STATISTICS IN TRANSITION new series, March 2017
Vol. 18, No. 1, pp. 91–114

TRADE PATTERN ON WARSAW STOCK EXCHANGE AND PREDICTION OF NUMBER OF TRADES

Henryk Gurgul[1], Artur Machno[2]

ABSTRACT

The main goal of this paper is to present the method for describing and predicting trade intensity on the Warsaw Stock Exchange. The approach is based on generalized linear models, the variable selection is performed using shrinkage methods such as the Lasso or Ridge regression. The variable under investigation is the number of trades of a particular stock 5-minute interval.
The main conclusion is that the number of trades during short intervals is predictable in the sense that the prediction, even based on relatively simple models, is with respect to statistical properties better than the prediction based on the random walk, which is used as a benchmark model.

Key words: high frequency data, daily trade pattern, Warsaw Stock Exchange, market microstructure.

JEL classification: C53, G17.

1. Introduction

Trade intensity in a high frequency setting is an interesting and important topic. Standard models for time series are invalid in a high frequency world for many reasons. For example, returns on stocks, in classic time series models, assume a continuous distribution. This assumption is clearly not met, because of the tick size, possible changes in price and multiple tick sizes. In standard time series analysis, e.g. daily or weekly data, this simplification is acceptable. However, for example for 1-minute returns, the probability of a return being equal to zero is very high, close to one, which makes this assumption materially incorrect. Another issue is the exact time of a transaction, for 1-minute data it

[1] Department of Applications of Mathematics in Economics, Faculty of Management, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: henryk.gurgul@gmail.com.

[2] Department of Applications of Mathematics in Economics, Faculty of Management, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: artur.machno@gmail.com.

might be significant if times of trades are reported with a margin of error of one or two seconds.

One approach to investigating the trade intensity is to analyze the time between trades. The most popular model based on this approach is the Autoregressive Conditional Duration (ACD) model. The ACD model is conceptually neat. Generally, the basic ACD model assumes a conditionally exponential distribution of the time to the next transaction. The assumption is not met for the trade of stocks. The involvement of more complex distributions makes an ACD estimation computationally heavy. Additionally, the U-shape of the daily trade pattern makes the estimation much more difficult. It is relatively easy to overcome this problem using a proper data transformation, although the results for transformed data are more difficult to interpret.

Those concerns have motivated us to analyze trade intensity in a different way. The proposed method is conceptually relatively easy, although most of the inference possible for ACD is achievable. The idea is to regress the number of trades of a particular stock during a 5-minute interval, the choice of the length of interval is arbitrary and chosen as an example. From the data analyst's point of view, this is a prediction problem with the independent variable being a count variable. Dependent variables in this situation can be chosen freely. In this paper we use the characteristics from the previous interval as regressors. They are the number of trades, the number of trades with a higher price, the number of trades with a lower price and the value of all transactions.

Additionally, the trade pattern is analyzed. This is not the core of this research, thus it is not discussed fully. The trade pattern is visible and not considering it in the prediction would be incorrect. We use statistical methods to cluster the intervals in such a way that in every cluster the trade intensity is statistically similar.

Prediction models are based on the linear and generalized linear model (GLM). The authors are aware that the results can be improved in terms of prediction power by using more sophisticated machine-learning techniques. However, the goal of this analysis is to show that the prediction is manageable and the authors try to present the relation between the number of trades and other factors. This inference would be difficult for machine-learning techniques, which are usually "black boxes".

2. Literature overview

The progress of computational techniques and trading methods has recently had a massive impact on the most important topics in financial research reflected in the financial literature. One of the most important directions of both theoretical and empirical research is the market microstructure. On the basis of intraday data numerous researchers try to describe the price, volatility and trading volume and the process of their formation. One of the first streams of research is represented in a paper by De Jong and Rindi (2009). Similarly to other researchers, the authors try to analyze the market structure and create market designs. The authors aim to find an effect of these factors on the intraday price formation. In the last decade researchers obtained access to tick-by-tick data and thus other intraday high frequency data. The availability of high frequency data has made it possible to prove the theories of market microstructure empirically.

One of the first topics analyzed empirically in the framework of market microstructure was intraday price dynamics. From a theoretical (statistical) point of view, the calendar-time distribution of stock price dynamics on small scales of time depends on both the distribution of price changes and the distribution of duration. These empirical studies are strongly interrelated with the observation of trading volume given by tick-by-tick data. The intraday behavior of stock prices was tested by Engle (2000). He noticed that the largest rise in the volume of transactions can be observed at the open and at the close of the market. On the basis of existing theories and this observation, Engle explained the U-shaped pattern of volatility over the course of the day. The research based on the intraday data supplied evidence that the size of the time intervals is also essential at long-time scales, which contradicts claims made using traditional stock price models. Numerous papers on market microstructure including Diamond and Verrechia (1987), Easley and O'Hara (1992), Engle and Russell (1998), Engle (2000), Dufour and Engle (2000), Manganelli (2005) and Cartea and Meyer-Brandis (2010) have confirmed that the duration at high frequencies between trades supplies constitutes important information about the intraday dynamics of tick-by-tick trades. This information reflects the behavior of the market, the differences in the market activity of uninformed or informed market participants, the volatility of price changes and the implied volatility from the option markets.

