme ranked set sample $\overline{X}_{ERSS}$ is more efficient than the usual sample mean $\overline{X}$ in the case of normal distribution but less efficient than the mean of ranked set sample $\overline{X}_{RSS}$.
4. The mean of the extreme ranked set sample $\overline{X}_{ERSS}$ is less efficient than the usual sample mean $\overline{X}$ in the case of exponential distribution.

The most important property for practitioners is that estimator of the population mean based on ERSS is better than the usual simple random sampling mean in the case of uniform and normal distribution. There is no any analogous theorem for skewed distribution, which are very common for real populations. So one has to be very careful with using ERSS, opposite to RSS, which is more efficient for estimating population mean than simple random sampling without any restrictions imposed on the population distribution of the variable of interest.

4. Applications of Sampling Schemes Based on Ranked Sets in Market and Consumer Surveys

4.1. Sales Estimation

4.1.1. Description of the Survey

A subpopulation of 1000 Polish pharmacies was examined and the sales of the pharmacies were measured. Also other characteristics were investigated but for the purpose of this study we can limit ourselves to sales estimation only. Such a research is very expensive because it requires a connection with the pharmacies' computers to have the exact data. Majority of cost is spent on the permission to use pharmacies' data. This method of data collection is characterized by lack of non-sampling errors, which are common for questionnaires.

Compared with the exact measurement, when data are taken by a connection with pharmacy's computer, the cost of a questionnaire is insignificant. Apart from that, market research agencies often conduct questionnaire studies for many different purposes and possess questionnaire data. And we also possess questionnaire data about sales for 2329 pharmacies.

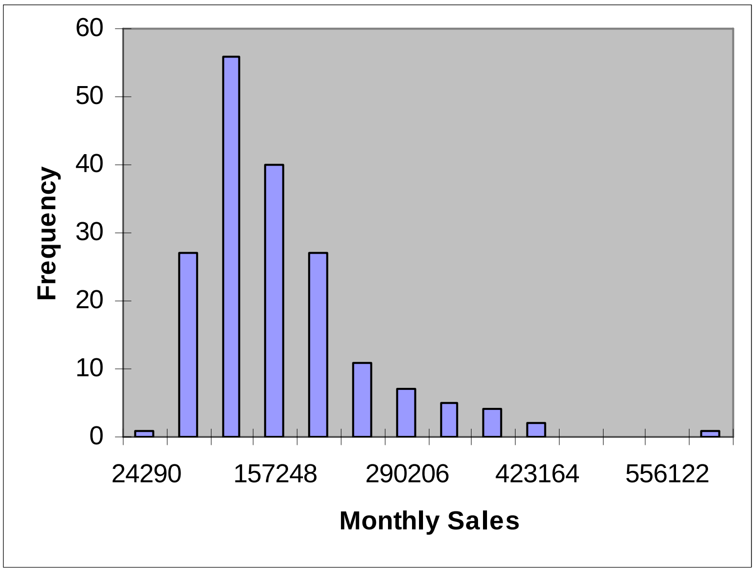
But the problem with questionnaires is that they are characterized by a considerable number of non-sampling errors. Almost all kinds of non-sampling errors are visible in questionnaire answers about sales: measurement errors (e.g. some items on the questionnaire may be poorly formulated and lead to misunderstanding on the part of the respondent), response errors (e.g. respondents may consciously or unconsciously give incorrect answers), telescoping errors (it is a tendency to allocate an amount that in actual fact belongs to a different period). The other problem with questionnaires is that we can only obtain on their basis qualitative data, more precisely grouped data. Because in practice in a questionnaire we can only ask "what bracket are you in?".

The problem to study is the following. We wish to investigate if questionnaire data can be employed to conduct ranking in sampling schemes based on order statistics (ranked sets) and if it still brings increase of the efficiency of the estimation. Two problems are connected with questionnaire data: they are burden with non-sampling errors and they are not quantitative, only grouped (qualitative) data – we cannot rank elements within classes.

4.1.2.Population Parameters

All data are used here with the kind permission of IMS Poland. The study variable is X – average monthly sales (in Polish zl). For a subpopulation of 181 pharmacies data about average monthly sales are available both from the exact measurement on a basis of computer data and from a questionnaire. For the purpose of this study 181 pharmacies are treated from now as a population from which a simulation study would be conducted. Population frequency distribution of X is the following.

Figure 1. Population frequency distribution of monthly sales



Population distribution is clearly asymmetrical (right-skewed distribution). Real population mean is $\overline{X} = 138385.65$ PLN. Data obtained from questionnaires are the following.

Table 1. Real sales vs questionnaire answers in %

	REAL SALES						
QUESTIONNAIRE ANSWERS	< 50 thous.	50-100 thous.	101-150 thous.	151-200 thous.	201-300 thous.	301-500 thous.	> 500 thous.
< 50 thous.	35.29	41.18	23.53				
50-100 thous.	5.56	54.17	29.16	2.77	4.17	4.17	
101-150 thous.		12.77	51.06	29.79	6.38		
151-200 thous.		11.11	11.11	51.85	25.93		
201-300 thous.		7.14	7.14		57.14	28.58	
301-500 thous.			50			50	
> 500 thous.		50					50

Generally in the questionnaire the true value of sales was given only by about 51.38% of inquired pharmacies. Let us analyze the fourth row of the table as an example. 51.85% of respondents who claimed to have their sales in the bracket 151-200 thousand were actually in the claimed bracket, almost 26 % of them underestimated their sales because their real sales were actually higher, 11.11% overestimated their sales and their real sales were actually lower, another 11.11% overestimated their sales even much more significantly because their real sales were less than 100 thousand.

4.1.3. Simulation Study and its Results

On the given population of N = 181 pharmacies we conduct a simulation study. The goal is to estimate on the basis of a sample the population mean, which real value is actually known and amounts 138385.65 zl. Thanks to this knowledge the possibility arises to compare different sampling schemes:

SRS – standard simple random sampling
RSS – ranked set sampling
ERSS – extreme ranked set sampling

In practice very often totals rather than means for domains, strata or population are needed. But when mean is assessed and number of units in a domain, strata or population is known the assessment of appropriate totals does not pose any problem. So for the purpose of this study without loss of generality we can focus on mean estimation only.

In the case of ranked set sampling and extreme ranked set sampling ranking is done on the basis of qualitative (grouped) data from questionnaires. All samples are of a size n =12.

5. Implementation of sampling schemes:

SRS – we randomly draw a sample of 12 different elements (pharmacies) from the population of 181 pharmacies. On the basis of this sample we calculate a sample mean as an estimate of the population mean by using real computer data.

RSS – from the population of 181 pharmacies we draw a sample of 12 elements. We rank the sample by using questionnaire data. When two or more pharmacies in a sample give the same bracket they are placed randomly within the given sample group. Next we choose for the final sample the element with the smallest rank, i.e. a pharmacy that give in a questionnaire the lowest bracket. To avoid repetitions we remove from the population the chosen pharmacy. So next we draw from the population of (181-1) pharmacies a sample of 12 elements. We rank the sample by using questionnaires and choose for the final sample the element with the second smallest rank and then we remove it from the population. We continue the procedure until we draw from the population of (181-11) pharmacies a sample of 12 elements, rank it by using questionnaires and choose for the final sample the element with the largest rank, i.e. a pharmacy that claims to be in the highest bracket. During all ranking procedures when two or more pharmacies in a sample give the same bracket they are ranked randomly within given bracket. We conduct actual measurements (based on real computer data not questionnaires) on the final sample of 12 elements only and calculate on their basis a sample mean as an estimator of the population mean total.

ERSS – from the population of 181 pharmacies we draw a sample of 12 elements. We rank the sample by using questionnaire data. When two or more pharmacies give the same bracket they are placed randomly within the given sample group. Next we choose for the final sample the element with the smallest rank. To avoid repetitions we remove from the population the chosen pharmacy. So next we draw from the population of (181-1) pharmacies a sample of 12 elements. We rank the sample by using questionnaires and choose for the final sample the element with the highest rank and then we remove it from the population. And so on. At the end we have the final sample of 12 elements, which are actually measured.

To compare different methods of estimation sampling is repeated 1000 times.

An estimate of the population mean obtained for the *j*th, $j = 1,2,...,1000$ sampling is denoted by t_j and the real population mean by $\overline{X}$. Simulation study results are presented in a table given below. The following notation is used:

- average absolute bias: $\frac{1}{1000}\sum_{i=1}^{1000}|t_i - \overline{X}|$
- MSE: $\frac{1}{1000}\sum_{i=1}^{1000}\left(t_i - \overline{X}\right)^2$
- variance: $\frac{1}{1000}\sum_{i=1}^{1000}\left(t_i - \bar{t}\right)^2$, where $\bar{t} = \frac{1}{1000}\sum_{i=1}^{1000}t_i$
- standard deviation: $\sqrt{\frac{1}{1000}\sum_{i=1}^{1000}\left(t_i - \bar{t}\right)^2}$

Table 2. Simulation Results

	SRS	RSS	ERSS
Average absolute bias of 1000 repetitions	18126.88	15899.11	26782.23
Average absolute bias in %	13.10	11.49	19.35
MSE	518332116	406010097	1117411514
Root mean square error	22766.91	20149.69	33427.71
Root mean square (relative standard) error in %	16.45	14.56	24.16
Variance	518305312	401734269	618514816
Standard deviation	22766.32	20043.31	24869.96
Standard deviation in %	16.45	14.48	17.97
Number of repetitions in which estimate gave the more accurate assessments	369	400	231

As it is seen from the result table RSS proved to be the most efficient sampling scheme for estimating population mean. Gain in precision compared with standard simple random sampling is respectively 2227.77 as far as average absolute bias is concerned, 2617.22 as far as relative standard error is concerned and 2723.01 as far as standard deviation is concerned. Relative gain in precision is $\frac{13.10-11.49}{13.10}100\% = 12.29\%$, $\frac{16.45-14.56}{16.45}100\% = 11.50\%$ and $\frac{16.45-14.48}{16.45}100\% = 11.98\%$ respectively.

6. Conclusions:

- Questionnaire data although qualitative (grouped) and burdened with errors can be employed for implementation of the ranking procedure.
- The use of ranked set sampling method leads to the increase of the efficiency of the estimation and gives more precise assessments than the widely used simple random sampling. The gain in precision despite of errors in ranking is achieved. The gain in precision is moderate but we have to remember that data from questionnaires are highly affected with errors (in almost 50%).
- When the population distribution is asymmetrical extreme ranked sampling method, although easier for implementation, should not be used. This method leads to the decrease of the efficiency as it is seen from table 2 [compare also section 3.2].

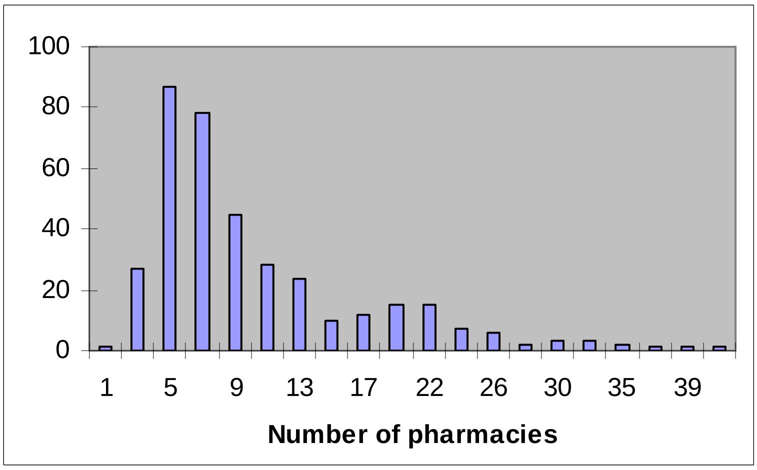
This was an example where additional cost of implementation of RSS was relatively small compared with actual measurements but it still existed. The next example is where there is no additional cost at all.

6.1. Estimation of the Number of Pharmacies

A domain of Polish towns of 10 000 – 100 000 inhabitants is taken into account. In Poland there are 368 such towns. Let us assume that we are interested in estimating the total number of pharmacies X in the domain and we are not able to measure the whole domain. We have to limit ourselves to a survey. So what can we do? We can select a sample of *n* elements by standard simple random sampling and measure the number of pharmacies in selected *n* towns. We can also select a sample of the same number of elements (towns for actual measuring) according to ranked set sampling. Ranking can be done on a basis of a concomitant (auxiliary) variable, which is correlated with the variable of interest *X* – number of pharmacies. A variable *Y*- number of inhabitants in a town should be an appropriate concomitant variable. The advantage of this choice is that this kind of auxiliary information can be obtained from the Central statistical Office and it does not require any additional cost.

Thanks to the IMS Poland database data about number of pharmacies in each town are obtained and on this basis we can conduct a simulation study which allows us to compare two different methods of estimation: based on simple sampling and on ranked set sampling. The real domain total is X = 3623 pharmacies. Domain correlation coefficient between the study X and the concomitant variable Y is $\rho = 0{,}9036$. Domain frequency distribution of a characteristic X is the following.

Figure 2. Population frequency distribution of the number of pharmacies



In section 3.1.3 theoretical statements about ranking by an auxiliary variable were given. Now we conduct a simulation study from a real population to illustrate the problem. From the given domain of 368 towns samples of 30 elements are drawn according to:

- SRS – standard simple random sampling
- RSS – ranked set sampling

Implementation of Sampling Schemes

SRS – we randomly draw a sample of 30 different towns from the population of 368 towns. We count the number of pharmacies in the selected towns and on this basis calculate a sample mean $\overline{X}$ and an estimate $X = 368\overline{X}$ of the population total.

RSS – from the population of 368 towns we draw a sample of different 30 towns. We rank the sample by using auxiliary variable Y and choose for the final sample the town with the smallest number of inhabitants. To avoid repetitions we remove the selected town from the population. Next we draw from the population of (368-1) towns a sample of 30 towns. We rank the sample with respect to auxiliary variable and choose for the final sample the element with the second smallest rank and then we remove it from the population. We continue the procedure until we draw from the population of (368-29) towns a sample of 30 towns, rank it by Y

and choose for the final sample the town with the largest number of inhabitants. We count the number of pharmacies only in the 30 towns from the final sample and calculate on this basis a sample mean $\overline{X}_{RSS}$ and an estimate $X_{RSS} = 368\overline{X}_{RSS}$ of the population total.

To compare two different methods of sample selection sampling is repeated 1000 times.

Table 3. Simulation Results

	SRS	RSS
Average absolute bias of 1000 repetitions	342.73	187.63
Average absolute bias in %	9.46	5.18
MSE	180779.79	57005.37
Root mean square error	425.18	238.76
Root mean square (relative standard) error in %	11.74	6.59
Variance	162786.46	49609.65
Standard deviation	403.47	222.73
Standard deviation in %	11.14	6.15
Number of repetitions in which estimate gave the more accurate assessments	319	681

As it is seen from the result table RSS proved to be much more efficient for estimating the total number of pharmacies in Polish towns of 10 000 – 100 000 inhabitant than standard simple random sampling. Relative gain in precision is very high and is equal to $\frac{9.46-5.18}{9.46}100\% = 45.24\%$ as far as absolute relative bias is concerned and $\frac{11.74-6.59}{11.74}100\% = 43.87\%$ as far as relative standard error is concerned.

The implementation of ranked set sampling procedure on the basis of auxiliary variable leads to the significant increase of the efficiency of estimation. Average absolute bias, root mean square error and standard deviation decreased in the case of RSS almost twice. The gain in precision is very high and, what is important, was gained in this example at no additional cost.

REFERENCES

COBBY, J. M., RIDOUT, M. S., BASSET, P. J. AND LARGE, R. V. (1985), AN Investigation into the Use of Ranked Set Sampling on Grass and Grass-Clover Swards, Grass Forage Sci. 40, 257—263.

DAVID H.A. and LEVINE D.N. (1972), Ranked Set Sampling in the Presence of Judgement Error, Biometrics, 28, 553—555.

DELL, T.R. and CLUTTER, J.L. (1972), Ranked Set Sampling Theory with Order Statistics Background, Biometrics 28, 545—553.

HALLS, L. S. and DELl, T. R. (1966), Trial of Ranked Set Sampling for Forage Yields, Forest Sci. 12 (1), 22—26.

JOHNSON, G. D., PATIL, G. P. and SINHA, A. K. (1993), Ranked Set Sampling for Vegetation Research, Abstracta Botanica 17,87—102.

MARTIN, W. L., SHARIK, T. L., ODERWALD, R. G. and SMITH, D. W. (1980), Evaluation of Ranked set Sampling for Estimating Scrub Phytomass in Appalachian Oak Forests, Publication No. FWS-4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and state University, Blacksburg, V. A.

McINTYRE, G. A. (1952), A Method of Unbiased Selective Sampling, Using Ranked Sets, Australian J. Agricultural Research 3, 385—390.

MUTLAK, H. A. and McDONALD, L.L. (1990a), Ranked Set Sampling with Size Biased probability of Selection, Biometrics 46, 435—445.

MUTLAK, H. A. and McDONALD, L.L. (1990b), Ranked Set Sampling with Respect to a Concomitant Variables and with Size Biased Probability of Selection, Commun. Statist. — Theory Math. 19 (1), 205—219.

PATIL, G. P. and Taillie, C. (1993), Environmental Sampling Observational Economy, and Statistical Inference with Emphasis on Ranked set Sampling, Encounter Sampling and composite Sampling, Bulletin ISI Proceedings of 49-th Session, Firenze, 295—312.

SAMAWI, H. M. MOHMMAD, S. And Abu-Dayyeh, W. (1996), Estimating the Population Means Using Extreme Ranked Set Sampling, Biom. Journal, 38, 577—586.

STOKES, S.L. (1977), Ranked Set Sampling with Concomitant Variables, Comm. Statist. Theory Methods A 6, 1207—1211.

STOKES, S.L. (1980), Estimation of Variance using Judgment Ordered Ranked Set Samples, Biometrics 36, 35—42.

STOKES, S.L. (1986), Ranked Set sampling. In: S. Kotz et al., eds., Encyclopedia of Statistical Sciences, Wiley, New York, 585—588.

TAKAHASI, K. and WAKIMOTO, K. (1968), On Unbiased Estimates of the Population Mean Based on the Sample Stratified Means of Ordering, Ann. Inst. Statist. Math. 20, 1—31.

YANAGAWA, T. and CHEN, S. H. (1980), the MG Procedure in Ranked Set sampling, J. Statist. Plann. Inference 4, 33—44.