Numerous empirical papers confirm that duration, as a random variable, can be considered as one of the most significant factors determining stock price behavior, especially with respect to short periods of time. As we just mentioned above, in the past this random variable was not usually taken into account in most asset pricing models with horizons of more than one day. Those who research stock market behavior on the basis of daily (or more aggregated) data assumed that any effect of durations is dissipated immediately. Nowadays, many trades are made within algorithmic trading processing on a tick-by-tick level. Thus, the duration between trades is included in the most important explanatory variables in most recent research. It delivers important information about the stock market behavior over short-time intervals.

The main goal of most trading strategies is to benefit from the price patterns and behavior of prices, volatility and trading volume over ever shrinking scales of time. Empirical studies suggest that the amount of time required to complete a trade has decreased in the last couple of years by a digit. Nowadays, most trades are conducted very quickly over short periods of time. This extremely fast trading has become one of the most popular kinds of trading (especially algorithmic trading) among market participants. The important aim of both theoretical and empirical research was the identification of the main factors which enhance the fast expansion of algorithmic trading. The impulse for speeding up the trades comes from essential changes in the market structure, observed especially in the last decade. In addition, in recent years the capacities of computers have significantly increased. However, the cost of even powerful computers has significantly decreased. These factors have increased not only the number of market participants but also raised the speed of trades on all stock markets.

Most researchers who write about duration just one paper of Engle and Russell (1998). Their autoregressive conditional duration (ACD) model maps the time of arrival of financial data. This model is a starting point for numerous authors who aim to generalize the ACD model in different ways. Probably the best known and most frequently applied examples are the logarithmic model of Bauwens and Giot (2000) and extended class of models by Fernandes and Grammig (2005). Some other generalizations refer to regime-switching and mixture ACD models. They are referenced in Maheu and McCurdy (2000), Zhang et al. (2001), Meitz and Terasvirta (2006), Hujer et al. (2002). The structural model for durations between events and associated marks is presented in a paper by Renault et al. (2012). An extensive review of different duration models which reflect duration can be found in Bauwens and Hautsch (2009).

Considering the literature as a whole, high frequency trade pattern is one of the most important topics from both the theoretical and empirical points of view. The trade pattern is also a very interesting topic, especially in the case of the stock markets in countries in transition like the Warsaw Stock Exchange (WSE). The Polish economy (and stock market) is the largest among the economies in transition from CEE. However, in relation to bigger economies such as the German economy it is small. According to recent empirical studies, all stock markets react to news from the US stock market in terms of price, volatility and trading volume performance. Trade on NYSE starts at 3 p.m. CET (Polish time), therefore in addition to the usual U-shape of the trade pattern, one can observe a fluctuation at 3 p.m. Trade pattern analysis is used later in this paper for a description of time in the model used for the predictions.

The goal of this article is to analyze the number of trades in short periods on WSE. The prediction is based on GLM and shrinkage methods. Recent research on GLM is presented by Friedman et al. (2010). The shrinkage methods which are used in this article are presented by Tibshirani (1996), a recent contribution on shrinkage methods can be found in Zou and Hastie (2005). An introduction to GLM models and shrinkage methods in a broader context can be found in Hastie et al. (2008). The analysis consists of two main parts. Firstly, a visualization of the trade pattern in terms of the number of trades is presented. Secondly, the task of predicting the number of trades during small intervals is undertaken. The analysis is conducted for 19 of the 20 biggest Polish listed companies in the period from January 1, 2014 to September 22, 2014. The only company that has been omitted was Orange Polska, which was preceded by Telekomunikacja Polska. The change of ownership and the name caused a technical problem and unreliable results, therefore the company has been removed from the analysis. The time interval that has been chosen for the analysis (frequency) is 5 minutes.

3. Data description and its daily pattern

3.1.Data description

The electronic system of the WSE has been changed since August 8, 2013; the most important change from the data analyst's point of view is the increased preciseness of the trade time from seconds to microseconds. This precision is very convenient from the researcher's point of view since in the period where microseconds are present no cases of two transactions taking place at the same time are reported. The trade hours on the WSE are 9:00 a.m. to 5:05 p.m. However, there is a break from 4:50 p.m. to 5:00 p.m. and during the last 5 minutes trade is not conducted as it is during normal hours, so transactions which took place after 4:50 p.m. are excluded from the analysis. This leaves us with 470 minutes of trade each day.

All transactions on the WSE are quoted in Polish zloty (PLN). At the time this publication is being prepared 1USD=3.95PLN and 1EUR=4.38PLN. In the text we present all values in PLN.

The stocks which form the main Polish equity index, WIG20, have been chosen for the analysis during the period January 1, 2014 to September 22, 2014. Table 1 summarizes the composition of WIG20; note that the analysis is performed for 19 out of 20 stocks from WIG20.

Table 1. The composition of the WIG20 index

Company's name	**Abbr. Name**	**Prime line of business**	**Index weightening (%)**
Alior Bank	ALIOR	Finance	1.67
Asseco Poland	ASSECOPOL	Software	1.94
Bank Pekao	PEKAO	Finance	12.21
Bank Zachodni WBK	BZWBK	Finance	5.62
Eurocash	EUR	FMCG	1.15
Grupa LOTOS	LOTOS	Oil and Natural Gas	0.84
Jastrzębska Spółka Węglowa	JSW	Mining	0.64
Kernel Holding	KERNEL	Food	0.59
KGHM Polska Miedź	KGHM	Mining	8.78
LPP	LPP	Trade	6.08
Lubelski Węgiel „Bogdanka" SA	BOGDANKA	Mining	1.91
mBank	MBANK	Finance	3.19
Orange Polska	ORANGEPL	Telecommunication	3.29
PKN Orlen	PKNORLEN	Fuels	6.85
PKO Bank Polski	PKOBP	Finance	14.44
Polska Grupa Energetyczna	PGE	Energy	6.72
Polskie Górnictwo Naftowe i Gazownictwo	PGNIG	Oil and Natural Gas	4.09
PZU SA	PZU	Insurance	14.09
Synthos	SYNTHOS	Chemistry	1.05
Tauron	TAURONPE	Energy	2.82

Note: The Orange Polska has not been used in the analysis.

The variable being studied is the number of trades in a 5-minute interval for a particular stock. Therefore, the analysis consists of 19 regression problems. The raw data has been transformed into 5-minute data with four variables for each of 19 assets:

- *Number*- total number of trades
- *Plus*- total number of trades with a higher price than the preceding trade
- *Minus*- total number of trades with a lower price than the preceding trade
- *Volume*- total number of shares traded

In addition, the categorical variable *time*, which indicates the time interval, is added and there are 92 levels of this variable. A more in-depth analysis of the time variable is given later in this paper. Additionally, the data set has been divided into training and testing datasets. The training sample is the sample which is used for estimation. The training set is used for the out-of-sample performance analysis and it is not touched during the estimation.

Table 2. Partition of the dataset

Dataset	No. of days	Beginning	End
Total	183	January 1, 2014	September 22, 2014
Training	120	January 1, 2014	June 25, 2014
Testing	63	June 26, 2014	September 22, 2014

Later in this paper, the results presented are derived from the training set if not stated otherwise. Table 3 consists of descriptive statistics for the number of trades for all assets considered. Figure 1 shows the distribution of all four numerical variables considered per asset. Note the logarithmic scale on the graph. The differences in the number of trades and the number of shared traded are considerable across the assets.

Table 3. Descriptive statistics of the number of trades in 5-minute intervals

Company	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PKOBP	0	14	27	37.71	49	520
PZU	0	9	18	25.94	34	459
PEKAO	0	8	16	22.29	29	420
KGHM	0	15	28	40.65	50	1075
PKNORLEN	0	5	12	18.68	24	432
PGE	0	11	22	29.09	38	472
LPP	0	0	1	4.60	5	227
BZWBK	0	3	9	14.20	19	551
PGNIG	0	5	11	17.67	22	483
MBANK	0	1	5	8.79	11	219
TAURONPE	0	3	7	10.98	14	246

Table 3. Descriptive statistics of the number of trades in 5-minute intervals (cont.)

Company	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ASSECOPOL	0	1	4	5.98	8	165
BOGDANKA	0	0	2	4.89	6	189
ALIOR	0	1	3	6.49	8	356
EUROCASH	0	2	7	12.27	16	489
SYNTHOS	0	1	4	7.28	9	273
LOTOS	0	1	4	7.52	9	371
JSW	0	4	8	11.68	15	325
KERNEL	0	2	6	11.38	14	374