YU. P. L. H. and LAM, K. (1997), Regression Estimator in Ranked Set Sampling, Biometrics, 53, 1070—1080.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 445—455

THE ASSESSMENT OF THE INVESTMENT ACTIVITY OF ENTERPRISES BASED ON THE BUSINESS TENDENCY SURVEYS

H. Sękowska, M. Święcka, K. Walkowska, I. Zagoździńska[1]

ABSTRACT

This paper presents the Polish CSO's business tendency surveys comprising investment activity of manufacturing industry and construction enterprises for years 1999-2004 when those surveys were fully harmonized with the European Commission's standards.

Investment survey includes, inter alia, size and directions of changes in investment activity as well as motives and sources of its financing. The results of the surveys confirm the notion of still difficult situation of Polish enterprises in the sphere of investment activity, however, larger units as well as those with foreign capital, are in mach better situation.

The main objective of investment activity is modernization of possessed fixed assets, to be able to compete on international markets in the wake of the EU's accession, while beginning of new investment is not so important thus far. The accuracy of enterprises' assessments of their investment activity, compared with proper indicators of quantitative statistics, is good in indicating the main directions of changes and predicting the turning points of growth cycles in Poland. Nevertheless, to assess future directions of development of investments in situation of Polish economy in transition and volatile general economic situation of the country was more difficult. The quality of gathered results is much better for a shorter term forecasts respectively.

Introduction

The collection by statistics of data on inclination of economic units to undertake investment activity and the size of those investments is one of the basic elements that allow to assess the situation of enterprises, their present financial situation and position on the market as well as their future possibilities of

[1] Central Statistical Office, Warsaw, Poland; E-mail: H.Sekowska@stat.gov.pl; K.Walkowska@stat.gov.pl ; I.Zagozdzinska@stat.gov.pl

development and strengthening their position in comparison with competing entities.

In the system of the Polish official business statistics, surveys on investment activity are conducted quarterly and yearly. The inclusion of questions from this field to the business tendency surveys was aimed at the supplementation of the quantitative data with new elements as well as collection of chosen data leading qualitative statistics.

The business tendency surveys deliver first assessment on the direction of changes of investments outlays made by enterprises in the previous and present year. For instance, primary yearly data on investment outlays are available from the qualitative statistics in April of the next year while from the business tendency surveys — one month earlier. Much more earlier first signals of the tendencies of the investment activity are obtained after the first and the third quarter of the given year.

The business tendency surveys are also a source of information on the forecasted investment activity in the present and future year so they could be used to the elaboration of the short term forecasts of the economic growth of the country.

Eventually in the business tendency surveys there are gathered data that are not collected by the quantitative statistics, for instance on factors influencing decisions on investments.

Moreover, the European system of investments surveys that is used in the business tendency surveys (wording of questions, timeliness and calculation of indicators) allows to make comparisons of results elaborated for Polish enterprises with results published for other countries.

The factors mentioned above cause that the assessments of entrepreneurs of present and future investment activity collected in the business tendency surveys become more and more important in analyses of present and future economic situation. There is no doubt that the possibilities of the use of those results will increase with the lengthening of time series of the collected data.

Such more and more wider use of the results of the business tendency surveys could be observed in Poland in the recent years, since the basic time series of information collected in those surveys are long enough to be used in the assessment of the present and future situation of enterprises (for three sectors of economy that is manufacturing industry, construction and retail trade time series are more that ten years long). There were several CSO papers on the contents and comparability of those data with results of adequate qualitative statistics, among others (Walkowska and Zagoździńska 1999, 2003), (Sękowska and Zagoździńska 2002), (Walkowska 2002).

At the present stage of those surveys it is interesting whether and on what scale the results of the surveys on investment activity of enterprises could be used in analyses of economic situation. The aim of this paper is to present the basic results on investment activity of Polish manufacturing and construction enterprises collected in the business tendency surveys and assessment of their usefulness for needs of analyses of present and future situation of enterprises.

Methodology of investment surveys in business tendency surveys

Business tendency surveys are conducted by Central Statistical Office since June 1992. As the first one, the business tendency survey in manufacturing industry was undertaken, one year later — surveys in construction and retail trade. Among other questions in quarterly questionnaires of manufacturing industry survey and monthly questionnaire of construction survey there were questions on investment activity of enterprises. In industry questionnaire they concerned directions of changes of investment outlays in a given quarter as well as investments forecasted for the future year and sources of their financing. In construction survey enterprises were asked if they intend to undertake the investment activity in the coming half of the year and what will be the sources of financing of this activity. The collected results were regularly published in *Business Tendency Survey in Manufacturing Industry, Construction and Retail Trade* — Quarterly Bulletin CSO and used for the needs of the country analysis.

In 1999 within the harmonisation of the Polish business tendency surveys with the European system modification of content and timing of investment survey in the business tendency surveys was made. In spite of questions asked in the quarterly (for industry) and monthly (for construction) surveys separate bi-yearly questionnaire was introduced. Those questionnaires are sent to the units taking part in the business tendency surveys in manufacturing industry and construction. In the system adapted by CSO the sample includes enterprises with 10 and more persons employed (surveys do not include so-called micro-enterprises). New questionnaire contains questions harmonised with the *Joint Harmonised European Union Programme of Business and Consumer Surveys* (European Economy No 6/1997, European Commission, 1998) and additionally two Polish questions asked already to respondents in the previous system: on kinds of investments and on sources of their financing. Entrepreneurs respond on questions according to the timing recommended by the European Commission i.e. in the first ten days of March (four questions) and November (7 questions). Results of surveys are available according to the kind of activity conducted by the enterprise (NACE classification) and according to the size classes (what allows to assess the situation of small, medium and large enterprises).

Because of the change of the wording of questions asked to the enterprises on their investment activity as well as timeliness and the horizon of the prognostic questions made in 1999, in this paper only results for the last five years are presented in *Business Tendency Survey in Manufacturing Industry, Construction and Retail Trade* — Current Information- Initial Data.

Size of the investment activity from the point of view of entrepreneurs

Before the results obtained in the business tendency surveys will be analysed it is worthwhile to show what was the overall economic situation of Poland at that time. In the table below there are shown three indicators presenting this situation: indices of the gross domestic product, domestic demand and gross fixed capital formation.

Table 1. Chosen yearly economic measures (in constant prices, previous year = 100)

Specification	1998	1999	2000	2001	2002	2003
Gross domestic product	104,8	104,1	104,0	101,0	101,4	103,8
Domestic demand	106,3	104,8	102,8	98,4	100,8	102,4
Gross fixed capital formation	114,2	106,8	102,7	91,2	94,2	99,1

Source: Statistical Bulletin, CSO, Warsaw 2004

This data show that in period analysed in this paper Polish economy did not develop regularly. In the first four years the worsening of the overall economic situation of the country and the lowest indicators were noted down in 2001. Since 2002 gradual improvement of the situation began. It is interesting how those tendencies were submitted by enterprises participating in the business tendency surveys in manufacturing industry and construction. Those assessments are one of aims of this elaboration.

The first, primary information on situation of enterprises on the investment activity collected from the business tendency surveys are data on the percentage of units undertaking or intending to start this kind of activity.

In this field the results obtained for manufacturing industry and construction enterprises show that in the period at issue the year 2001 was the most difficult for the Polish entrepreneurs. In manufacturing industry almost 20% of surveyed units did not undertake investment activity, in construction — even twice as much. In the following three years this percentage has been diminishing (to 10% in manufacturing industry in March 2004 and 24% in construction). It is characteristic at the same time that the assessments made at the beginning of the year are worse than assessments from the end of the same year. It might indicate that in reality from 2002 the situation of the Polish enterprises in the year comes out to be better than it was predicted by enterprises what influences the diminishing of the pessimism of respondents.

Second characteristic occurrence is much worse situation of small and medium enterprises in comparison with the situation of big ones. In 2001 every second small enterprise in manufacturing industry did not plan investment activity. In the group of medium enterprises the percentage was 30% while in the group of large ones it did not exceed 15%. In 2004 no investment activity was planned by 37% of small enterprises, 13% of medium and only 4% of big ones.

Analogical results for construction also confirm much worse situation of small units. In this sector of economy in the most difficult period every second small firm did not intend to invest, 38% of medium enterprises, 27% of large ones. This data additionally show that situation of construction enterprises was worse than manufacturing industry units. Its improvement in the following years was also slower than in industry because at the beginning of 2004 43% of small units did not intend to invest, 19% of medium and 17% of large ones.

In manufacturing industry the highest percentage of units that do not intend to invest is among enterprises producing wearing apparel (55%), furniture (40%), wood and products of wood (32%), radio, television and communication equipment (32%), textiles (31%), other non-metallic mineral products (30%). The improvement noticed in the following years was the most significant in units producing radio, television and communication equipment (only 2% of units did not intend to invest in 2004), wood and products of wood (12%) and furniture (15%). The highest percentage of enterprises that did not invest is for enterprises producing wearing apparel (38%).

First three questions of the business tendency survey questionnaire deliver data on the assessment of the percentage change of the investment outlays. Those questions concern three points in time:

- In the first question — posed in March of the year n — the respondents write down indices of the changes of those outlays in the year (n-1) in comparison with the year (n-2);
- In the second question — posed twice a year in the year n i.e. in March and November — the respondents assess the indices of the changes of those investments in the year n in comparison with the year (n-1);
- In the third question — posed in November of the year n — the respondents assess the forecasted size of outlays for the year (n+1) in comparison with the year n.

Answers obtained on the first of those questions are the first information on investments in the year (n-1), preceding– as it was stated before — the quantitative statistics. Final indices of investment outlays in manufacturing industry and construction are shown in the table below:

Table 2. Indices of the investment outlays in the national economy (previous year = 100)

sector	1998	1999	2000	2001	2002
manufacturing activity	117,9	92,6	90,8	88,5	92,8
construction	125,6	103,7	107,4	83,7	93,8

Source: Fixed assets in national economy in 2002, CSO, Warsaw 2003

Data collected from the business tendency surveys confirm the presented above directions of trend of macroeconomic situation. They indicate that in the first years of the period at issue investment outlays were increasing more and more slowly. In manufacturing industry the lowest increase was observed in 2001 and then this tendency turned round. In construction slow increase has been observed one year earlier. It should be reminded that the business tendency surveys do not include the smallest units (micro-enterprises) that are much more significant in construction than in manufacturing industry and at that time their situation was clearly worse than that of the bigger ones. That is why the results above do not show the situation in the whole manufacturing industry and in the whole construction to the same degree.

Answers of the enterprises on that question could be also treated as an assessment of the correctness of replies given on the second question. The respondents' assessments of changes of investment outlays in the year n estimated in the following years, in November of that year in most cases were lower than indicators presented half a year earlier that is three months after the end of the reporting year. It could be explained by the significant size of the investment activity in the last quarter of each year, but the scale of this occurrence shows that the situation of the Polish enterprises is not yet stabilised.

Analysing answers to those questions it is worth to pay attention to the relation between assessments of the respondents made in March and November of a given year. In manufacturing industry first of those assessments — closer to the forecasts in their character — were usually higher than the second ones. It also confirms other dependencies between prognostic and diagnostic assessments in other fields of the business tendency surveys: higher optimism of entrepreneurs while formulating their forecasts than the later realisation of the indicator. March assessments should be then interpreted with being aware of the dependencies mentioned above.

It is also interesting that in construction there is inverse dependence: correction of the first March assessment presented in November was more optimistic. Higher prudence of those assessments could be caused by the seasonal character of the works conducted by the construction enterprises and the fact that real yearly situation of the examined units is more difficult to assess.

Data collected as answers for the third question concerning investments forecasted for the future year seem to show that it is difficult for the entrepreneurs to forecast the situation in turning points. In the analyzed period it particularly concerns forecasts formulated in manufacturing enterprises for the year 2003. Those forecasts came out to be much more prudent than the real size of the investment activity undertaken later. In case of construction, noted down period prudent forecasts was longer what probably was connected with the longer difficult period in this sector of the economy.

At the end of this part of the elaboration the differences noted in the investment activity by enterprises of different size classes should be also underlined. Larger units clearly formulate better assessments of their activity in this field what confirms their better situation and position on the market than those of small units. The latter ones — in manufacturing industry as well as in construction — are those that submit the further decrease of their investment activity in comparison with the previous year even for the year 2004.

Factors influencing decisions on investment activity

Answers on questions on factors influencing inclination of enterprises for investing give interesting information on behaviour of entrepreneurs in comparison with their present situation and present and future overall economic situation.

In manufacturing industry main indicators influencing decision on undertaking the investment activity are the forecasted level of demand for produced goods and technical factors: development of technology, adaptation to technical requirements specified in proper regulations, increase of qualifications of labour force.

In case of demand, assessments made by entrepreneurs are derivated of the overall economic situation of the country in the recent years: diminishing demand in the years 1998-2001 justifies decrease of importance of this factor noted in the business tendency surveys tendency of manufacturing industry enterprises. The next year brought clear reversion of that tendency. This direction of changes has been noted down particularly in large units, and according to the kind of activity — mainly in units producing furniture, units producing wood and wood products (providers of materials for the latter ones), chemicals, chemical products and man-made fibres, other non-metallic mineral products and in enterprises conducting publishing activity.

Importance of the demand increased also in construction in the last year. It is caused by the similar tendency observed in enterprises producing other non-metallic mineral products of which construction is the main recipient. It is self-evident for construction that the forecasted increase of demand influenced more the investment activity of small and medium enterprises than those of the large ones.

Second of the most important factor — technical factor — became also of much more importance in the last year of the period at issue. The reasons for this occurrence should be found not only in increasing demand but also in expected by entrepreneurs accession of Poland to the European Union and connected with this harmonisation to the European technical technological norms as well as eagerness to cope with international competition and entrance on the Union markets. It is confirmed by the fact that this is the most underlined factor by producers of food and beverages that is those for which the implementation of the European technological and sanitary norms was the most important before 1 May 2004.

In construction importance of technical factor for all examined period is on similar and significant level.

Aims of investment activity

Important information on present situation of enterprises and on possibilities of their future development is answer on question on aims of planned investment activity.

In the examined period among investing enterprises positively more units undertake activity aimed at modernisation of already possessed assets than at new investments.

In manufacturing industry a little more than 80% of enterprises bear expenditures on modernization. New investments were planned in 2004 by about 70% of units while this percentage was increasing systematically since 2002 (in 2001 it amounted to the lowest — 45% — level in the whole period at issue). This tendency confirms reported above gradual improvement of the situation of manufacturing industry enterprises in the last few years.

Also in this case collected data show faster improvement of the situation of large units in comparison with small and medium ones. In 2004 new investments were planned by 80% of large units, 65% of medium and only 50% of small ones.

In particular sectors of industry in 2004 the highest percentage of units planning investments aimed at modernisation was among enterprises producing food and beverages, chemicals and chemical products, other non-metallic mineral products, metals and metal products, electrical machinery and apparatus. Only in three divisions of industry more units plan new investments than modernising investments: producers of wood and wood products, electrical machinery and apparatus and radio, television and communication equipment. New investments will constitute the most significant percentage for producers of wood and wood products, chemicals and chemical products and radio, television and communication equipment.

In construction — where overall situation of enterprises is worse than in manufacturing industry — modernising investments were undertaken by about 70% of units in the whole examined period. New investments were started by less enterprises till 2003 but this difference was gradually becoming smaller and in 2004 totally disappeared.

Also in this field of economy the development plans of the greatest enterprises were more optimistic then of the smaller ones. First of all it is shown by higher percentage of units planning modernising investments in the first group (in 2004 there were 60% of small units, 67% — medium and 72% large ones) as well as new ones (accordingly — 53%, 66% and 80%). It should be also noticed that in the group of big units the interest in new investments increases more visibly than the interest in the modernising investments.

The supplement to the information on the aims of investment activity presented above are answers on subsequent question of the questionnaire harmonised with the system passed by the European Commission. They confirm that both manufacturing industry and construction enterprises (total and divided to the size classes) plan the exchange of used or disused machinery and apparatus more often than introduce automatisation, mechanisation or new techniques.

Sources of financing investment activity

Both in manufacturing industry and construction enterprises dominant source of financing investment activity are own resources of enterprise. They are indicated by more than 90% of respondents. It should be also underlined that this source is dominant in all groups of enterprises independently from size class and kind of activity.

The second source is domestic bank credit but it is used by less than 40% of enterprises in both sectors of economy. In this case this source is a little more important in larger and stronger financially units to which banks grant credits more willingly.

Foreign credit is only a marginal source of financing investment activity both for manufacturing and construction enterprises. In the latter ones it is used mainly by units from divisions to which the foreign capital poured in, including producing wood and products of wood, chemicals and chemical products and motor vehicles.

Those results are confirmed by the quantitative statistics. In the total amount of investment outlays borne by the enterprises the dominant are own sources of investor and domestic bank credit that significantly prevail over foreign bank credit.

Table 3. Financing of investment outlays, financial elements of fixed assets and intangible assets in units with more than 50 employees — sources of financing (in %)

Year	total investment outlays	of which:		
		own sources of investor	domestic bank credit	foreign bank credit
1999	100	61,3	16,8	2,9
2000	100	69,3	14,3	3,1
2001	100	66,2	17,7	2,2
2002	100	71,3	11,8	3,3

Source: own calculations based on CSO data *Fixed assets in national economy in 2002*, CSO, Warsaw 2003

Conclusions

To sum up first it should be stated that the analysis made with use of data on investment activity gathered in the business tendency surveys indicates that it is a good, additional source of information on situation of enterprises and general economic situation. Respondents are able to assess size and directions of present and future investment activity with good approximation. They had some difficulties with this assessment in a not stabilised general economic situation of the country, especially in turning points. That is why forecasts of investment activity formulated by entrepreneurs (similarly as other forecasts) are more accurate for shorter periods on present stage of economic development of Poland.

Second conclusion coming from the analysis concerns situation of enterprises in presented sections according to size classes and kind of activity. Both in manufacturing industry and construction situation of smaller units is significantly more difficult than of large ones. It causes that they are not able to develop their activity on expected scale, it is more difficult for them to find sources of investments and to improve their position towards competition.

A significant percentage of larger units undertake investment activity and those are more often new investments that should lead to implementation of new production techniques than modernisation, they have also easier access to bank credits to finance their investment activity.

In sections divided according to kind of activity it is obvious that units connected with foreign markets, especially those with foreign investors are stronger. It should be also kept in mind that in the first phase of inflow of foreign capital to Poland investors were mainly interested in the strongest units with perspectives of future development also on international scale.

REFERENCES

WALKOWSKA, K. and ZAGOŹDZIŃSKA, I. (1999), *Analysis and Causes of Microeconomic Disequilibria Based on Business Tendency Survey Results*, Selected Papers submitted to the 23rd CIRET Conference 1997 in Helsinki, CIRET 1999.