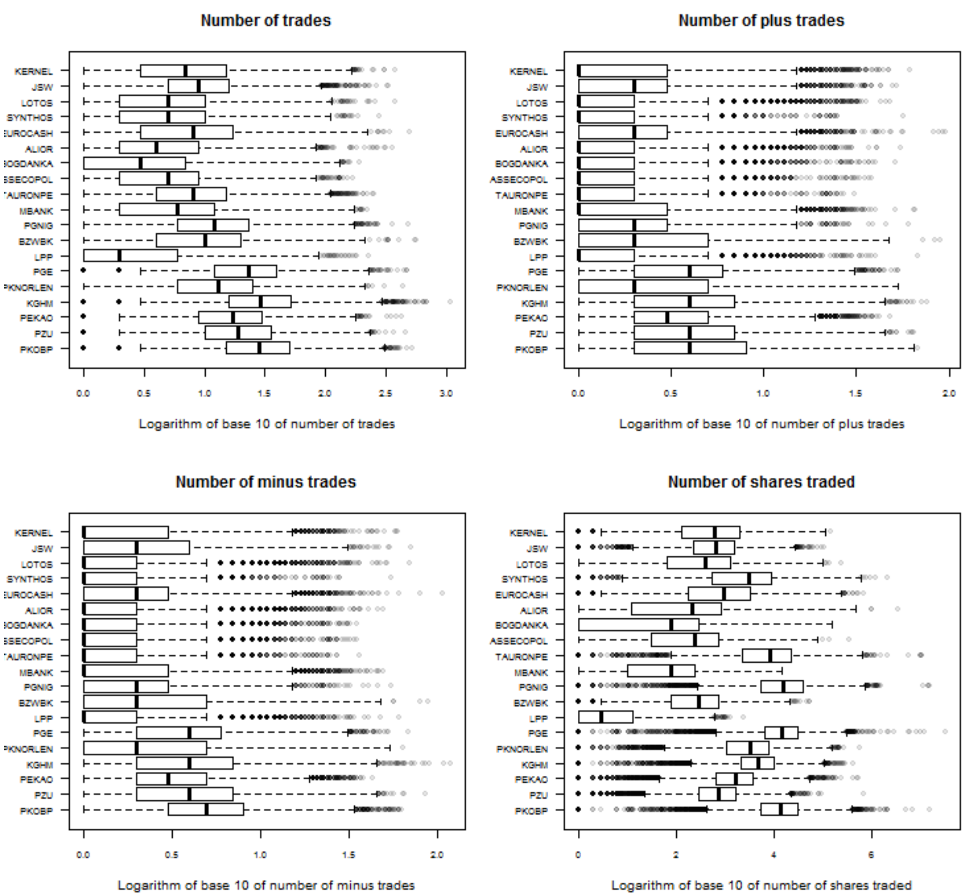

Figure 1. Boxplots of the four analyzed numerical variables for all assets.

Note: "Plus trade"- trade with a higher price than the preceding trade, "minus trade"- lower.

Logarithms are taken of raw values enlarged by one. Outliers are plotted with 90% transparency if the area is black, 10 or more points are plotted in it.

3.2. Daily pattern on WSE

The number of transactions displays a U-shape, which is characteristic of the daily pattern. Just after opening, trade intensity is very high, it drops significantly after a couple of minutes and then a slow downward trend is observed until midday. After that the trend changes to a slow upward one and starts to increase quickly around 3 p.m. The last five minutes are again significantly more intensive. A similar pattern is observable for the other variables considered (plus, minus and volume).

Figures 2 and 3 display the trade pattern for two sample companies, the biggest one - PKOBP and the smallest one - KERNEL. The PKOBP possesses a clear U-shape for all variables, although the KERNEL does not have visibly higher values in the morning (excluding the first interval).