SĘKOWSKA, H. and ZAGOŹDZIŃSKA, I. (2002), *Situation of the Polish manufacturing industry based on qualitative business tendency surveys,* Surveys on Polish economy. Present state and the perspectives of development, Warsaw School of Economics, Warsaw 2002.

WALKOWSKA, K. (2002), *Survey on situation of Polish construction based on results of business tendency survey conducted by Central Statistical Office*, Surveys on Polish economy. Present state and the perspectives of development, Warsaw School of Economics, Warsaw 2002.

WALKOWSKA, K.and ZAGOŹDZIŃSKA, I. (2003) *Present state of the business tendency surveys conducted by the Central Statistical Office*, paper prepared on EU Workshop on recent developments in Business and Consumer Surveys, Brussel 2003.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 457—474

AGGREGATION OF EXCHANGE RATE DATA AND LONG MEMORY MEASURES

Ewa Marta Syczewska[1]

ABSTRACT

Presence of long-memory in the series is an important feature to take into account in applied work. There are several methods and indicators of the long-memory, based on nonparametric and parametric methods (Hurst indicator, estimates of fractional integration parameter with periodogram regression of GPH type, ARFIMA and FIGARCH models, etc.)

We compare results of applying three of them (the Geweke and Porter-Hudak method, the modified rescaled range statistic, the Robinson's periodogram regression) to time-series data of exchange rates, with logarithms of daily data, weekly and monthly averages, and to daily, weekly and monthly returns. The series in question are the daily average exchange rates of the National Bank of Poland.

Results presented show that results of the long-memory estimation depend on aggregation of data.

***Key words*:** exchange rate, long memory, periodogram regression, fractional integration, aggregation.

1. Aim and structure of the paper

We present here results concerning testing of possible long-memory behaviour of a time series[2]. This task is important from applied econometrician point of view, e.g., in financial econometrics, as long-memory effects influence not only behaviour of a series, but also statistical properties of various statistics, used to test, e.g., financial market efficiency, to formulate forecasts etc. We compare results of three methods, used to detect or reject the presence of a long-term dependence in the series:

---

[1] Warsaw School of Economics, Institute of Econometrics, Al. Niepodległości 162, 02-554 Warsaw, Poland, E-mail: Ewa.Syczewska@sgh.waw.pl

[2] The first version of this paper was presented at the 27th CIRET Conference, Warsaw, September 2004, Session: "Economic Situation in Countries of Central and Eastern Europe".

- The modified rescaled range statistics, introduced by Lo (1991),
- The Geweke and Porter-Hudak periodogram regression method,
- The modified Robinson's periodogram regression.

As our main interest is in searching a piece of advice for a practitioner, we apply these methods to an actual set of data, namely the PLN exchange rates versus over twenty currencies, published by the National Bank of Poland. We use logarithms of daily data, weekly and monthly averages, and also logarithmic returns both for daily data and their weekly and monthly averages.

The results presented here are part of an ongoing research, part of which would be comparison of several other methods (wavelet estimates of fractional integration parameter, Philips methods, Hurst exponent etc) for exchange rate series.

In section 2, a short description of the three methods and date used are presented. Section 3 contains results of comparison of the methods for three level of aggregation. Section 4 concludes.

2. Methods and data

Due to importance of long-memory behaviour, there are many methods and tests of detecting this feature. Robinson (2003) remarks: "Indeed, the long memory literature spans the range from abstract probability theory, through mathematical statistics, to empirical studies in many areas of science, besides economics" (p. 1). One of earliest indicators of long-memory behaviour is the rescaled range statistic, introduced by Hurst (1951), later refined by Mandelbrot (1972) and others, with an important modification introduced by Lo (1991). Several methods are based on estimation of so-called fractional integration parameter, d. Among them, relatively easy to use and implemented in several econometric and statistical packages is the Geweke and Porter-Hudak (1983) method. Due to its deficiencies, a modification have been proposed by Robinson (1995). Another approach is based on maximum likelihood estimation of ARFIMA or FIGARCH models.

In earlier empirical work, concerning fractional integration parameter estimation for Polish zloty exchange rates, the Geweke and Porter-Hudak estimators and a wavelet transformation method was applied to daily and monthly data for a group of exchange rates. The results obtained suggested that it is worthwhile to check whether results for a particular variable are sensitive to a level of aggregation of that variable. In Syczewska (2002), both the GPH method and the wavelet transform method were applied to estimation of the fractional integration parameter for the USD/PLN exchange rates. In case of logarithms of daily data, the estimates range from 0,954 to 1,093, depending on estimation methods and subsample used. In case of logarithmic returns, d estimates ranged from –0,068 to 0,039. For the wavelet transform, the values of d ranged from –0,036 to –0,018 and shown smaller variance. In Syczewska (2004), the Hurst exponent, estimates of the fractional integration parameter, computed for a subsample and recursively, were applied to the exchange rates of the Hungarian forint, Latvian lat and Polish zloty. Hence it is worthwhile to check the fractional integration parameter estimates for greater set of exchange rates. We present here results of the Geweke and Porter-Hudak method, and also of the modified rescaled range of Lo and Robinson's method, for a set of actual PLN exchange rate daily data, for their weekly and monthly averages, and returns.

We use daily data on average nominal exchange rate, published by the National Bank of Poland. The available sample of the daily nominal exchange rates data covers period 4th January 1993 — 31th May 2004, i.e. 2878 observations. This set of daily data contains the Austrian shilling, Australian, Canadian and US dollar, Luxemburg, Swiss and French franc, Czech korona, German mark, Danish krona, Spanish peseta, Finnish mark, British and Irish pound, Greek drachma, Hungarian forint, Italian lira, Japanese yen, Portuguese escudo, Swedish and Finnish korona, SDR, and of course the Euro exchange rate. All the rates are expressed as number of the Polish zlotys per unit, or in several cases, per 100 units of the foreign currency. Due to joining the Euro zone by most of European countries, several series are shorter than others.

To extend the series for the Euro backwards, we use the euro conversion rates (see Table A1 in Appendix) to derive values of an hypothetical exchange rate, computed as an average of the earlier values of several currencies.

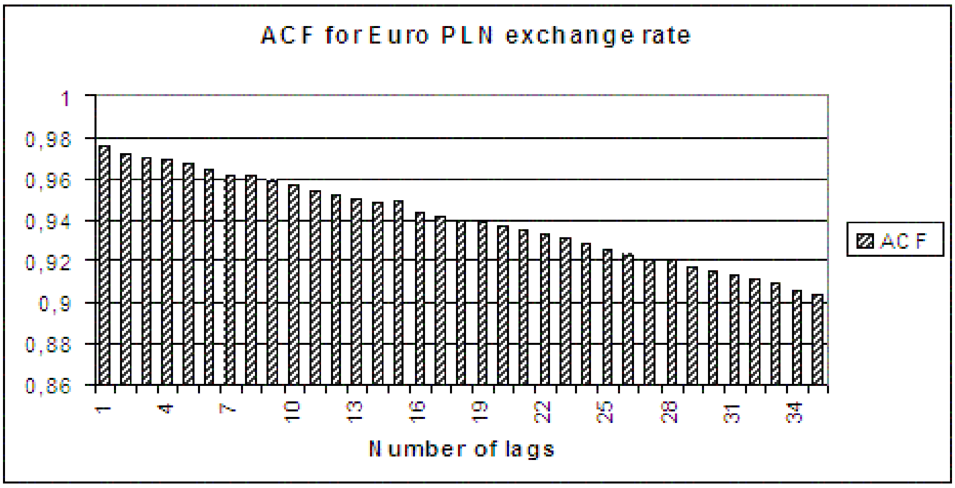
For each of the series we computed the logarithm of an exchange rate and also the logarithmic returns, i.e.

$$r_t = \ln(e_t) - \ln(e_{t-1})$$

To check influence of aggregation, for each of the daily exchange rate series, we compute weekly averages and monthly averages of logarithms, and weekly and monthly returns.

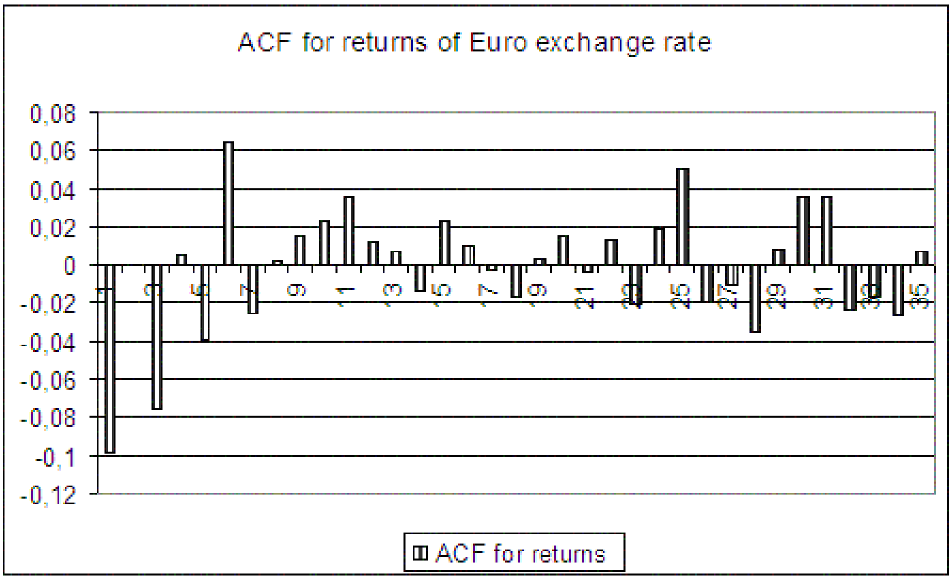
We compute empirical autocorrelation functions and partial autocorrelation functions for both logarithms and logarithmic returns of the EURO exchange rate. The results are presented in Table A2 in the Appendix. This illustrates one of characteristics of a long-memory series: the rate of decrease of autocorrelation coefficients is slower than in case of a series with short memory. Visual inspection suggests that — as can be expected — the exchange rate itself might be an integrated of order 1 or of a fractional order, whereas behaviour of the logarithmic returns seem to be quite resembling that of a stationary series.

Figure 1. The empirical autocorrelation function for an exchange rate series



Source: Author's computations based on the NBP data

Figure 2. Empirical autocorrelation function for logarithmic returns



Source: Author's computation for the NBP data

We compare here the following three methods, used for detecting long- and short term behaviour of a series, namely the modified rescaled range statistics, proposed by Lo (1991), and two versions of periodogram regression. The first was introduced by Geweke and Porter-Hudak (1983), and is still widely used; the second is a modified version of this method, proposed by Robinson (1995). As our aim is to provide a bit of advice for a practitioner, we use the following procedures, easily available as Stata programmes:

- lomodrs — the procedure for a modified Lo rescaled range statistics, written by Christopher F. Baum and Tairi Room, Boston College;
- gphudak — the procedure for the Geweke and Porter –Hudak method, by Christopher F. Baum,
- roblpr — the procedure for the modified Robinson's periodogram regression, by Christopher Baum, Boston College, and Vince Wiggins, Stata Corporation.

2.1. Rescaled range statistic and Lo modification

One of earliest tools for estimating of the fractional integration parameter is the well-known rescaled range statistic, R/S, introduced by Hurst (1951) and studied by Mandelbrot (1975) and Mandelbrot and Taqqu (1979), among others. However, Robinson (2003), p. 8—9, remarks that "while it behaves well with respect to long-tailed distributions, its limit distribution is non-standard and difficult to use in statistical inference, while it has no known optimal efficiency properties with respect to any known family of distributions, The continued popularity of R/S, for example in the finance literature, may rest in part on an inadequate appreciation of rival procedures".

Let $X_1, X_2, \ldots, X_n$ denote a time series of exchange rates or returns, and $\frac{1}{n}\sum_{i=1}^{n} X_n = \overline{X}$ – its sample mean. The classical rescaled range statistics is defined as

$$\overline{Q_n} = \frac{1}{S_n}\left[\max_{1\le k\le n}\sum_{j=1}^{k}(X_j - \overline{X}) - \min_{1\le k\le n}\sum_{j=1}^{k}(X_j - \overline{X})\right]$$

where S_n is the standard deviation estimator

$$S_n = \left[\frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2\right]^{0.5}$$

This statistic is sensitive to a short–term dependence in the series (see Lo (1991)), hence Lo (1991) introduced a modified version,

$$Q_n = \frac{1}{\hat{\sigma}_n}\left[\max_{1\le k\le n}\sum_{j=1}^{k}(X_j - \overline{X}) - \min_{1\le k\le n}\sum_{j=1}^{k}(X_j - \overline{X})\right]$$

with a consistent estimator of the variance

$$\hat{\sigma}^2(q) = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2 + 2\sum_{j=1}^{q}\omega_j(q)\hat{\gamma}_j$$

where $\omega_j(q) = 1 - \frac{j}{q+1}, q < n$, are the Bartlett weights. The truncation parameter q can be chosen by a data dependent formula, suggested by Andrews (1991): q is the greatest integer less than or equal to

$$k_n = \left(\frac{3n}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}2}\right)^{2/3}$$

where $\hat{\rho}$ is a first-order autocorrelation coefficient of the data.

Lo normalizes the computed value, dividing it by a square root of number of observations. The null hypothesis is of short memory in the series, i.e., lack of long-term memory. A computed value of the Lo modified statistics is compared with appropriate critical values, at 95% the interval is [0.809, 1.862]. We reject the null of lack of long memory, if a computed value $Q_n/\sqrt{n}$ is not in this interval.

An example give by Lo (1991) presents application of his statistics to daily and monthly stock returns of financial indices for a data from the Center for Research in Security Prices. Daily observations span was from 3 July 1962 to 31 December 1987, monthly indices (equal-weighted or value-weighted) were computed for a period from 30 January 1926 to 31 December 1987. Lo compared the classical R/S statistics with his modified version, which had lower values. The statistical insignificance of the modified statistics indicated that daily stock returns were consistent with the null hypothesis of short-memory. For first half of his sample, the classical rescaled range statistic was significant, leading to a different answer. For the monthly data, the equal-weighted and value-weighted indices were insignificant.

2.2. Fractional integration parameter

The fractional integration parameter is defined as a real number d, for which differences of a non-stationary series $\{y_t\}$ are stationary, i.e. $\Delta^d y_t = \varepsilon_t$. For d =1 we have an I(1) process. For a general d , fractional differences are defined as

$$\Delta^d = (1-L)^d = \sum_{k=0}^{\infty}\binom{d}{k}(-1)^k L^k = \sum_{k=0}^{\infty}\frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}L^k ,$$

where L denotes usual lag parameter, and coefficients are expressed with use of Gamma function.

Properties of a series depend on the value of integration parameter of the series. If d =1, the process is integrated, I(1), with infinite variance. For d >1 a process also has infinite variance, and shocks effects increase with time.

- For 0.5 $\leq d < 1$, the process has infinite variance, hence it is non-stationary, but in very long run it is mean-reverting (see Hosking (1981). Influence of external shocks lasts for a quite long time.
- For 0 < d < 0.5 the process is stationary, is mean-reverting in long run and has finite variance.
- If d = 0, the process is mean-reverting in short run, has finite variance, and shock effects diminish in short run.

2.3. The Geweke and Porter-Hudak method and Robinson's modification

One group of methods for estimating long-memory parameter is based on semiparametric periodogram regression. This is based on a relationship between spectra of processes *X* and *u*:

$$f_x(\omega) = |1 - \exp(i\lambda)|^{-2d} \ f_u(\lambda)$$

This expression is valid for a stationary X giving its spectrum, and for non-stationary series it is a limit of expectation of its periodogram (see Phillips (1999a). Taking logarithms gives

$$\ln(f_x(\lambda)) = -2d\ln(|1-\exp(i\lambda)|) + \ln(f_u(\lambda))$$

Instead of the spectrum, the periodogram ordinates are used for a set of frequencies $\lambda_s = 2\pi s / n$. Let w_a denote a discrete Fourier transform of a series and w_a^* its complex conjugate. Then the periodogram ordinates used in estimation are equal to

$$I_a(\lambda_s) = w_a(\lambda_s) w_a^*(\lambda_s)\text{, } s = 1,2,\ldots,m \tag{1}$$

where *m* is a properly chosen truncation parameter. This form of equation was used by Phillips to formulate properties of estimates in case of stationary and of integrated process. Both Phillips and Geweke and Porter-Hudak use similar form of least squares regression of $I_x(\lambda_s)$, namely

$$\ln(I_x(\lambda_s)) \cong \ln c - 2d \ln \lambda_s + u_s$$

In Geweke and Porter-Hudak (1983), a regression is of the form

$$\ln(I_x(\lambda_s)) = \ln c - 2d|1 - \exp(i\lambda_s)| + u_s$$

In their original paper, they suggest using the exact Fourier transform to compute periodogram ordinates. Phillips (1999) proposes instead a corrected version of the equation (1)

$$I_a(\lambda) = v_a(\lambda) v_a^*(\lambda)$$

with

$$v_x(\lambda_s) = w_s(\lambda_s) + [\exp(i\lambda s)X_n]/[(1-\exp(i\lambda s))\sqrt{2\pi n}]$$

The modified periodogram regression is especially useful for non-stationary series. Phillips (1999a) explains that for $d < 0.5$ the usual version of periodogram regression is also appropriate. For d=1 the correction is exact for all frequencies. For d >1, the usual periodogram is inconsistent, hence the modified version is even more useful.

Robinson (1995) introduced a modified version of the periodogram regression method, for a multivariate case, with both parameter estimates and test statistics. The periodogram regression, both its original GPH version and modified one, gives also standard errors, hence there is a possibility of testing whether $d = 1$ (in case of a series) or whether $d = 0$ (in case of returns).

3. Empirical results for exchange rates

3.1. The Lo's modified rescaled range statistics

The empirical results obtained for the exchange rate series are the following[1].

The Lo's modified rescaled range statistics for almost all daily data takes values higher than a critical value for 5% significance level. Only for Greek drachma, Czech korona, and Hungarian forint the statistic is insignificant. However, from our point of view more important is whether results depend on level of aggregation.

For most of currencies, in case of daily data and weekly and monthly averages, the Lo statistics suggests presence of long-run dependence, as it is significant. A few cases of different results are the following:

- For Canadian dollar, weekly averages, the Lo statistic is insignificant,
- For Czech korona, only for daily data,
- For Spanish peseta and British pound, only for weekly averages,
- In case of Hungarian forint, daily and weekly result is insignificant, monthly statistic is significant,
- For Japanese yen, daily result is insignificant, and significant for both weekly and monthly averages.

Note also that the proportion of monthly to daily values of the modified rescaled range statistics is quite high, with minimum 0,7, average equal to 2.3 and median 1.9, and maximum of 5.7.

The Lo statistic results for returns are much more consistent, with average and median 1.1, minimum 0.9 and maximum 1.2. Only for Finnish mark the value was significant. This suggests that the Lo statistics value does depend on the

[1] Full set of estimates is available from the author on a request.

aggregation level, in smaller extent when applied to a series with a fractional integration parameter close to 0.

The values of the Lo modified rescaled range statistics are presented in Table 3 in the Appendix. Figures 3 and 4 illustrate differences between statistics values at the three levels of aggregation.

3.2. The GPH estimates of fractional integration parameter

For the GPH estimates of the fractional integration parameter the situation is in a sense opposite. The estimates for the daily data and their weekly and monthly averages do not differ much. But the estimates for the logarithmic returns seem to be influenced by the aggregation levels (see figures 5 and 6a and b). Moreover, for the returns we obtain negative estimates. Proportion of absolute value of the fractional integration parameter estimates for *monthly* returns to an absolute value of this estimate for *daily* returns ranges from 0 to 78, for most currencies is around 10 (with average 16 and median 10), see table 1. The last column of the table shows the proportion of the absolute value of *d* for the monthly returns to its absolute value for the daily returns.

Table 1. Comparison of the GPH fractional parameter estimates for the logarithms and the logarithmic returns.

The GPH estimates	Daily data	Weekly averages	Monthly averages	Daily returns	Weekly returns	Monthly returns	Proportion
Average	1,024	1,039	0,925	0,043	0,008	0,336	16
Min	0,925	0,866	0,754	-0,114	-0,144	-0,393	0
Max	1,166	1,354	1,126	0,197	0,237	0,677	78
Median	1,016	1,027	0,927	0,026	-0,004	0,383	10

Source: Author's computations.

3.2. The Robinson's modified estimates of fractional integration parameter

In case of the Robinson's modified estimate of the fractional integration parameter we also obtain quite similar values in case of daily data, and their weekly and monthly averages, but the estimates for returns are influenced by a level of aggregation, extremely so in case of the Japanese yen, for which proportion of absolute value of estimate for monthly returns to the absolute value for daily returns is 117.

Table 2. Comparison of the Robinson's estimates of fractional integration parameter

Robinson's estimate	Daily data	Weekly averages	Monthly averages	Daily returns	Weekly returns	Monthly returns	Proportion
Average	0,92	0,89	0,83	-0,05	0,11	0,25	10 (6[1])
Min	0,90	0,85	0,78	-0,14	0,06	0,12	2
Max	0,94	0,91	0,90	0,02	0,17	0,40	117 (15[1])
Median	0,92	0,89	0,83	-0,05	0,11	0,25	5

[1]After skipping extreme values for yen.

4. Conclusion

Presence of the negative estimates for a fractional integration parameter in case of the two periodogram regression methods is important, as the positive values of *d* correspond to a process with the long memory and negative values of *d* correspond to a slightly different behaviour, namely anti-persistency. We tested the null hypothesis that the fractional integration parameter $d = 0$. In cases of the negative GPH estimates, this null hypothesis has been rejected (see Table 4). Unfortunately, in case of Robinson's negative estimates for daily returns, H_0: $d = 0$ in most cases could not be rejected[1].

Table 3. Negative estimates for GPH daily returns are rejected

Symbol	GPH estimate	t-ratio for H0: d=0
Dem	-0.0059	-0.0582
Dkk	-0.0424	-0.4250
Esp	-0.114	-1.0818
Euroa	-0.042	-0.4451
Frf	-0.0158	-0.1396
Luf	-0.0129	-0.1185
Nlg	-0.0128	-0.1254
Pte	-0.0387	-0.3452

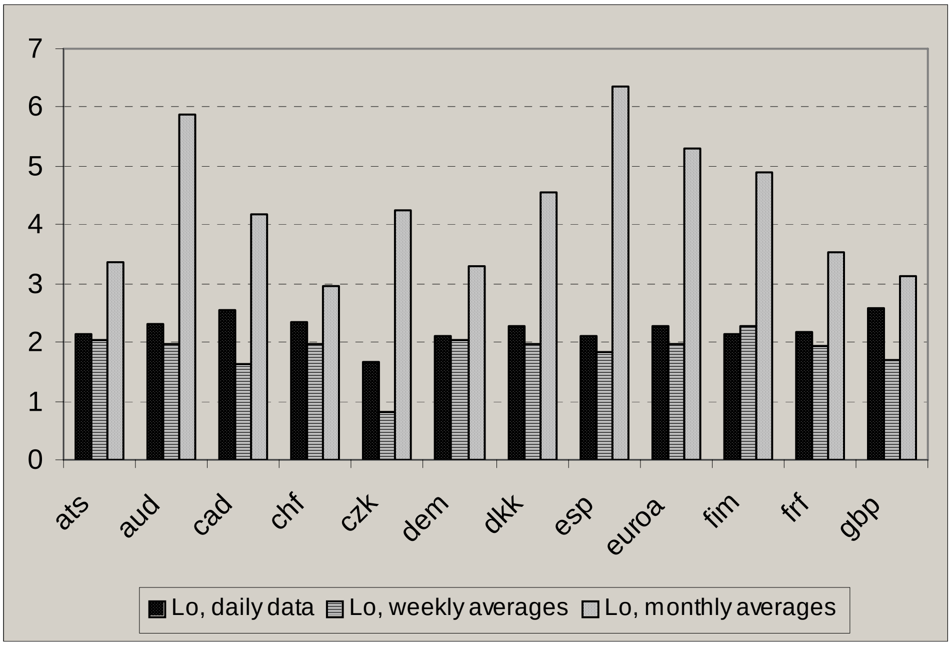
Source: Author's own computation

The results presented here for the daily, weekly and monthly data for the set of 26 exchange rate versus Polish zloty show that a question of appropriate choice of long-memory behaviour tests and checks merits careful attention, and not only its theoretical properties (which have been thoroughly analysed by Mandelbrot, Phillips, Geweke, Robinson and others), but from the point of view of a

[1] Table with all the Robinson's method results available on request.

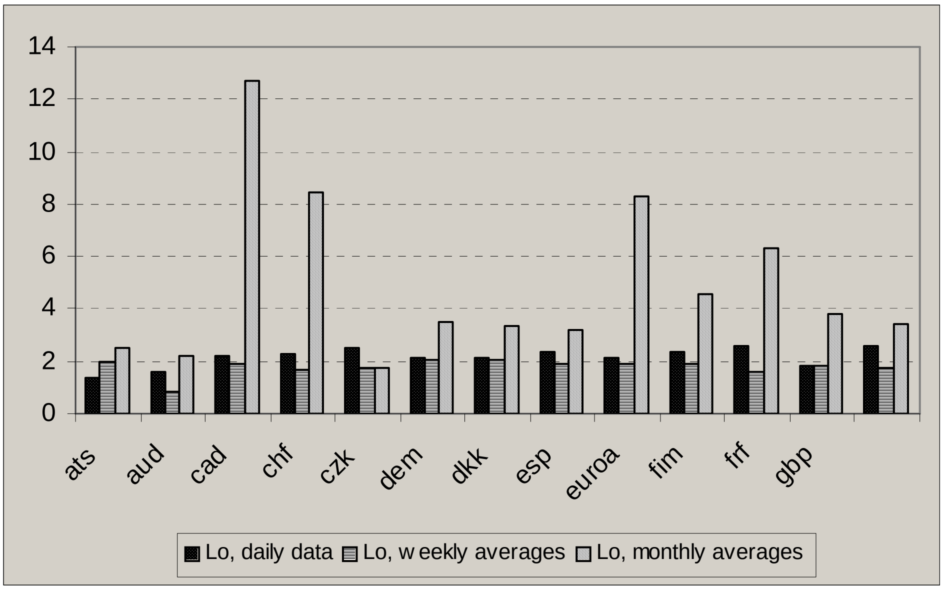
practitioner. Undergoing research, not presented here, concerns comparison of results for estimation of fractional integration parameter with wavelet methods (see, for example, Daubechies (1991)), which are reported to be more accurate theoretically.

Figure 3. The Lo modified rescaled range statistics for daily data and weekly and monthly averages, part 1



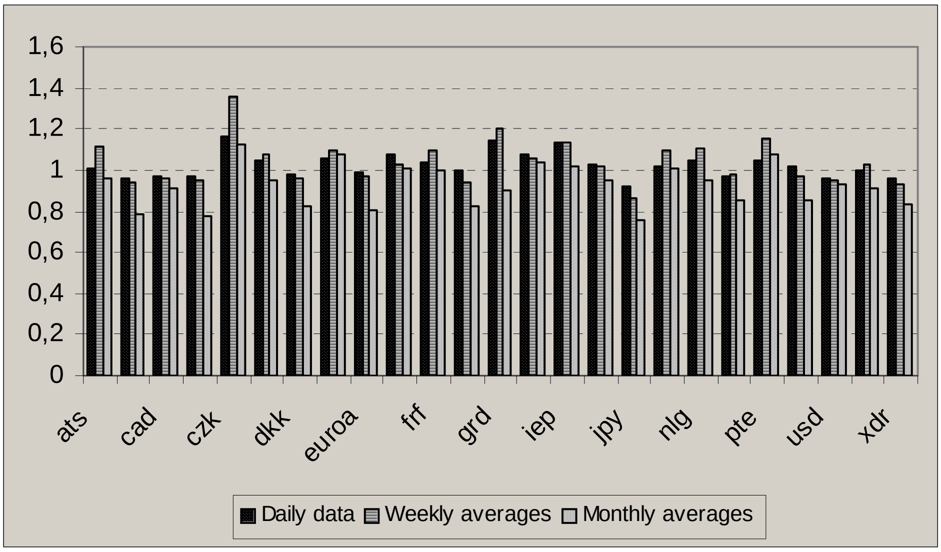
Source: Author’s computations

Figure 4. The Lo modified rescaled range statistics for daily data and weekly and monthly averages, part 2



Source: Author's computations

Figure 5. The GPH estimates of a fractional integration parameter



Source: Author's computations.

Figure 6. The GPH estimates of a fractional integration parameter for returns, part 1.

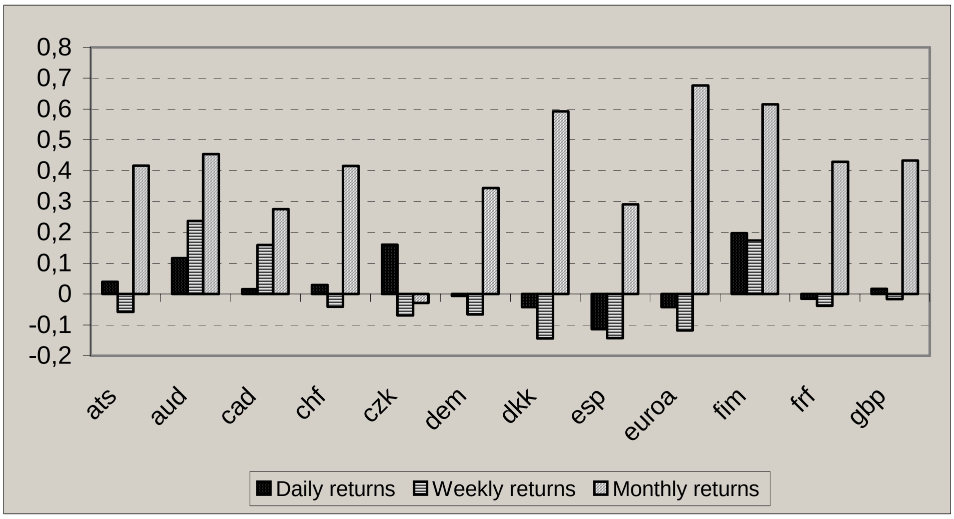
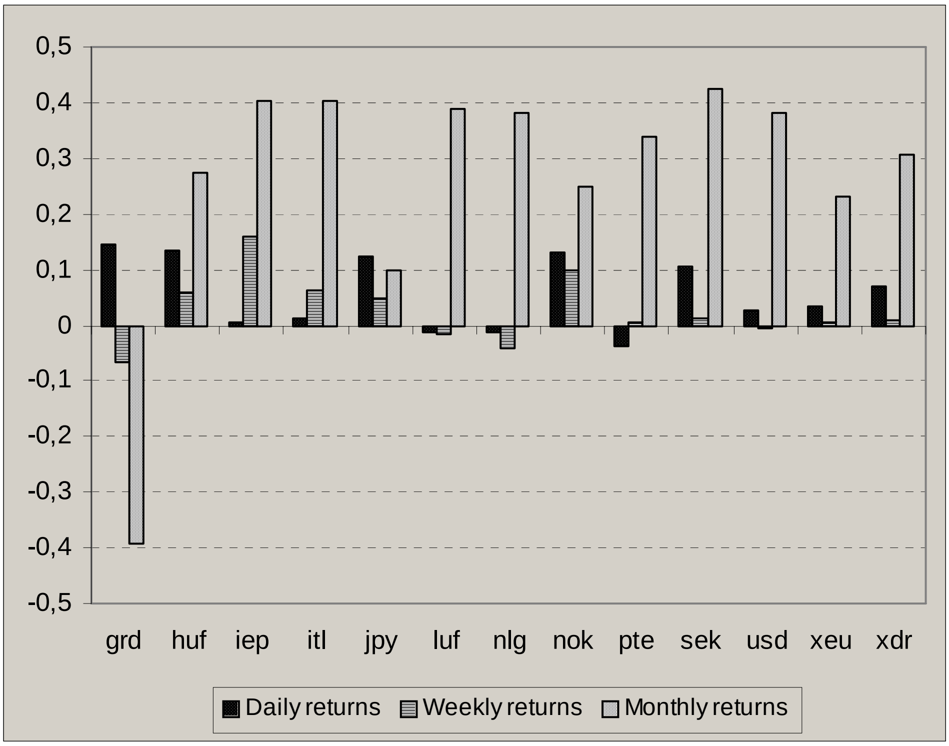


Figure 7. The GPH estimates of a fractional integration parameter for returns, part 2.



REFERENCES

ANDREWS, D. 1991: Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817—858.

DAUBECHIES, I. 1998: Orthonormal Bases of Compactly Supported Wavelets, *Communications in Pure and Applied Mathematics*, 41, 909–996.

GEWEKE, J. , PORTER-HUDAK, S. 1983: The estimation and application of long-memory time series models, *Journal of Time Series Analysis*, 4, 221—228; reprinted in: Robinson (2003a).

HOSKING, J.R.M. 1981: Fractional differencing, *Biometrika*, 68(1), 165—176.

HURST, H. 1951: Long term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, 116, 770—799.

LO, A. H. 1991: Long-term memory in stock market prices, *Econometrica*, 59, 1279—1313, reprinted as chapter 5 in: Robinson (2003).

MANDELBROT, B. 1972: Statistical methodology of non-periodic cycles: from the covariance to the R/S analysis, *Annals of Economic and Social Measurement*, 1, 259—290.

MANDELBROT, B. 1975: Limit theorems on the self-normalized range for weakly and strongly dependent processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31, 271—285.

ORĘZIAK, L. 2002: *Euro – new currency*, (in Polish), The PWN Scientific Publishers, Warsaw.

ROBINSON, P. M. 1995: Log-periodogram regression of time series with long range dependence, *Annals of Statistics*, 23, 1048—1072.

ROBINSON, P. M. 2003: Long-memory time series, chapter 1 in: Robinson (2003a), 4—32.

ROBINSON, P. M. (editor) 2003a: *Time series with long memory*, Oxford University Press, Oxford.

SYCZEWSKA, E. M. 2002: Nonstationarity analysis for USD/PLN exchange rate based on daily and monthly data (in Polish), in: *Roczniki Kolegium Analiz Ekonomicznych* “Decision analysis in eonomic applications”, ed. by T. Szapiro, 10/2002, 159—175.

SYCZEWSKA, E. M. 2004: Comparison of exchange rate behaviour for Hungary, Latvia and Poland, *Macromodels' 2003*, Łódź 2004, 39—61.

Appendix

Table A1. The Euro conversion rates

1 euro =	Currency	
13.7603	Austrian shilling	ATS
40.3399	Belgian franc	BEF
2.20371	Dutch gulden	NLG
5.94573	Finnish mark	FIN
6.55957	French frank	FRF
1.95583	Deutschmark	DEM
0.787564	Irish pound	IEP
1936.27	Italian lira	ITL
40.3399	Luxemburg franc	LUF
200.482	Portuguese escudo	PTE
166.386	Spanish peseta	ESP

Source: Table 3, page 78 in: Oręziak, 2002.

Table A2. Empirical autocorrelation and partial autocorrelation functions for the Euro exchange rate

Lags	Logarithm of euro			Logarithmic returns		
	AC	PAC	Q-Stat	AC	PAC	Q-Stat
1	0.976	0.976	2747.0	-0.098	-0.098	26.780
2	0.972	0.402	5471.1	0.000	-0.010	26.781
3	0.970	0.251	8183.9	-0.075	-0.077	42.612
4	0.969	0.192	10892.	0.005	-0.010	42.680
5	0.967	0.120	13590.	-0.039	-0.042	47.031
6	0.964	0.048	16273.	0.064	0.051	58.456
7	0.962	0.047	18945.	-0.025	-0.016	60.246
8	0.962	0.092	21620.	0.002	-0.006	60.259
9	0.959	-0.017	24274.	0.015	0.022	60.867
10	0.956	-0.011	26914.	0.023	0.023	62.299
11	0.954	0.006	29545.	0.035	0.045	65.810
12	0.952	0.002	32166.	0.012	0.019	66.222
13	0.950	0.000	34776.	0.007	0.018	66.375
14	0.948	0.010	37376.	-0.014	-0.004	66.929
15	0.949	0.097	39983.	0.023	0.024	68.432
16	0.944	-0.076	42562.	0.010	0.018	68.713
17	0.942	-0.011	45132.	-0.003	-0.003	68.733
18	0.940	-0.002	47691.	-0.017	-0.014	69.515

Lags	Logarithm of euro			Logarithmic returns		
	AC	PAC	Q-Stat	AC	PAC	Q-Stat
19	0.938	-0.005	50241.	0.003	-0.001	69.537
20	0.937	0.030	52787.	0.015	0.015	70.148
21	0.935	0.024	55324.	-0.004	-0.009	70.201
22	0.933	-0.002	57850.	0.013	0.009	70.696
23	0.931	-0.020	60366.	-0.020	-0.018	71.803
24	0.928	-0.027	62866.	0.019	0.015	72.806
25	0.926	-0.016	65355.	0.050	0.054	79.776
26	0.923	-0.011	67833.	-0.019	-0.015	80.758
27	0.921	-0.002	70301.	-0.011	-0.009	81.094
28	0.921	0.065	72770.	-0.035	-0.034	84.545
29	0.917	-0.052	75218.	0.008	0.004	84.732
30	0.915	-0.029	77654.	0.035	0.035	88.165
31	0.913	0.014	80081.	0.035	0.030	91.680
32	0.911	0.012	82500.	-0.024	-0.013	93.249
33	0.909	-0.004	84907.	-0.017	-0.018	94.083
34	0.906	-0.016	87300.	-0.026	-0.022	95.955
35	0.904	0.003	89685.	0.007	-0.004	96.090
36	0.903	-0.005	92061.	-0.010	-0.018	96.366

Source: Author's own computations.