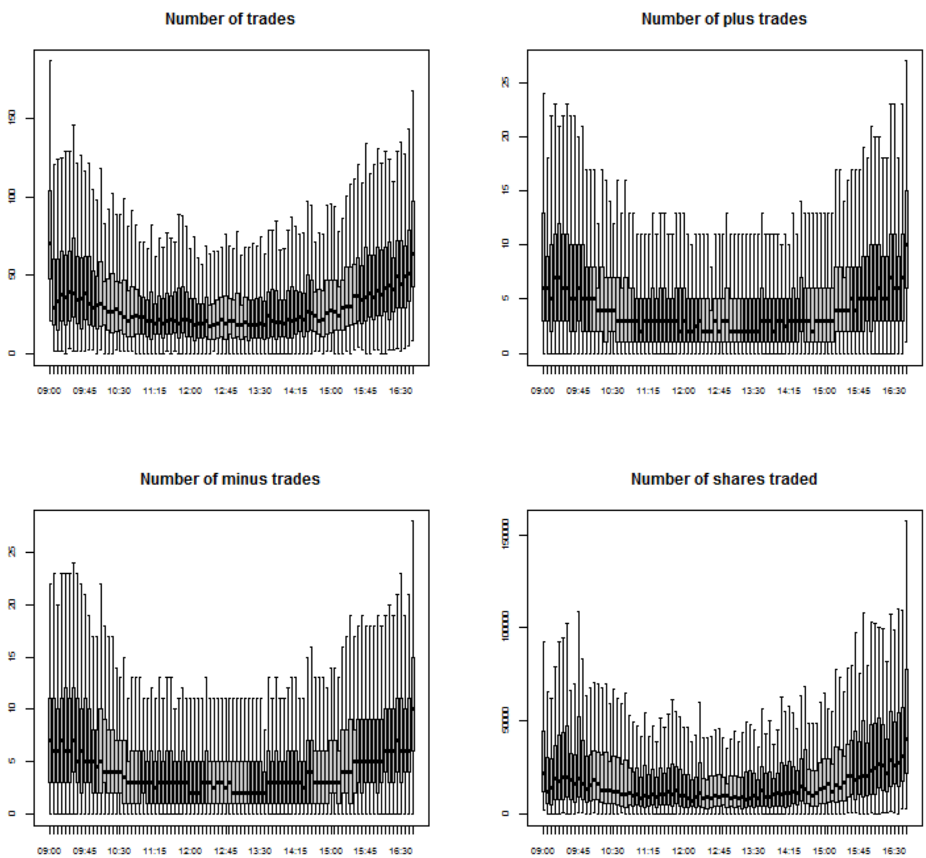

Figure 2. Boxplots of the four considered numerical variables by time for PKOBP.

Note: "Plus trade"- trade with a higher price than the preceding trade, "minus trade"- lower. The outliers have not been plotted.

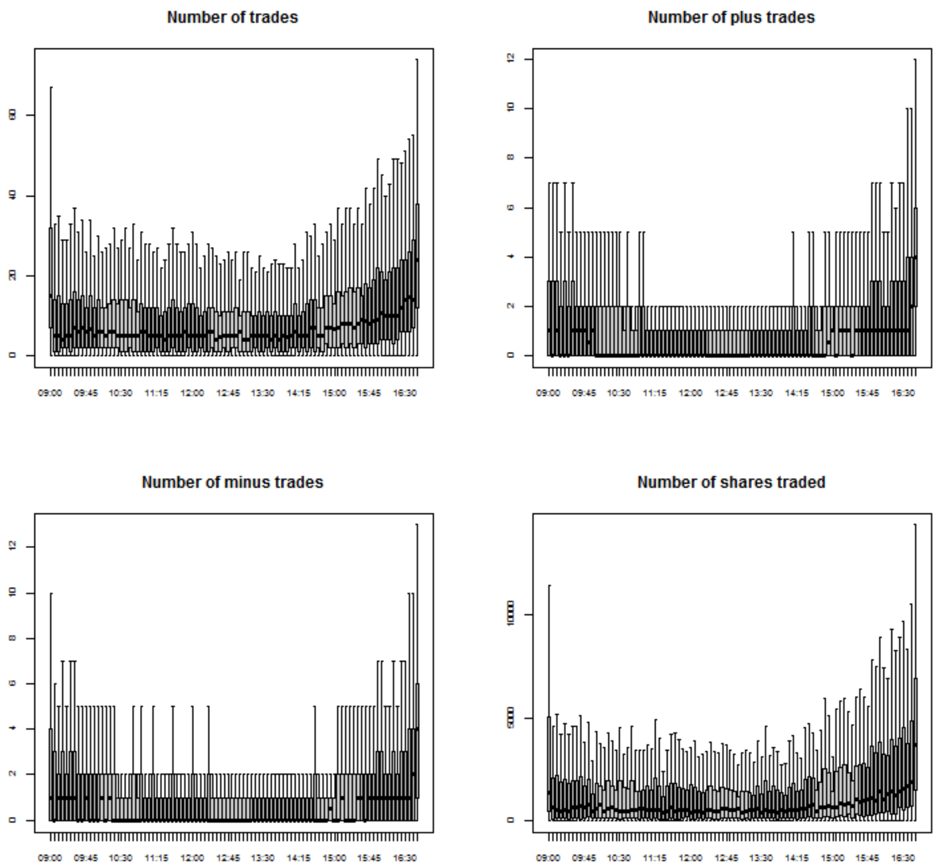

Figure 3. Boxplots of the four considered numerical variables by time for KERNEL.

Note: "Plus trade"- trade with a higher price than the preceding trade, "minus trade"- lower. The outliers have not been plotted.

The main goal of this section is to construct a factor time variable. A factor variable with 92 levels is not practical since if one wanted to use it in a regression model it would have 92 dummy variables. It seems that trade intensity is not significantly different, e.g. during intervals 10:00-10:05 and 10:05-10:10. However, we need a technique to group intervals.

In order to capture the daily pattern differences across names and across variables we use the following procedure:

I. Standardize 76 numerical variables (with zero mean and unit variance); logarithms of number, plus, minus and volume for 19 stocks.

II. Perform Principal Component Analysis (PCA) on standardized variables.

III. Take the first component as a new variable, named PC1, of the PCA.
IV. Perform hierarchical clustering of time intervals (92 levels of the time factor variable) using the PC1 as an explanatory variable.
V. Cluster the intervals by PC1 using K-means clustering.
VI. Divide a cluster if it does not contain a joint set of intervals.

Standardization is a common preliminary step in the PCA, it is done in order to avoid the PCA being driven by variables which have the highest absolute values. The data is skewed and thus the logarithms of variables are taken. PCA is performed on all variables in order to capture potential specific features of particular stocks or variables. Step IV is mainly used in order to visualize the clustering and choose the number of clusters. The complete linkage and average linkage are used and both suggest 4 clusters at most. K-means clustering results in the following clusters:

A. 9:00-9:05 and 16:45-16:50, 2 intervals.
B. 9:05-11:00 and 14:50-15:50, 35 intervals.
C. 11:00-14:50,46 intervals.
D. 15:50-16:45, 11 intervals.

Finally, clusters A and B are divided in order to construct clusters in the form of joint intervals.

Figure 4 summarizes the clustering of time variables. The result is very intuitive and expected. Both extreme intervals constitute separate factors. There are two moderately intensive periods, 9:05-11:00 and 14:50-15:50. The period of the lowest intensity for one consistent cluster is 11:00-14:50. The period 15:50-16:45 is characterized by a significant number of outliers in comparison with others.

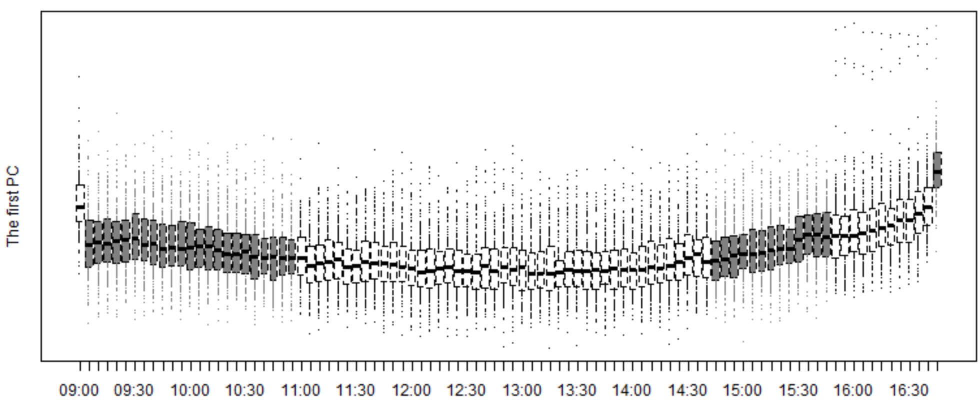

Figure 4. Boxplots of the PC1 variable by time.

Note: The clustering into 6 factor variables is presented by changes in color. Boxplots do not have whiskers, all observations outside quartile ranges are plotted as dots.

4. Forecasts by selected models

The main goal of this article is to verify whether the number of trades during short intervals is predictable. However, we do not aim to find the best algorithm for prediction. The models chosen are based on the GLM. We use a relatively large number of independent variables. In order to avoid overfitting we use two shrinkage methods: ridge regression and lasso regression. The biggest advantage of the linear model is its simplicity and ease of interpretation.

We test four models and compare the performance to the simple random walk. Let $\{N_{it}\}_{t=1}^{T}$ be the series of the number of trades in the t-th 5-minute interval for the i-th company; analogously $\{P_{it}\}_{t=1}^{T}$-the number of trades with a higher price than the preceding trade; $\{M_{it}\}_{t=1}^{T}$- the number of trades with a lower price than the preceding trade; $\{V_{it}\}_{t=1}^{T}$- volume (number of shares traded).

For the sake of simplicity, let $\left\{X_{jt}\right\}_{t=1}^{T}$ be the set of all observed variables, $j = 1, \dots, 92$.

4.1. Random Walk

The random walk strategy takes the most recent number of trades as the best prediction of future change. The model is defined by:

$$N_{it+1} = N_{it} + a_{it}; \tag{4.1}$$

whereN_t is the actual number of trades at period t and a_{it} is white noise.

Therefore

$$\widehat{N}_{it+1} = N_{it}; \tag{4.2}$$

where$\widehat{N}_{t+1}$ is the forecast of the number of trades for the following period.

4.2. Ridge linear regression

The linear regression prediction is defined by:

$$\widehat{N}_{it+1} = \beta_0^i + \sum_{j=1}^{92} \beta_j^i X_{jt} + \sum_{k=2}^{6} \tau_k^i T_{kt+1}\,; \tag{4.3}$$

where $\beta_0^i, \dots, \beta_{92}^i$ are coefficients for the numeric variables; T_{kt} is a dummy variable which equals one if the t-th observation belongs to the k-th interval defined in section 3.2 and visualized on Figure 4, and zero otherwise; $\tau_2^i, \dots \tau_6^i$ are intercept changes in comparison with the first interval (9:00-9:05).