Table A3. The Lo's modified rescaled range statistics*

Symbol	Daily data	Weekly averages	Monthly averages	Daily returns	Weekly returns	Monthly returns
Ats	2.13	2.05	3.37	1.42	1.66	1.52
Aud	2.32	1.96	5.87	1.44	1.67	1.43
Cad	2.54	**1.62**	4.18	1.26	1.35	1.32
Chf	2.33	1.98	2.94	1.34	1.46	1.49
Czk	**1.68**	**0.83**	4.24	1.17	1.37	1.20
Dem	2.12	2.04	3.29	1.41	1.54	1.50
Dkk	2.28	1.97	4.56	1.39	1.59	1.58
Esp	2.12	**1.82**	6.36	1.39	1.44	1.42
Euroa	2.29	1.96	5.29	1.36	1.50	1.51
Fim	2.13	2.27	4.88	1.69	1.87	1.46
Frf	2.16	1.94	3.54	1.43	1.61	1.50
Gbp	2.59	**1.71**	3.13	1.56	1.59	1.58
Grd	**1.36**	1.99	2.53	1.10	1.47	1.21
Huf	**1.63**	**0.835**	2.24	1.28	1.65	1.38
Iep	2.24	1.87	12.7	1.59	1.74	1.57
Itl	2.32	**1.65**	8.41	1.30	1.43	1.41
Jpy	2.5	**1.73**	1.78	1.38	1.37	1.20
Luf	2.11	2.08	3.49	1.41	1.67	1.54
Nlg	2.12	2.07	3.31	1.49	1.63	1.50
Nok	2.36	1.91	3.18	1.35	1.60	1.47
Pte	2.14	1.92	8.33	1.41	1.53	1.45

Symbol	Daily data	Weekly averages	Monthly averages	Daily returns	Weekly returns	Monthly returns
Sek	2.33	1.91	4.6	1.42	1.69	1.55
Usd	2.58	**1.61**	6.28	1.59	1.60	1.60
Xeu	**1.86**	**1.81**	3.77	1.24	1.33	1.44
Xdr	2.55	**1.75**	3.44	1.14	1.64	1.41

Source: Author's computation

*Insignificant values marked in bold.

Table A4. The Geweke and Porter-Hudak estimates of fractional integration parameters for daily data and weekly and monthly averages

Symbol	GPH daily	GPH weekly averages	GPH monthly averages	GPH daily returns	GPH weekly returns	GPH monthly returns
Ats	1.009	1.115	0.963	0.039	-0.058	0.416
Aud	0.960	0.937	0.784	0.116	0.237	0.454
Cad	0.974	0.964	0.908	0.016	0.159	0.275
Chf	0.972	0.948	0.773	0.029	-0.041	0.415
Czk	1.166	1.354	1.126	0.159	-0.069	-0.029
Dem	1.044	1.078	0.954	-0.006	-0.067	0.344
Dkk	0.979	0.958	0.822	-0.042	-0.144	0.593
Esp	1.055	1.094	1.080	-0.114	-0.143	0.291
Euroa	0.993	0.969	0.801	-0.042	-0.118	0.677
Fim	1.078	1.027	1.007	0.197	0.174	0.615
Frf	1.034	1.099	0.996	-0.016	-0.038	0.429
Gbp	1.000	0.938	0.826	0.017	-0.016	0.433
Grd	1.140	1.199	0.897	0.147	-0.067	-0.393
Huf	1.078	1.058	1.036	0.135	0.059	0.275
Iep	1.132	1.138	1.015	0.005	0.158	0.403
Itl	1.027	1.021	0.952	0.014	0.064	0.403
Jpy	0.925	0.866	0.754	0.123	0.048	0.098
Luf	1.016	1.095	1.012	-0.013	-0.016	0.390
Nlg	1.050	1.109	0.948	-0.013	-0.043	0.383
Nok	0.973	0.977	0.858	0.132	0.097	0.249
Pte	1.048	1.155	1.079	-0.039	0.006	0.337
Sek	1.021	0.970	0.854	0.105	0.014	0.423
Usd	0.964	0.948	0.927	0.026	-0.004	0.383
Xeu	1.001	1.031	0.911	0.033	0.006	0.232
Xdr	0.961	0.934	0.838	0.069	0.010	0.308

Source: Author's computations.

Table A5. The Robinson's modified periodogram regression estimator of fractional integration parameter

Symbol	Daily data	Weekly averages	Monthly averages	Daily returns	Weekly returns	Monthly returns
Ats	0.919	0.874	0.810	-0.088	0.068	0.229
Aud	0.928	0.905	0.864	-0.032	0.118	0.286
Cad	0.925	0.903	0.843	-0.022	0.116	0.172
Chf	0.926	0.889	0.818	-0.040	0.138	0.243
Dem	0.924	0.893	0.855	-0.052	0.100	0.285
Dkk	0.925	0.884	0.818	-0.061	0.130	0.213
Esp	0.913	0.890	0.845	-0.047	0.109	0.124
Euroa	0.928	0.889	0.834	-0.047	0.115	0.244
Fim	0.942	0.905	0.878	-0.057	0.092	0.338
Frf	0.921	0.873	0.808	-0.061	0.112	0.179
Gbp	0.920	0.887	0.862	-0.048	0.171	0.198
Iep	0.925	0.867	0.904	-0.021	0.119	0.300
Itl	0.934	0.878	0.819	-0.063	0.112	0.271
Jpy	0.913	0.885	0.829	-0.002	0.143	0.274
Luf	0.922	0.883	0.823	-0.077	0.061	0.192
Nlg	0.926	0.879	0.835	-0.054	0.090	0.281
Nok	0.926	0.885	0.809	-0.066	0.094	0.281
Pte	0.904	0.871	0.818	-0.073	0.111	0.196
Sek	0.937	0.910	0.847	-0.053	0.100	0.404
Usd	0.921	0.904	0.826	0.021	0.142	0.319
Xeu	0.912	0.853	0.783	-0.064	0.169	0.173
Xdr	0.925	0.896	0.824	-0.136	0.108	0.261

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp.475—500

BUSINESS SURVEYS AND OFFICIAL STATISTICS IN RUSSIA. WHICH ONE IS BETTER?

Serguey Tsukhlo[1]

ABSTRACT

A great number of similar regular surveys on enterprises certainly fuels competition between Russian organisations that carry them out. And our respondents' replies testify to a strong competition between survey organisers for the "voices of enterprises". This necessitates monitoring of their attitudes towards surveys, particularly in the event their organiser is an NGO that does not have an administrative resource. Results of this monitoring demonstrated a high efficiency of the classical survey procedure that IET was implementing. Not less than a half of the enterprises took part in the surveys, because they received a useful information in exchange, and they could clearly admit that. Establishing informal relations with respondents, providing their voluntary participation in surveys opens opportunities for receiving assessments of reliability of the official statistics. The national enterprises' official reporting is partially unreliable. The unreliability of reporting, as a rule, does not depend on a particular sector, size or property form of an enterprise. Survey results have proved to be quite competitive vis-à-vis the official statistical data. Surveys can received high scores from respondents, providing their organizer is aware of their needs and takes those into account while preparing the results to be sent back to the enterprises. Questions are, what the organizer needs and what enterprises are interested to know, while the returned data should be primarily those that the enterprises are interested in.

***Key words*: surveys about surveys, Eastern Europe, correct economic data, official statistics.

[1] Head, Business Surveys Department, Institute for the Economy in Transition, Moscow, Russia;
E-mail: tsukhlo@iet.ru

Introduction

The history of business surveys (BS) in Russia already allows to draw first conclusions and identify new opportunities this particular source of economic information provides to transition economies.

First, regular surveys on enterprise heads are of interest to numerous organisations. In Russia, already seven of them have launched their own surveys on the federal tier, i.e. their surveys cover more than one or several Russian regions. Those are: the Institute for the Economy in Transition, the Institute for National Economic Forecasting, the Service for Special Communication and Information under the Federal Secret Service, the Centre for Economic Juncture under the RF Government in co-operation with the Federal Government Statistics Service, the Central Bank of Russia, and NTC Research (UK). The surveys by the noted structures have much in common and at the same time show fundamental peculiarities. What they have in common is their regular conduct, repetitiously asked major questions, similarity in their formulation, and attempts to ensure the feedback from enterprise heads. At the same time, we believe, their peculiarities lie with differences in material and financial resources, existence/absence of the so-called "administrative resource", the extent to which the industrial sector is covered, respondents' motivation and their position in a company. As a result, the competition for "voices" of participant enterprises has become more intense, and the first part of the paper will be dealing with this specific challenge.

Second, providing enterprises take part in a survey voluntarily and acknowledge the fact, the survey organiser enjoys an opportunity to establish and maintain an informal contact with them. This in turn opens fundamentally new horizons for studying into a number of sensitive problems and, particularly, the assessment of trustworthiness of the national official statistics.

Third, the specifics of regular enterprise surveys has already earned the general recognition. Once arranged in their classical form, BS appear a more operative source of economic information vis-à-vis the government statistical data. They allow tracking down a number of fundamentally new indicators that appear inaccessible to the official statistics. So a direct comparison of the official statistics and BS outcomes has formed the third problem the present paper addresses.

As an empirical base, the paper employs results of the "surveys on surveys" the IET holds in the course of regular BS of heads of industrial enterprises.

Why Enterprises Take Part in the IET Surveys

Originally, IET started monitoring of reasons for industrial enterprise management’s participation in business surveys. The necessity of such monitoring is dictated by all the enterprises' voluntary participation in monthly business surveys that IET has conducted according to the harmonized methodology since

1992. Whereas IET has no administrative mechanisms to exercise influence on enterprises, their contribution to the surveys depends solely on the respondents' interest in results that are sent back to the enterprises along with the next questionnaire. Such a motivation is also suggested by the classical survey procedure developed by the father of business surveys, IFO (Germany). The workability of the procedure is indirectly proved by the existence in various countries of business surveys national NGOs hold. Notably, according to the contemporary Russian terminology, they have no administrative resource. At the same time, in the West, various surveys have long become quite widespread. That is why it cannot be excluded that Western managers respond to business surveys following the "plebiscite" custom that has long took its roots in the Western countries: that is, one is keen to express their views for the benefit of the society as a whole rather than their own. The similar motivations may also underlie the Russian directors' behavior. However, while in the West such motivations are steered by the long-established traditions, in Russia, the eagerness to express one's personal view may be fueled by a long absence of democratic institutions and the keenness to fill in the vacuum. It was back in 1996 when the IET survey first comprised the question on the reasons for the respondents' participation. As the first try has proved to be a success, the question was asked annually. At the same time, the formulation of questions and answers remained unchanged, which allowed a maximum comparatibility of results (perhaps, thus lowering the quality). The respondents' answers have proved to be a good feedback channel for the survey organizers, as they showed the evolution of the enterprises' attitude towards the results disseminated among enterprises and formed the sole argument of their participation in the IET surveys.

Results of the monitoring demonstrated a high efficiency of the classical survey procedure that IET was implementing. Not less than a half of the enterprises took part in the surveys, because they received a useful information in exchange, and they could clearly admit that. Most likely, the proportion of such enterprises may be even greater, as a part of the respondents did not cite this particular reason not to over-praise the organizers. The reverse situation (i.e. enterprises exaggerating the usefulness of the surveys) appears highly unlikely. The proportion of such responses grew up to 63% by 2003, with a consequent drop to 57%, while the analogous dynamics was found for another "egocentric" variant of response: that is: "This is the pretext to think over the situation at the enterprise". Between 1996 to 2003 the proportion of this kind of responses grew from 28 to 37% and fell consequently to 33%. The addition of the proportions of enterprises that opted for the first and third variants of the response less the double count, i.e. the enterprises that gave both responses, provides yet more impressive results: the proportion of enterprises that recognize, in one form or another, the usefulness of the surveys grew from 65% in 1996 up to 75% in 2003, with a consequent fall to 69%. Hence two-thirds of the respondents as a minimum participate in the IET surveys, because they are keen to receive their results.

Table 1. Why Do You Response on Our BS Questionnaires? (% to the Number of Answers)

	1996	1997	1998	1999	2000	2001	2002	2003	2004
1. so as to change the response for useful information	50	53	54	58	61	60	61	63	57
2. awareness of “social” use of such interviews	45	42	37	38	35	38	36	35	36
3. it’s a good pretext to think over the performance of own enterprise	28	31	33	34	35	35	36	37	33
4. it’s a mere habit to react to any inquiry sent to my enterprises	10	11	10	12	11	11	11	10	11
5. my managers charged me responsible for this	5	8	8	9	11	13	16	15	16
6. because of inquisitiveness	6	6	8	6	6	5	6	7	5
7. hard to say	2	2	3	2	2	2	2	3	4
8. other	2	2	2	1	1	1	2	1	1

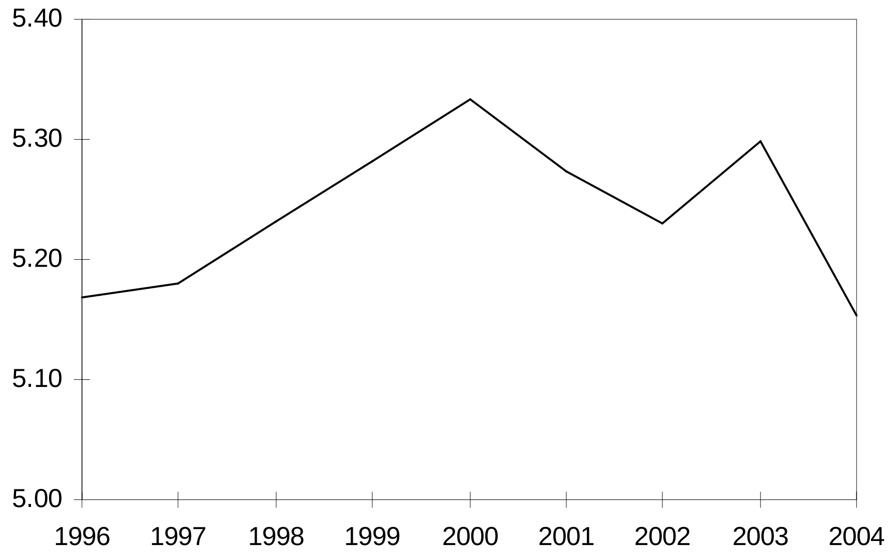
Sources: Surveys by the Institute for the Economy in Transition (Russia).

The reasons for the enterprises' participation in the surveys can be ranked according to the extent the respondents take personal interest in the results they receive, as follows:

1. hard to say
2. it’s a mere habit to react to any inquiry sent to my enterprises
3. my managers charged me responsible for this
4. awareness of “social” use of such interviews
5. because of inquisitiveness
6. it’s a good pretext to think over the performance of own enterprise
7. so as to change the response for useful information

The higher is the rank of the reasons for participation in the list, the more qualitatively business surveys are organized and the receipt of respondents' reaction is guaranteed. The ranking allows building the simplest (which appears adequate to the general ideology of the surveys) aggregate index of the enterprises' interest in the surveys that takes into account all the reasons for their participation. The value of the index may fluctuate between 7 (keen to receive a useful information, all the enterprises take part in the surveys) to 1 (all the participants have failed to identify their attitude towards the surveys). The dynamics of the aggregate index of interest in the surveys showed that between 1996 through 2000 the IET had been consistently successful in ensuring a stable improvement of the structure of the interest in the surveys. The structure then began to deteriorate and, following the rise in its improvement in 2003, it fell to its absolute minimum in 2004. However, during the whole period of observations the value of the index was fluctuating within rather a narrow range: between 5.15 to 5.30. This undoubtedly testifies to the prevalence of "positive" reasons for participation in the surveys. The "negative" reasons (i.e. " Because of the custom to react to any incoming inquiry" and "I was assigned by a senior" were cited by 21% of enterprises on average, while "positive" ("Receiving a useful information in exchange" and " The pretext to think over the situation at the enterprise"- by 70%.

Figure 1. Aggregate Index of the Enterprises' Interest in the Surveys



Numerous reasons may be cited to explain the change in the interest in the IET survey results and, accordingly, in the surveys themselves.

First, while at the outset of reforms such surveys had appeared a new and exotic phenomenon, they consequently have lost the aura, and their recipients have grown accustomed to them. In all likelihood, at first enterprises had a fairly vague understanding of pluses and minuses of independent regular surveys. At the time, the major incentive was the comprehension of the fact that some and evidently respected Moscow-based organization asked directors (not demanding, as it would happen before) to provide an assessment of their enterprises' current situation and to inform of their plans and forecasts. That might be especially true as long as non-Moscow and smaller enterprises were concerned. Unfortunately, the reasons for the participation were not included in the surveys during the first four years of their existence (1992—1995).

Second, the rise in the number of surveys also decreased the enterprises' interest and is likely to irritate them because of another two reasons: an intensifying flow of mandatory and non-compulsory inquiries and reporting forms objectively lays a greater burden on enterprises, and hence a natural negative reaction. As the number and size of such inquiries and forms is most likely to grow, one should not expect the enterprises' better attitude towards them.

Plus, a great similarity of questions comprised therein also contributes to the above. Of the noted seven organizations four have employed the European harmonized BS methodology and send out questionnaires that appear more different in design and the number of questions than in the contents. This causes confusion, and enterprises begin to return questionnaires to a wrong sender or to send different questionnaires to the same sender. To assess the level of the competition in the market for "enterprises' responses", in April 2004 IET complemented its regular survey with the following question, "Do you respond personally to analogous monthly questionnaires by other organizations (for instance, the Center of Economic Juncture under the RF Government, Goskomstat of RF, the Central Bank of RF, the Russian Economic Barometer, NTC Research?" The question implied two variants of the answer: 1) "Yes, regularly"; 2)"No".

We believe that our respondents' replies testify to a strong competition between survey organizers for the "voices of enterprises" (See Table 2). Almost 60% of the IET respondents also reply to other organizations' questionnaires. This means they have a chance to compare both the questionnaires (primarily from the perspective how comfortable they are to complete them and whether they are overloaded with questions or not) and results of the earlier conducted surveys. It should be emphasized that, having no administrative mechanisms to influence the correspondents, for the sake of sustaining competition, IET can use only an optimization of the size of its questionnaire and the quality of results to be returned to the respondents. Notably, even constrained as it is, IET so far has performed well.

More specifically, the comparison of the involvement in other surveys of respondents holding different positions in a company shows that enterprise directors more often than other managers prefer only the IET surveys, notwithstanding the fact the director is a clear minority at an enterprise (he is always one), and he has the power to re-assign the task of completing a questionnaire to any junior. Needless to say, the questionnaire the enterprise has received for the first time may well not be brought to his attention. It is deputy directors and chief specialists that carry the burden of their completion. Most likely, they try to concentrate in their hands information flows to prepare economic decisions, that is why questionnaires and survey results can be found on their desks, particularly because the latter (providing they are disseminated among enterprises) provide a compact and concentrated wrap-up of results of the previous year. Heads of divisions (primarily economic ones) also favor the IET surveys: 44% of them do not respond to other questionnaires. At the same time, the rise in the level of participation of their deputies and other staff in other surveys is most likely to be associated with the respective assignment rather than their personal interest.

Table 2. Do You Respond Personally to Analogous Monthly Questionnaires by Other Organisations, 2004. (% to the Number of Answers)

	Yes	No
All respondents	59	41
Respondents' positions		
1. Directors	54	46
2. Vice-directors	63	37
3. Department heads	56	44
4. Deputy department heads	76	24
5. Other stuffs	62	38
Size of firms, employees		
1-200	48	52
201 -1000	58	42
1001 -5000	62	38
>5000	72	28

Sources: Survey by the Institute for the Economy in Transition (Russia).

The dependence of the level of the enterprises' participation in/ignorance of other questionnaires on their size is determined, primarily, by the number of enterprises. As the largest companies, whose number is limited, are of interest to all survey organizers, this particular group demonstrates the most intense contribution to various surveys, while the smaller is the enterprise, the greater number of them can be found and, accordingly, both the survey organizers' interest in them and the probability of falling different organizations' questionnaires on the given enterprise tends to fall.

The data on the respondents' participation in/ignorance of other surveys allows to seek an answer to the question as to whether the attitude towards the IET surveys changes, providing the respondent can compare different questionnaires and, most importantly, the respective results. Computations show that the attitude towards the IET survey has not undergone any fundamental changes (see Table 3). Perhaps, the contents of survey results that different organizations disseminate among enterprises does not differ dramatically, and the enterprises find it sufficient to receive (or read) results of a sole survey. However, not all the survey organizers send out survey results in the form the enterprises are comfortable with. This primarily is true, as long as the government agencies are concerned, as they are in possession of administrative mechanisms, thus not caring much of the enterprises' interests. By contrast, NGOs of course send their results to enterprises.

The rise in the proportion of answers "My managers charged me responsible for this ", providing the given respondent participates in more than one survey, would mean that the growing "survey burden" on an enterprise has resulted in a lower rank of the actual respondent and, accordingly, the fall in the quality of the feedback. This also means that a certain staff has been assigned to deal with completing questionnaires and, perhaps, he/she gradually acquires the respective experience.

Table 3. Why do You Response on our BS Questionnaires? (2004, % to the Number of Answers)

	Do you take part in other surveys?	
	Yes	No
1. so as to change the response for useful information	57	61
2. awareness of "social" use of such interviews	38	39
3. it's a good pretext to think over the performance of own enterprise	37	35
4. it's a mere habit to react to any inquiry sent to my enterprises	12	11
5. my managers charged me responsible for this	18	11
6. because of inquisitiveness	5	5
7. hard to say	3	4
8. other	1	0

Sources: Surveys by the Institute for the Economy in Transition (Russia).

The third reason underlying the decline in the interest in surveys may become the rise in the enterprises' informational awareness. The drastic transition from the planned economy to market in the 1990s put Russian companies in a very complex situation, particularly from the perspective of informational backup of the decision-making processes. However, in the course of time and especially in the wake of the 1998 default and the consequent start of a normal rise in output and improvement of their financial and economic shape, enterprises are most likely to address informational challenges and establish much-needed analytical and forecast divisions. Given the above, the value of BS for enterprises should fall, and at worst it may lead to the refusal to take part in voluntary NGOs' surveys, while at best — to the lower status of the respondent. However, results of yet another IET monitoring have failed to show a rise in the national enterprises informational awareness (see Table 4).

The lack of the sectoral information was always on the top of the list, followed by the lack of information on new consumers (10 percentage points down) and the need to have information on competitors (another 10 p.p. down). The need in the information in the form of general economic reviews (and most likely, survey results fall under this particular category) was holding the 9th position on average out of 10 categories of information. Almost minimum lack of such information can be explained both by little need in it and its high availability. At this point, perhaps, a regular dissemination of survey results has also played its positive role. These data also suggest the possibility to boost the enterprises' interest in surveys: that is, to shift the focus of the analysis onto the sectoral level.

Table 4. The Shortage of Which Economic Information Do You Currently Perceive? (% to the Number of Answers)

	2000	2001	2002	2003
1. General economic/reviews	20	13	16	13
2. Sectoral	68	71	70	67
3. Regional	22	22	23	19
4. Financial	25	22	19	20
5. Legal	27	25	22	21
6. On stock market	9	7	6	6
7. On new (potential) suppliers	41	36	35	35
8. On New (Potential) consumers	58	62	60	59
9. On competitors	46	48	55	52
10. On investors	39	35	33	32
11. Which else	3	3	3	2

Sources: Surveys by the Institute for the Economy in Transition (Russia).

Is Russian Official Statistics Trustworthy?

The assumption that Russian enterprises deliberately distort their financial reports for the sake of minimizing tax withdrawals has first found its public proof in words (speeches and interview) by some directors. They argued quite openly that enterprises conducted three kinds of accounting: for the state, for themselves, and for an outsider investor. Such state of affairs both derails the confidence in macrostatistics and lowers Russian companies' investment attractiveness, albeit they are very much in need for outsider investment. However, it is either impossible, or extremely complex (costly) for public agencies to collect large-scale statistical assessment of this particular phenomenon. Perhaps, the only rational method of studying the problem is to add questions on trustworthiness of enterprises' accounting and reporting to the regular and voluntary BS on enterprise heads that an NGO conducts for a long period of time. It is only by such a means that can one hope for collecting data which would be maximally close to the actual situation. There are several reasons for that.

First, the regularity and the panel organization of BS teaches enterprises to react to questionnaires and builds trust between the survey organizer and his respondents. Second, the panel organization of BS allows to establish and maintain personified relations with the respondents. Third, the confident relations with the respondents allow an easy conduct of additional surveys that would ensure a high guarantee of questionnaire return. Fourthly, the non-government status of the survey organizer allows to hope for a greater frankness of responses to sensitive questions, including the problem of trustworthiness of the enterprises' official statistical reporting. Naturally, it would be naive to hope for a full frankness of such responses, however, the noted comments on the nature of BS allow the assumption that, collected in such a fashion, assessments of trustworthiness of enterprises' official reporting would become most close to the actual state of affairs. In addition, while being challenged by the above, one also faces another problem: that is, how to get competent assessments of the trustworthiness of enterprises' financial reporting. The competence of the received assessments depends on the respondent's position within the company. It would be logical to assume that the higher his position is, the more complete picture of to what extent the official reports of the enterprise he manages matches its actual position, for it often happens that it is only the director and the chief accountant of an enterprise that are aware of very delicate "adjustments".

In 2000, for the first time ever the IET BS contained the question about the trustworthiness of reporting by Russian industrial enterprises. The question was put as follows, "Do you think one can use the enterprises' official (i.e. reported to the public agencies) reporting while analyzing an actual state of affairs in the national industrial sector?" The enterprises were offered two variants of the response: 1) Yes, we do, and 2) No, it would be better no to do it.

Such a formulation of the question, on the one hand, appears quite informative, while not causing respondents' concern and their avoiding answering to it. Evidently, once a direct (straightforward) form of the question is used, it cannot ensure satisfactory results. The formulation we used also excludes from consideration the problem of lack of completeness of the official information. The respondent immediately guesses that the question implies solely the appropriateness of the use of enterprises' reports, while not meaning that the official reporting appears sufficient for evaluation of the state of affairs in the national industrial sector.

Over 1,000 enterprises gave their assessments of the trustworthiness of the official reporting in the national industrial sector, with only 27 of them (2.6% of those who sent back the questionnaires) not daring to reply. The results of the survey showed that the problem of trustworthiness of the official reporting indeed was in existence in the national industrial sector, and enterprises recognized it, and they were ready to share their assessments with the organizers.

Over one-fourth of the surveyed respondents could (dared to) recognize that it would be better not to use the official reporting by Russian industrial companies for the analysis of the actual state of affairs in the Russian industrial sector (see Table 5). And those are minimum assessments, for some part of respondents apparently could not give a frank reply or they evaded it, or responded "yes". But even such figures appear impressive (in negative sense, of course). It appeared that the most trustworthy reporting could be found in the sector for electricity: only 21% of enterprises in this particular sector believed that their reporting could not be trusted. The fuel sector formed the opposite pole, with the number of such enterprises doubled — 45%. But this value makes the sector unique vs. the others, where the respective index "at best" hits 31%. Such a confession by the fuel companies is most likely to be associated with the considerable export capacity of the sector and consequently with their excellent financial performance. Hence, they have something to conceal from the state. Holding the second position, the non-ferrous metallurgy, forestry and wood-working and paper and pulp sector are also to a significant extent oriented towards export and thus can also enjoy good financial performance.

Table 5. Do you think one can use the enterprises' official reporting while analyzing an actual state of affairs in the national industrial sector? 2000. (% to the Number of Answers)

	Yes, we do	No, it would be better no to do it
All respondents	73	27
Sectors		
Electrical energy	79	21
Fuel industry	55	45
Ferrous metals	76	24
Non-ferrous metals	69	31
Chemicals & Petro-Chemicals	77	23
Engineering	77	23
Wood, furniture and pulp	69	31
Construction materials	73	27
Light industry	72	28
Food industry	71	29
Size of firms, employees		
1-50	60	40
51-200	69	31
201-500	73	27
501-1000	73	27
1001-2000	71	29
2001-5000	73	27
5001-10000	86	14
10001-20000	90	10
>20000	86	14
Respondents' positions		
1. Directors	73	27
2. Vice-directors	75	25
3. Department heads	72	28
4. Deputy department heads	71	29
5. Other stuffs	69	31

Sources: Survey (May 2000) by the Institute for the Economy in Transition (Russia).

The data presented in the Table shows that the bigger is the enterprise, the more reliable its reporting is. Such a correlation can be explained by the fact that larger companies find themselves under the constant supervision on the part of public agencies and have enough qualified accountants on hand. By contrast, smaller enterprises are challenged by the problem of recruiting and paying to accountants capable of mastering complex and altering accounting and reporting forms. Speaking of the respondent's position with the company, even if it has had an impact on the assessment of trustworthiness of the enterprises' official reporting, it was very insignificant and rather reverse. Such an effect is likely to be explained by the fact that junior staff enjoys greater freedom and the absence of external responsibility.

However, statistical tests have not proved the dependence of trustworthiness of the enterprises ' official reporting on the noted characteristics. First, criterion Chi-Square failed to prove the existence of a correlation for each pair (See Table 6).

Table 6. Chi-Square in the Course of Assessing the Correlation between Trustworthiness of Reporting and Main Characteristics of Enterprises

	Pearson Chi-Square			Likelihood Ratio		
	Value	df	Asymp. Sig. (2-sided)	Value	df	Asymp. Sig. (2-sided)
Firm size	10.051	8	0.261	11.245	8	0.188
Sector	11.844	13	0.540	11.164	13	0.597
Respondent position	1.730	4	0.785	1.714	4	0.788
Ownership form	2.174	4	0.704	2.468	4	0.650

Second, the hierarchic loglinear analysis procedure has equally failed to identify significant correlations in the model that comprised pair interactions between all the studied factors with assessments of trustworthiness of the official reporting. The algorithm consistently excluded all the correlations set in the original model and kept only major factors. The quality of the final model without interactions of factors remained high (see Table 7).

Table 7. The Description of the Hierarchic Loglinear Analysis Procedure

Iteration	Factor whose interaction with the dependent variable is excluded at the next iteration	Characteristics of the received loglinear model		
		G2	df	Sig
0	Without excluding	1427.74369	7136	1.000
1	Respondent position	1429.45792	7140	1.000
2	Sector	1440.62231	7155	1.000
3	Ownership form	1443.09146	7159	1.000
4	Firm size	1454.33567	7167	1.000

Notes. G2 – likelihood ratio Chi square, DF – degree of freedom, Sig – observed significance level.

Whereas the sectoral data on trustworthiness of the official reporting provided grounds to assume that distortions can arise in the event enterprises enjoy a good financial and economic health, to test the hypothesis, we used assessments of their actual financial and economic state. The respective data are collected quarterly, in the course of business surveys and may be used for the purpose of analysis. Criterion Chi-Square has not refuted the existence of a correlation between the trustworthiness of the reporting and the enterprise's current financial and economic state.

The loglinear analysis proved the above conclusions. Now, the variables used in the previous hierarchic loglinear analysis model were complemented by another three ones: namely, the actual assessment of the enterprise's state, envisaged changes of that, and the assessment of the export demand. Results of computations showed that the only factor that proved to correlate with assessments of trustworthiness of the enterprises' official reporting was the assessment of their actual financial and economic state. Correlations of all other factors with the assessment of trustworthiness of the official reporting were excluded from the model.

The loglinerar model with the linear interaction between the trustworthiness of reporting and assessments of the enterprises' actual financial and economic state enjoyed a very high quality of adjustment: the observed level of significance accounts for 1. The coefficient of the model that assesses the linear interaction between the two parameters appears statistically significant, but positive, meaning that the enterprises' official reporting becomes more reliable, once their financial and economic state improves. This conclusion contradicts our hypothesis, however, let us try to interpret it. An objective (actual) improvement of the enterprise's state makes it unnecessary to adjust its reporting for the sake of minimizing tax withdrawals. Now (with effective demand, output, and profit soaring) the enterprise has found it easier to honor its obligations towards the budget and employees rather than to take risks with adjusting its reports. However, as long as the enterprises with a poorer performance are concerned, the possible benefits (economies) from the "improvement" of their reporting overbalances the risks to bee caught on that.

Once the first try (the 2000 survey) to assess the trustworthiness has proved to be quite a success (the enterprises did not evade answering, and the interpretation of the results appeared sensible), the decision was made to proceed with monitoring. The only adjustment was that the question hence should not concern the reporting on the whole, but the trustworthiness of its major kinds: a) output and shipment; b) employment and salaries and wages; c) financial state; d) capital investment; e) stock distribution, while the variants of the reply were complemented with the "Hard to assess" option. The modifications of the question were introduced to receive more specified and grounded assessments of trustworthiness of the reporting.

Results of the 3-year-long monitoring highlighted an overall gloomy picture (see Table 8).

Table 8. Do You Think One Can Use the Enterprises' Official Reporting while Analyzing an Actual State of Affairs in the National Industrial Sector? 2001—2003. (% to the Number of Answers)

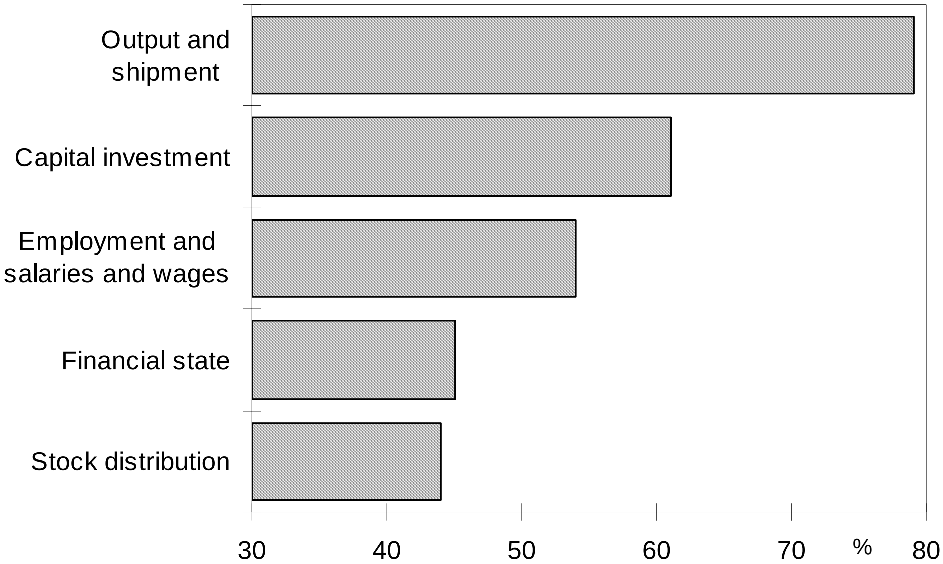
Type of reports	2001	2002	2003
Output and shipment			
yes, we do	80	80	78
no, it would be better no to do it	9	10	10
hard to assess	12	10	12
Employment and salaries and wages			
yes, we do	58	52	52
no, it would be better no to do it	24	28	31
hard to assess	18	20	17
Financial state			
yes, we do	45	44	45
no, it would be better no to do it	27	31	31
hard to assess	28	26	24
Capital investment			
yes, we do	61	59	62
no, it would be better no to do it	11	13	14
hard to assess	29	27	24
Stock distribution			
yes, we do	44	41	46
no, it would be better no to do it	21	21	19
hard to assess	35	37	36

Sources: Surveys by the Institute for the Economy in Transition (Russia).

Note: The account of the answers to the question on stock distribution concerned only AOs' responses.

First, the enterprises quite clearly confess that some kinds of their reporting are subject to adjustment more often than other. Most likely, this testifies to purposeful and regular operations by the respective divisions in this particular area. Second, there is no progress (in terms of abandonment of the shadow zone): the proportion of enterprises that consider their reporting trustworthy has practically remained unchanged. Third, the trustworthiness of the key reporting in the market conditions (on financial state and settlements, and stock distribution) has proved to be the lowest one, with less than half of the enterprises that responded to the question considering its reliable (see Fig.2).

Figure 2. The Average Trustworthiness of the Enterprises' Official Reporting for 2001—2003.