The ridge regression is an estimation method which was originally derived in order to solve the problem of colinearity in the multiple linear regression model. This type of regression is similar to the ordinary least squares (OLS), except that the coefficients are estimated by minimizing a different quantity. The ridge regression coefficient estimates are the values that minimize:

$$RSS_i + \lambda \sum_{j=1}^{92} {\beta_j^i}^2 ; \tag{4.4}$$

where

$$RSS_i := \sum_{t=1}^{T-1} \left(N_{it+1} - \beta_0^i - \sum_{j=1}^{92} \beta_j^i X_{jt} - \sum_{k=2}^{6} \tau_k^i T_{kt+1} \right)^2 ; \tag{4.5}$$

whereλ is a tuning parameter. The tuning parameter is chosen in order to minimize in-sample mean squared error (MSE).

4.3. Lasso linear regression

The prediction for lasso linear regression is the same as for ridge linear regression, see (4.3). The difference is in the estimation method. In fact, the difference is in the shrinkage penalty. The estimates of the lasso regression coefficients are the values that minimize:

$$RSS_i + \lambda \sum_{j=1}^{92} \left| \beta_j^i \right| . \tag{4.6}$$

From a practical point of view, the main difference between the ridge and lasso penalty functions is that the latter works as a variable selection method. When lambda increases, the number of coefficients which equal zero also increases. In the case of the ridge regression, when lambda increases, the absolute value of the coefficients decreases to zero, but is not equal to zero. We refer to Tibshirani (1996) for more details on this method.

4.4. Ridge Poisson regression

The number of trades is an integer variable, thus a linear regression approach is not the natural one. Firstly, the fitted values might be negative. Secondly, the fitted values in linear regression are continuous, meaning that the prediction is not an integer. In fact, using linear regression, one assumes normal distribution of the forecast with certain parameters, specifically with the mean given by (4.3).

The distribution of the dependent variable under the Poisson regression is given by:

$$P(\widehat{N}_{it+1} = k) = \frac{\Lambda^k}{k!} e^{-\Lambda}; \tag{4.7}$$

where

$$log(\Lambda) = \beta_0^i + \sum_{j=1}^{92} \beta_j^i X_{jt} + \sum_{k=2}^{6} \tau_k^i T_{kt+1}. \tag{4.8}$$

The Poisson regression is simply GLM with the Poisson distribution and the exponential link function. The standard estimation is performed using the maximum likelihood (ML) method. Combining the ML with the ridge shrinkage method, the coefficients are the values which minimize:

$$LogLik_i + \lambda \sum_{j=1}^{92} {\beta_j^i}^2 ; \tag{4.9}$$

where

$$LogLik_i \coloneqq \sum_{t=1}^{T-1} \left(N_{it+1}\Theta_i - e^{\Theta_i} - \log(N_{it+1}!) \right); \tag{4.10}$$

and

$$\Theta_i \coloneqq \beta_0^i + \sum_{j=1}^{92} \beta_j^i X_{jt} + \sum_{k=2}^{6} \tau_k^i T_{kt+1}. \tag{4.11}$$

This setting is mathematically quite complex, although it is very intuitive. Formula (4.7) tells us that the number of trades is conditionally a Poisson distribution. Formula (4.8) defines the link function; the logarithm of the parameter in the Poisson distribution is assumed to be linearly linked to the independent variables. Formula (4.9) is similar to (4.4), except that the OLS is not a valid estimation method in the case of GLM, thus the log-likelihood function is used instead. Formulas (4.10) and (4.11) rewrite (4.7) and (4.8) in terms of estimates for GLM.

4.5. Lasso Poisson regression

The prediction for the lasso Poisson regression is the same as for the ridge Poisson regression, see (4.7). The difference is in the estimation method. The lasso Poisson regression coefficients estimates are the values that minimize:

$$LogLik_i + \lambda \sum_{j=1}^{92} \left| \beta_j^i \right|. \tag{4.12}$$

We compare the methods presented in terms of prediction power in the following section.

We refer to Hastie et al. (2008) for an extensive introduction to the methods presented, to Friedman et al. (2010) for recent research on GLM and to Zou and Hastie (2005) for a recent paper on shrinkage methods.

5. Empirical results

We estimated four models described in section 4 for 19 stocks. The presentation and interpretation of 76 models was challenging. Most attention was paid to the prediction ability of those models. Four models were compared with respect to prediction quality. Additionally, an analysis of the estimates was undertaken; the number of parameters combining all models was very high. The linear regression with lasso penalty function is the easiest to analyze estimates from. Thus, the estimates for this model, for all 19 stocks, are presented in Tables 6-9 in the Appendix.