The enterprises believe that reports on output and shipment are most reliable. Perhaps, this is associated with the fact that adjusting the physical volume of output and its spatial moving is the most complex challenge. However, 10% of respondents argue that this is still possible and some enterprises practice it. By contrast, the trustworthy data on the enterprises' financial state are among most unavailable ones for the state. Nearly in all the sectors not more than half of enterprises recognize the trustworthiness of such data, with the reporting on stock distribution on the top of the list. Such a result was likely to be predictable, for the fact that the majority of Russian companies lack transparency in their property structure is common knowledge. However, at this point, sectoral results, perhaps, have highlighted new tendencies. According to the enterprises' estimates, it is the industry of building materials and forestry complex that are unbeatable in bogus reporting, while metallurgy, the fuel and electricity sectors form the opposite pole. Interestingly the percentage of a certain unreliable reporting (the share of responses "No, better not") in the construction sector is, at least, as much as twice superior to the analogous indicator of the sectors that are on the opposite pole. There may be several reasons underpinning this, of which the following forms possible and the most interesting explanation: it is not a secret to anyone that the metallurgical, fuel and electrical companies have much further advanced in their market development than others, and they have long passed through the stage of the “wild privatization" and property redistribution. Nowadays, they are challenged by the need to shape a civilized corporate image, which primarily implies such a crucial component as transparency of property. By contrast, other industry branches- and in all likelihood, primarily, the construction sector and forestry- have not yet passed through the early development stages characteristic of Russian private companies and still have to address different challenges.

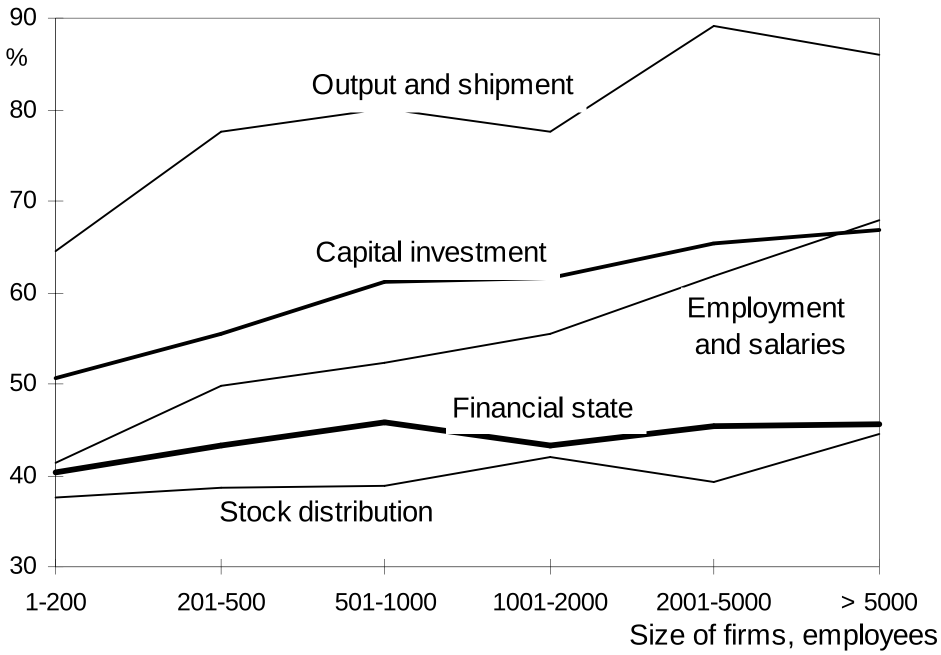
However, a statistical test with the use of Chi-square has failed to find the dependence of trustworthiness of the official reporting of any kind on the enterprise's sectoral attribution. In all the cases the value of the criterion was not great enough to argue about a correlation between the sector and the reliability of the reporting (see Table 9). But the negative result of the test in our case means that all the sectors roughly to the same extent distort their reporting.

Table 9. Chi-Square in the Course of Assessing the Correlation between Trustworthiness of Reporting and Enterprise's Sector

	Pearson Chi-Square			Likelihood Ratio		
	Value	df	Asymp. Sig. (2-sided)	Value	df	Asymp. Sig. (2-sided)
2001 г.						
Output and shipment	13.966	13	.376	15.450	13	.280
Employment and salaries and wages	8.064	12	.780	8.577	12	.739
Financial state	11.115	12	.519	11.906	12	.453
Capital investment	10.671	12	.557	11.566	12	.481
Stock distribution	3.464	11	.983	3.371	11	.985
2002 г.						
Output and shipment	12.690	11	.314	14.785	11	.193
Employment and salaries and wages	15.463	12	.217	15.547	12	.213
Financial state	11.378	12	.497	11.955	12	.449
Capital investment	8.710	11	.649	7.726	11	.738
Stock distribution	17.719	11	.088	22.521	11	.021
2003 г.						
Output and shipment	18.346	12	.106	18.939	12	.090
Employment and salaries and wages	17.204	12	.142	17.307	12	.138
Financial state	5.612	12	.934	5.678	12	.931
Capital investment	6.834	12	.868	11.261	12	.507
Stock distribution	12.235	12	.427	14.861	12	.249

The assumption that the trustworthiness of the reporting may depend on the enterprise's size was specified in the course of the evaluation of assessments of reliability of single kinds of reporting basing on the 2001-03 surveys results. The assessments of reliability averaged through the period are presented in Fig.3 and demonstrate their rise from a smaller to a bigger enterprise (see Fig.3). The greater overfall between the peak and minimum values was found between the reporting on employment and wages and the one on output and shipment. In the first case, the value accounts for 27 percentage points, and 25 p.p. for the second case. The reliability of the reporting on finance and the one on stock distribution appears almost independent on the enterprise's size (the overfall accounts for 5 and 7 p.p., respectively), but this fact is unlikely to be conceived as a positive one, for these particular kinds of reporting are most unreliable in the Russian industrial sector. In other words, enterprises of all sizes equally conceal these data from the state.

Figure 3. The Dependence of the Official Reporting Trustworthiness (Average Value for 2001—2003) on the Firms' Size



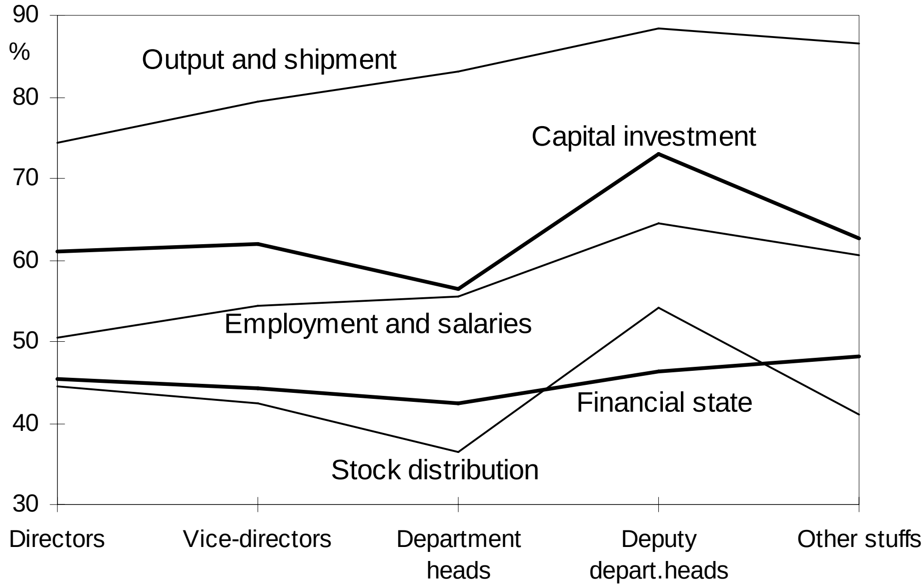
The statistical testing of the dependence of reporting on the enterprise's size proved the earlier made conclusions (see Table 10). First, the reliability of the reporting on output and shipment demonstrated a maximum correlation with the enterprise's size, followed by the trustworthiness of the reporting on employment and wages. Second, the reliability of other kinds of reporting does not depend on the enterprise's size, which becomes especially notable in the case of the reporting on finance and the one on stock distribution.

Detalization of the problem of reliability of different kinds of reporting also revealed some dependence of assessments on the respondents' position in the company (see Fig.4). As concerns the reporting on output and the one on employment and wages, one can argue that the higher is the respondent's corporate status, the more pessimism can be found in his assessments, or, at least, that the directors corps's assessments appear most pessimistic. As long as other kinds of reporting are concerned, there is no such trend in place, while all the employees appear equally pessimistic with regard to the finance reporting – the assessments of its reliability fluctuate between 42 to 48%. The maximum of such assessments belongs to regular (not senior) employees who, perhaps, are the least informed of the enterprise's financial health staff.

Table 10. Chi-Square in the Course of Assessing the Correlation between Trustworthiness of Reporting and Firm's Size

	Pearson Chi-Square			Likelihood Ratio		
	Value	df	Asymp. Sig. (2-sided)	Value	df	Asymp. Sig. (2-sided)
2001 г.						
Output and shipment	29.169	8	.000	22.551	8	.004
Employment and salaries and wages	20.212	8	.010	20.249	8	.009
Financial state	2.910	8	.940	2.895	8	.941
Capital investment	15.560	8	.049	13.604	8	.093
Stock distribution	5.786	8	.671	5.765	8	.674
2002 г.						
Output and shipment	21.134	8	.007	20.009	8	.010
Employment and salaries and wages	18.441	8	.018	18.292	8	.019
Financial state	8.008	8	.433	8.102	8	.424
Capital investment	9.327	8	.315	9.424	8	.308
Stock distribution	4.382	8	.821	4.476	8	.812
2003 г.						
Output and shipment	38.418	8	.000	36.074	8	.000
Employment and salaries and wages	15.920	8	.044	17.168	8	.028
Financial state	10.310	8	.244	10.347	8	.242
Capital investment	7.676	8	.466	8.682	8	.370
Stock distribution	11.327	8	.184	11.570	8	.171

Figure 4.The Dependence of the Official Reporting Trustworthiness (Average Value for 2001-2003) on the Respondent's Position



However, the statistical testing has failed to find a significant correlation between the trustworthiness of most kinds of reporting and the respondent's position in the company (see Table 11). It was only in 2002 that the assessments of the reporting on output and shipment and the one on employment and wages were associated with the respondents' positions. However, judging other years of observations, this particular correlation could hardly be stable. Albeit negative at first sight, such a result, anyway, has some positive effects. If the assessments of reliability of the official reporting do not depend on the respondent's position, then the fact that the staff other than the enterprise heads participate in the surveys does not introduce distortions to the results of this delicate monitoring.

The respondents representing enterprises of different property forms have proved to be yet more unanimous while assessing the reliability. The results suggests that public enterprises are not at all leaders in terms of assessing the reliability, except for they provided the highest assessments of the reporting on capital investment only. In other cases leaders were open-end joint stock companies, while the lowest assessments of the reliability of the reporting were received from the group of limited liability companies as a whole. Chi-square criterion has also failed to identify the correlation between the trustworthiness of the reporting and the enterprise's status.

Table 11. Chi-Square in the Course of Assessing the Correlation between Trustworthiness of Reporting and Respondent's Position

	Pearson Chi-Square			Likelihood Ratio		
	Value	df	Asymp. Sig. (2-sided)	Value	df	Asymp. Sig. (2-sided)
2001 г.						
Output and shipment	0.831	4	.934	0.869	4	.929
Employment and salaries and wages	6.799	4	.147	8.309	4	.081
Financial state	0.850	4	.932	0.849	4	.932
Capital investment	3.267	4	.514	5.471	4	.242
Stock distribution	2.539	4	.638	2.529	4	.639
2002 г.						
Output and shipment	12.722	4	.013	14.300	4	.006
Employment and salaries and wages	10.068	4	.039	10.170	4	.038
Financial state	0.478	4	.976	0.477	4	.976
Capital investment	3.342	4	.502	3.622	4	.460
Stock distribution	3.934	4	.415	3.861	4	.425
2003 г.						
Output and shipment	9.168	4	.057	11.355	4	.023
Employment and salaries and wages	1.166	4	.884	1.156	4	.885
Financial state	2.607	4	.626	2.611	4	.625
Capital investment	2.625	4	.622	2.883	4	.578
Stock distribution	5.876	4	.209	6.090	4	.193

The hierarchic loglinear analysis models have demonstrated insignificance of correlations between the majority of enterprises' characteristics and the reliability of the kinds of reporting in question. The enterprise's size has always correlated with the reliability of the reporting on output and shipment, once – with reliability of the reporting on employment, and once- on stock distribution. At the same time, the enterprises' financial and economic state has always correlated with reliability of the reporting on employment and wages. The testing of the sign of the noted correlations by means of loglinear models in all the cases showed the presence of a positive correlation, which quite matches the earlier drawn conclusions, except for the results that concern the enterprise's size. The assumption that the reliability of reporting grows depending on the enterprise's size was not proved by a simultaneous employment in the model of all the characteristics in question. The existence of a positive linear interaction shows that, given other conditions being equal, the reliability of reporting decreases depending on the enterprise's size.

Official Statistics or Business Surveys: Which is Better?

A direct comparison of the official statistics and BS results formed the third avenue for the IET research basing on the "surveys on surveys", albeit it is now possible to use the term "surveys on surveys and statistics". To compare such different sources of economic information, of course, requires courage, for too great are the differences, primarily in resources, coverage, and objectives. But we decided to make such a comparison on some aspects anyway, engaging our correspondents. We proceed from the following prerequisites:

First, the overwhelming majority of them respond to our questionnaires voluntarily, that is why they can consciously assess pluses and minuses of this, relatively new, source of information. Secondly, they deal with economic analysis and are likely too work with the official statistical data and other sources. Thirdly, they complete questionnaires by themselves and have (due to their position in the company) the idea on what the national industrial enterprises' official reporting is worth. Fourth, while responding to our questionnaires, they have no reason to conceal their assessments of these two kinds of economic information. Fifth, in order to avoid focusing their attention on a direct comparison, the respondents were suggested to assess yet another well-known data sources on the state of the national industrial sector. As a result, the respondents reviewed: 1) the Goskomstat's and local statistical agencies' data; 2) specialized sectoral reviews; 3) marketing research; 4) results of non-governmental surveys on enterprises; 5) analytical articles in papers, journals, radio and TV; 6) contacts with other enterprises; 7) contacts with representatives of the government agencies.

The monitoring started in 2000, and its last results are available as of 2004.

At the first stage, the respondents were offered a question on the use of the above sources of economic information, with the following variants of the answer: 1) Actually use; 2) Would like to use; 3) Refused to use. Such answers allow identification of both their current attitude towards the sources of economic information and that of the past (refused to use), as well as their intentions (would like to use).

Already the first years of monitoring (2000-04) have demonstrated interesting results (see Table 12). First, enterprise directors do not use all the data sources in an equal fashion. The average (over 5 years) frequency of the usage changes from 79% (papers, radio, TV) to 33% (marketing research). Such a difference in the frequency of the actual use, of course, can be explained primarily by to what extent the noted sources are available. But, since our major objective is to compare the official statistical data with results of non-governmental surveys, we will not focus then on assessments of other sources.

Table 12. The Dynamics of Usage of Main Sources of Information by the IET Respondents in Their Work, as % to the Number of Those Who Responded.

	2000	2001	2002	2003	2004
Goskomstat's and local statistical agencies' data					
Actually use	44	43	52	55	55
Would like to use	44	39	35	30	31
Refused to use	12	18	14	15	15
Specialized sectoral reviews					
Actually use	35	36	48	50	45
Would like to use	63	62	49	47	51
Refused to use	2	2	2	3	4
marketing research					
Actually use	28	27	34	40	35
Would like to use	68	68	61	55	61
Refused to use	4	5	5	5	4
Results of non-governmental surveys on enterprises					
Actually use	56	54	34	32	39
Would like to use	36	39	46	50	46
Refused to use	8	7	20	18	15
Analytical articles in papers, journals, radio and TV					
Actually use	79	82	76	79	81
Would like to use	15	12	12	12	11
Refused to use	6	5	12	9	8
Contacts with other enterprises.					
Actually use	63	70	69	68	71
Would like to use	36	30	29	30	27
Refused to use	1	1	2	2	2
Contacts with representatives of the government agencies					
Actually use	46	35	45	45	45
Would like to use	40	43	38	41	39
Refused to use	15	22	17	14	17

Second, the enterprise heads use the BS results on average more rarely than the official statistical data - 43 vs. 50. But the level of the usage of the BS results has undergone drastic changes over the period of monitoring. Given that in 2001 54% of respondents used them, in 2002- only 34%, which means that the surveys lost 20 percentage points! By contrast, the official statistics gained 9 p.p. over the same period. Sectoral reviews likewise experienced drastic changes – their use grew from 36 to 48%. These facts may prove that enterprises feel a growing need in the profile information, which has discontinued to appear in the Russian

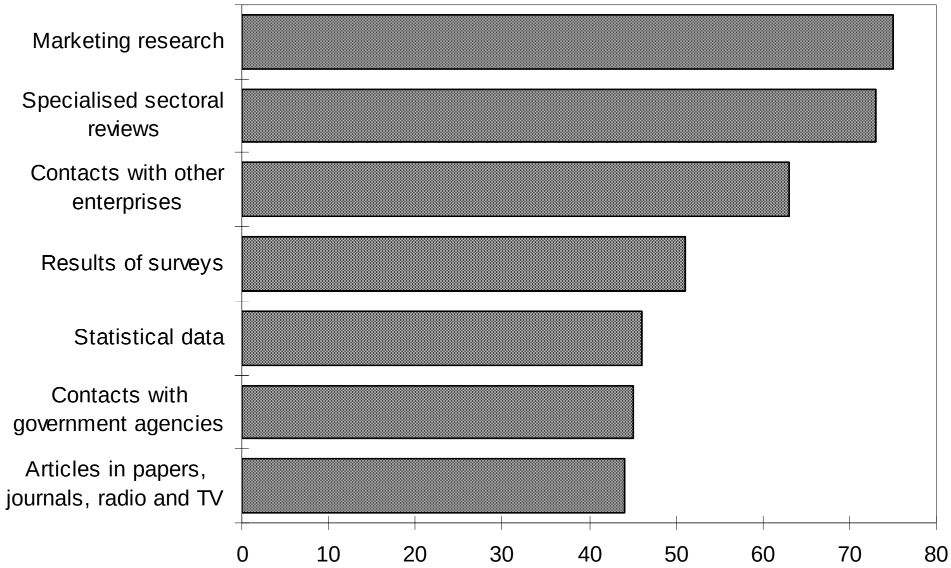
Business Survey Bulletin, or it proved to be insufficient therein. That has compelled the enterprises to focus on the official statistics and pay more attention to sectoral reviews. Indeed, for six years, until August 2001, every month IET has been disseminating results of surveys among enterprises. Such papers included several charts with dynamics of indicators across 25-30 industries, among others. Then the charts have discontinued to appear in the Bulletin, and, in all likelihood, the enterprises were in need of them. Until 2002 the level of using BS results coupled with sectoral charts was superior to the level of using the official statistical data.

Third, as long as the refusal to use in the work (i.e. the proved inefficiency) is concerned, the respondents mostly cited contacts with government agencies, followed by the official statistical data. This testifies to the fact that the RF government is not seriously concerned about information backup to the national economy. The BS results hold the third position in this negative rating — again, because of the refusal to use the sectoral charts. While prior to their exclusion the refusal rate had accounted for 7—8%, it consequently soared up to 15-20%. Had the sectoral charts still been available to enterprises, BS=s' position in the rating would be two lines down vs. the present one.

Fourth, the answer "Would like to use" points out to the failure to use or under-use of the information source, which is of a certain interest to enterprise heads, nonetheless. The champion, of course, are marketing research that, because of financial constraints and nascent business of this kind in most of Russia's regions, are hardly affordable to most of national industrial producers. They are followed by sectoral reviews, which, again, points out to the necessity of using this category of information to ensure enterprises' involvement in surveys and keeping them in the survey panel. Non-governmental BS=s hold the third position. Between 2000 through 2004 the averaged 43% of the respondents reported the desire to more extensively (vigorously) use surveys. Once the sectoral charts were excluded from the disseminated results, the index in question rose by another 9 p.p. Finally, the need in the official statistical data holds just the humble fifth position.

Fifth, the aggregate index of using economic information sources (see Fig.5) witnesses the superiority of BS vis-à-vis to the Russian official statistics (at least, in the eyes of the enterprise heads). The index was built basing on the diffusive index formula: the proportion of responses "Would like to use" plus a half of the proportion of responses "Actually use", plus a share of responses "Refused to use". The greater is the index values, the more extensive the source has already been used and the greater potential it has. For the 5-year monitoring period, it was only once, in 2002, that the annual values of the index for BS were beaten by those of the index of the official statistical reporting. Coincidentally, that was the first year after the sectoral charts had been excluded from the disseminated results. That entailed an immediate 18- point downfall in the value of the index.

Figure 5. The Aggregate Index of Using Economic Information Sources for 2001—2004



At the second stage, the respondents were proposed to assess the usefulness of the same information sources using the Russian school 5-score scale. The higher the score is, the more useful is the source of economic data for the enterprises. The ranking of the sources from the perspective of their usefulness differs from the one by their use. The IET respondents consider contacts with other enterprises to be most useful (see Table 13) — this particular source of economic information has always earned the highest score and fluctuated just slightly, with its average score accounting for 4.23. The contacts with government agencies formed the opposite pole — they received minimum score three times out of five, and the average result was 3.22.

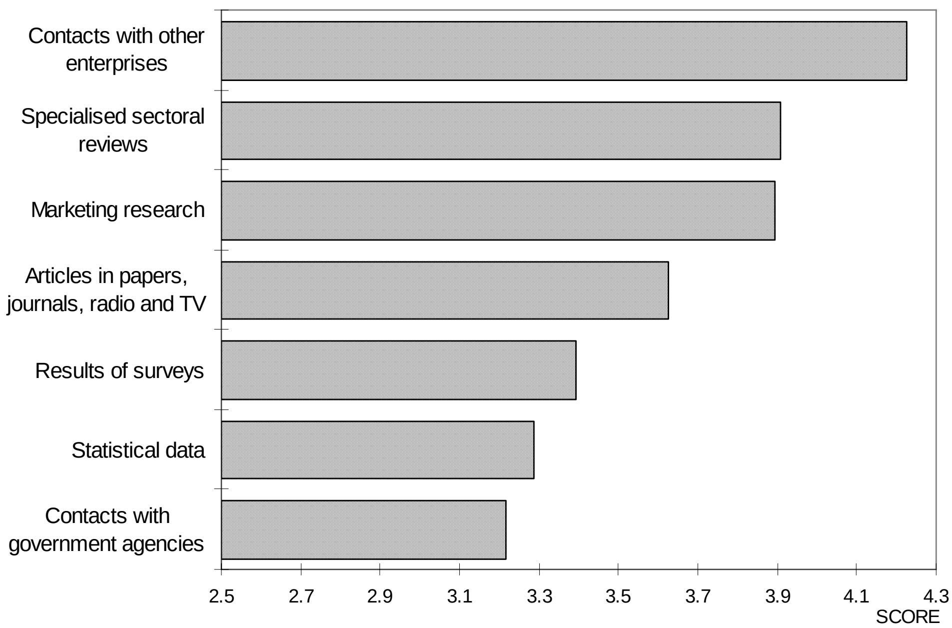
Table 13. The Usefulness of the Information Sources (school 5-score scale)

	2000	2001	2002	2003	2004
Goskomstat's and local statistical agencies' data	3.31	3.11	3.30	3.42	3.29
Specialised sectoral reviews	3.92	3.90	3.94	3.91	3.87
Marketing research	3.94	3.83	3.84	3.91	3.95
Results of non-governmental surveys on enterprises	3.59	3.63	3.14	3.34	3.26
Analytical articles in papers, journals, radio and TV	3.72	3.80	3.42	3.58	3.61
Contacts with other enterprises.	4.17	4.23	4.22	4.30	4.22
Contacts with representatives of the government agencies	3.22	2.96	3.37	3.34	3.18

The official statistics was notably lagging behind the non-governmental surveys until 2002, i.e. until the IET surveys (as we assume) had contained sectoral charts. In 2002, the usefulness of the survey results fell drastically and became lower than the one of statistics. The gap just narrowed over the consequent years. Today, the enterprises practically equally value the usefulness of the surveys and the official statistics, however the chances for the former seem more promising. IET and other NGOs find it much easier to modify the contents of their survey results, while the government statistical agencies can hardly turn friendlier towards enterprises and the society on the whole. At the same time, the steadily high rating of contacts with other enterprises testifies to the fact that the surveys basing on an informal communication with enterprise heads have a great growth potential. For example, one may conduct several surveys to specify, step by step, contents of the information that respondent collects while contacting other enterprises. It becomes possible then to complement regular questionnaires with questions that are of interest to enterprises and, most importantly, to disseminate the respective results to enterprises.

The final rating of usefulness still demonstrates a slight superiority of the non-governmental surveys vs. the official statistics (see Fig.6).

Figure 6. The Average Assessments of Information Sources Usefulness between 2001—2004.



Conclusions and Recommendations

A great number of similar regular surveys on enterprises certainly fuels competition between Russian organizations that carry them out: namely, the competition for enterprises' "voices". This necessitates monitoring of their attitudes towards surveys, particularly in the event their organizer is an NGO that does not have an administrative resource.

While regularly (annually) monitoring the reasons for enterprises to participate in surveys, one can also monitor the evolution of their attitude to those and adjust methods that ensure their further participation, primarily by modifying forms and contents of the surveys subject to return to the enterprises.

Monitoring and taking into account the structure of participants' positions is a fairly simple mission. These data form important characteristics of the quality of the survey information.

Establishing informal relations with respondents, providing their voluntary participation in surveys opens new opportunities in terms of collection of survey statistics. For example, this ensures receiving assessments of reliability of the official statistics, and the IET's several years of practice with respect to such responses has proved that such opportunities are quiet real.

The national enterprises' official reporting is partially unreliable, and they recognize it themselves, as well as the fact that different kinds of reporting appear unreliable to a different extent. It is the financial reporting and that on property structure that are most unreliable. The unreliability of reporting, as a rule, does not depend on a particular sector, size or property form of an enterprise.

Survey results have proved to be quite competitive vis-à-vis the official statistical data. Surveys can received high scores from respondents, providing their organizer is aware of their needs and takes those into account while preparing the results to be sent back to the enterprises. Questions are, what the organizer needs and what enterprises are interested to know, while the returned data should be primarily those that the enterprises are interested in, otherwise the surveys may well evolve towards the traditional statistics, meaning the enterprise's junior staff are compelled to complete huge reporting forms.

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 501—505

Book Review

Household Sample Surveys in Developing and Transition Countries,

Studies in Methods, Series F, No. 96, Statistics Division,
United Nations, New York, 2005, 655 pages

United Nations Statistics Division (UNSD) has already published several monographs related to household surveys to improve the quality of survey methodology and strengthen the capacity of national statistical systems. The most comprehensive of these initiatives over the last two decades has been the National Household Survey Capability Programme (NHSCP). The aim of the NHSCP was to assist developing countries to obtain critical demographic and socio-economic data through an integrated system of household surveys, in order to support development planning, policy formulation, and programme implementation. Furthermore, the NHSCP supported methodological work leading to the publication of several technical studies and handbooks. The Handbook of Household Surveys (Revised Edition) provided a general overview of issues related to the design and implementation of household surveys. It was followed by a series of publications addressing issues and procedures in specific areas of survey methodology and covering many subject areas, including:

- Sampling Frames and Sample Designs for Integrated Household Survey Programmes, Preliminary Version, New York, 1986.
- Sampling Errors in Household Surveys, New York, 1993.
- Survey Data Processing: A Review of Issues and Procedures, New York, 1982.
- Non-sampling Errors in Household Surveys: Sources, Assessment and Control: Preliminary Version, New York, 1982.
- Development and Design of Survey Questionnaires, New York, 1985.
- Household Income and Expenditure Surveys: A Technical Study, New York, 1989.
- Guidelines for Household Surveys on Health, New York, 1995.
- Sampling Rare and Elusive Populations, New York, 1993.

This publication updates and extends the technical aspects of the issues and procedures covered in detail in the above publications, while focusing exclusively on their applications to surveys in developing and transition countries.

The present publication presents the “state of the art” on several important aspects of conducting household surveys in developing and transition countries, including sample design, survey implementation, non-sampling errors, survey costs, and analysis of survey data. The main objective of this handbook is to assist national survey statisticians to design household surveys in an efficient and reliable manner, and to allow users to make greater use of survey generated data.

The publication's 25 chapters have been authored by leading experts in survey research methodology around the world. Most of them have practical experience in assisting national statistical authorities in developing and transition countries. Some of the unique features of this publication include:

- Special focus on the needs of developing and transition countries;
- Emphasis on standards and operating characteristics that can applied to different countries and different surveys;
- Coverage of survey costs, including empirical examples of budgeting for surveys, and analyses of survey costs disaggregated into detailed components;
- Extensive coverage of non-sampling errors;
- Coverage of both basic and advanced techniques of analysis of household survey data, including a detailed empirical comparison of the latest computer software packages available for the analysis of complex survey data;
- Presentation of examples of design, implementation and analysis of data from some household surveys conducted in developing and transition countries;
- Presentation of several case studies of actual large-scale surveys conducted in developing and transition countries that may be used as examples to be followed in designing similar surveys.

The project that led to this publication went through many interrelated phases and involved many players. The first step was to develop a proposal that clearly stated the goals of the publication, and provided a provisional table of contents with proposed chapter titles and brief descriptions of the topics to be covered, and the manner in which they were expected to be covered. Second, authors and reviewers were recruited from among well-known experts in the relevant fields. In the recruitment process, due consideration was given not just to technical expertise, but also relevant experience in developing and transition countries.

Each chapter was reviewed by two referees. The revised chapters were then assembled to produce the first draft of the publication, which was critically reviewed at an expert group meeting (EGM) organized by the United Nations Statistics Division in New York, in October 2002. At the end of the EGM, an editorial board was established to review the whole volume and make final recommendations about its structure and contents. This phase of the review process led to a restructuring and streamlining of the whole volume to make it more coherent, more complete, and more internally consistent. A total of 25 chapters were selected for the final version of the publication. Each of these chapters then went through a third round of review by two referees before a final decision was taken on whether or not to select it for inclusion in the publication.

A team of editors then did a final review of the whole volume, ensuring that the material presented was technically sound, internally consistent, and faithful to the primary goals of the publication. A detailed description of the contents of the publication is provided in the following overview.

The publication's 25 chapters have been authored by leading experts in survey research methodology around the world. As has been stressed, most of them have practical experience in assisting national statistical authorities in developing and transition countries.

The publication is divided into two parts. **Part one** consists of ***21 chapters*** dealing with survey design, implementation, and analysis. It is divided into five sections, labelled A through E. The following is a summary of the contents of each section of part one.

Section A: ***Survey design and implementation***. This section contains three chapters. Chapter II presents an overview of various issues pertinent to the design of household surveys in the context of developing and transition countries. Chapters III and IV, discuss issues pertaining to questionnaire design and issues pertaining to survey implementation, respectively, in developing and transition countries.

Section B: ***Sample design***. This section contains an introductory note and three chapters dealing with the specifics of sample design. Chapter V deals with the design of master samples and master frames. The use of design effects in sample design and analysis is discussed in chapter VI and chapter VII provides an empirical analysis of design effects for surveys conducted in several developing countries.

Section C*:* ***Non-sampling errors***. This section contains an introductory note and four chapters dealing with various aspects of non-sampling error measurement, evaluation, and control in developing and transition countries. Chapter VIII deals with non-observation error (non-response and non-coverage). Measurement errors are considered in chapter IX. Chapter X presents quality assurance guidelines and procedures with application to the World Health Surveys, a programme of surveys conducted in developing countries and sponsored by the World Health Organization (WHO). Chapter XI describes a case study of measurement, evaluation, and compensation for non-sampling errors of household surveys conducted in Brazil.

Section D: ***Survey costs***. This section contains an introductory note and three chapters. Chapter XII provides a general framework for analysing survey costs in the context of surveys conducted in developing and transition countries. Using empirical data, chapter XIII describes a cost model for an income and expenditure survey conducted in a developing country. Chapter XIV discusses issues pertinent to the development of a budget for the myriad phases and functions in a household survey and includes a number of examples and case studies that are used to draw comparisons and to illustrate the important budgeting issues discussed in the chapter.

Section E: ***Analysis of survey data***. This section contains an introductory note and seven chapters devoted to the analysis of survey data. Chapter XV provides detailed guidelines for the management of household survey data. Chapter XVI discusses basic tabular analysis of survey data, including several concrete examples. Chapter XVII discusses the use of multi-topic household surveys as a tool for poverty reduction in developing countries. Chapter XVIII discusses the use of multivariate statistical methods for the construction of indices from household survey data. Chapter XIX deals with statistical analysis of survey data, focusing on the basic techniques of model-based analysis, namely, multiple linear regression, logistic regression and multilevel methods. Chapter XX presents more advanced approaches to the analysis of survey data that take account of the effects of the complexity of the design on the analysis. Finally, chapter XXI discusses the various methods used in the estimation of sampling errors for survey data and also describes practical data analysis techniques, comparing several computer software packages used to analyse complex survey data. The strong relationship between sample design and data analysis is also emphasized. Further details on the comparison of software packages, including computer output from the various software packages, are contained in the CD-ROM that accompanies this publication.

Part two of the publication, ***containing four chapters*** preceded by an introductory note, is devoted to case studies providing concrete examples of surveys conducted in developing and transition countries. These chapters provide a detailed and systematic treatment of both user-paid surveys sponsored by international agencies and country-budgeted surveys conducted as part of the regular survey programmes of national statistical systems. ***The Demographic and Health Surveys*** (DHS) programme is described in chapter XXII; the ***Living Standards Measurement Study*** (LSMS) surveys programme is described in chapter XXIII. The discussion of both survey series includes the computation of design effects of the estimates of a number of key characteristics. Chapter XXIV ***discusses the design and implementation of household budget surveys,*** using a survey conducted in the Lao People’s Democratic Republic for illustration. Chapter XXV discusses ***general features of the design and implementation of surveys conducted in transition countries,*** and includes several cases studies.

Household surveys are an important source of socio-economic data. Important indicators to inform and monitor development policies are often derived from such surveys. In developing countries, they have become a dominant form of data collection, supplementing or sometimes even replacing other data collection programmes and civil registration systems.

The English version of this publication has now been web-published on the UNSD website: http://unstats.un.org/unsd/hhsurveys/. As this is the PDF file of the camera-ready version submitted for printing, it is possible to use it directly. The translation into the other 5 official UN languages (French, Spanish, Arabic, Chinese and Russian) will take a bit longer.

I am confident that this publication will help numerous colleagues in developing and transition countries, and other interested parties, in designing and implementing household surveys.

Prepared by Jan Kordos

STATISTICS IN TRANSITION, September 2005
Vol. 7, No. 2, pp. 507—508

ANNOUNCEMENT

European Conference on Quality in Survey Statistics (Q2006)
Cardiff, UK, 24—26 April 2006

Hosted jointly by the UK Office for National Statistics and Eurostat, the Q2006 conference will be the third in a series of scientific gatherings covering important methodological and quality-related topics of relevance to the European Statistical System.

The conference will focus on topics relating to quality criteria such as: accuracy, relevance, timeliness, accessibility, comparability and coherence of survey statistics.

With quality a major concern for all statistical organisations and users, this conference provides an ideal forum for the exchange of ideas and experiences.

The Director Generals of Eurostat and the European National Statistical Institutes have committed to a strategy of continuous quality improvement, as a result of the work accomplished by the European Leadership Expert Group (LEG) on Quality in 2001.

A clear need to improve co-operation and exchange in the field of quality in statistics was identified. Q2006 will build on the successful gatherings at the International Conference on Quality in Official Statistics in Stockholm (Q2001) and the European Conference on Quality and Methodology in Official Statistics in Mainz (Q2004).

These conferences will now be held biennially.

Call for Abstracts

The Q2006 Programme Committee invites anyone interested in presenting a contributed paper to submit their abstracts via email to q2006@ons.gov.uk

Abstracts should not exceed 500 words and should contain up to three key words, as well as your contact details. The submission deadline is 7 November 2005. Accepted abstracts will be announced in December 2005 and will be inserted in the delegate handbook for distribution at the conference.

The Organizers hope to receive contributed papers on a variety of quality topics, including:

Eurostat developments

- Status of the implementation of the Leadership Expert Group on quality recommendations

Measuring and improving the quality of survey statistics

- Measuring quality
- Measuring accuracy in complex surveys
- Quality reporting and quality indicators
- Tools for the assessment and management of data quality
- Improving process quality

Measuring and improving the quality of statistical organisations

- Implementation of quality management models
- The user-producer dialogue/service level agreements
- Relationships with data suppliers
- The quality of statistical systems
- Information management in statistical institutes
- Training for quality

Survey design

- Mixed mode data collection
- Data collection methods and technology
- Questionnaire testing methods
- Confidentiality and disclosure control
- Data processing
- Data analysis
- Metadata and documentation

Administrative data

- Quality and quality assurance in the use of administrative data
- Boundary crossing data use

Key Dates:

7 November 2005	Deadline for contributed abstracts
16 December 2005	Notification of accepted abstracts
17 February 2006	Closing date for discounted registration
10 March 2006	Closing date for discounted accommodation bookings
17 March 2006	Closing date for registration

If you have any questions about the conference please contact the Organizers by: email: q2006@ons.gov.uk