5.1. Prediction accuracy

The quality of prediction is a very broad topic. Most importantly, the in-sample and out-of-sample predictions should be distinguished. It is of course important for the model to work well on the estimation sample (in-sample prediction), although the goal is usually the prediction on the basis of data for which the outcome is unknown (out-of-sample prediction).

Technically, the verification of the prediction abilities of the model focuses on the analysis of residuals, the difference between the prediction and the actual value. The most obvious goal for the model is to produce small residuals in terms of the absolute value. In this article, we only analyze this part of the prediction abilities. Examples of more detailed analyses, which go beyond the scope of this article, are sensitivity analysis or analysis of the distribution of residuals.

Table 4 Prediction quality

	Root Mean Squared Error					Mean Average Error				
	Random walk	Ridge LM	Ridge PR	Lasso LM	Lasso PR	Random walk	Ridge LM	Ridge PR	Lasso LM	Lasso PR
PKOBP	38.2	33.6	32.6	33.6	32.7	25.8	21.8	20.8	21.8	20.8
PZU	23.6	21.3	20.6	21.4	20.6	15.6	14.6	13.9	14.5	13.9
PEKAO	22.4	20.1	19.6	20.2	19.6	15.2	13.4	13.1	13.5	13.1
KGHM	34.1	31.6	30.2	31.5	30.2	22.6	21.7	19.7	21.5	19.7
PKNORLEN	22.0	19.8	19.3	19.8	19.3	14.5	12.2	11.9	12.2	11.9
PGE	26.8	25.1	24.4	25.1	24.5	17.0	16.8	16.4	16.8	16.5
LPP	14.1	14.1	13.5	14.1	13.5	8.8	6.9	6.7	6.9	6.7
BZWBK	21.7	20.0	18.9	20.0	18.9	11.8	10.6	10.4	10.6	10.4
PGNIG	19.6	18.3	17.4	18.3	17.4	11.7	11.3	10.8	11.3	10.8
MBANK	14.8	13.8	13.1	13.7	13.1	8.6	7.4	7.4	7.4	7.4
TAURONPE	14.0	12.9	12.4	13.0	12.3	8.4	7.7	7.5	7.6	7.5
ASSECOPOL	8.8	8.0	7.6	8.0	7.6	5.3	4.6	4.6	4.7	4.6
BOGDANKA	8.6	7.9	7.1	7.9	7.1	4.5	4.0	3.9	4.0	3.9
ALIOR	14.6	13.5	12.3	13.5	12.3	7.1	5.9	5.8	5.9	5.8
EUROCASH	16.5	15.5	15.1	15.5	15.1	9.6	9.3	9.1	9.3	9.1
SYNTHOS	11.2	10.0	9.4	10.0	9.4	6.1	5.5	5.4	5.5	5.4
LOTOS	13.7	12.7	12.2	12.7	12.2	6.8	5.8	5.7	5.8	5.7
JSW	17.1	15.5	14.7	15.5	14.7	9.9	8.4	8.2	8.4	8.2
KERNEL	12.4	11.7	10.5	11.7	10.5	6.8	7.8	7.0	7.7	7.0

Note: LM stands for linear model, PR for Poisson Regression. Shaded cells indicate the minimum value per measure and asset out.

Table 4 summarizes prediction power of the proposed models. Firstly, of the chosen models Poisson regression with the ridge penalty function seems to outperform others, although Poisson regression with the lasso penalty function gives very similar results. Secondly, the performance of the prediction is better for more liquid stocks. Comparing MAEs and MSEs in Table 4 to descriptive statistics, especially the mean and median, in Table 3, relative MAEs and relative MSEs are lower for more liquid stocks. Similarly, in the case of less liquid stocks, MAEs and MSEs are not much lower for the predictions of the models in comparison with the naïve strategy. The MAE for KERNEL is even lower using random walk as a predictor. Note that the performance was validated out-of-sample, thus even small differences are meaningful, although they might not be statistically significant. It is impossible for the model to perform worse than the naïve strategy in-sample.

5.2. Estimation results

The second goal of the analysis performed is an evaluation of the predictive power of each of the chosen variables. There are 4 models for each of 19 analyzed stocks with 82 parameters each; this results in 6232 parameters. We present estimates for the linear model with the lasso penalty function. It does not possess the best properties with respect to quality of forecast, although it is easiest to infer from. Estimates in linear model have very natural and clear interpretation: a value of the variable one unit higher has been observed with, on average, the independent variables changed by the value of the coefficient. For the sake of comparability and validity of shrinkage methods, we estimated models using standardized variables, with zero mean and unit variance. Lasso penalty function works as a variable selection, thus it even simplifies the interpretation. The zero coefficients mean that the respective predetermined variables are insignificant.

Table 5. Estimates for time variable.

	09:05-10:55	**11:00-14:45**	**14:50-15:45**	**15:50-16:40**	**16:45**
PKOBP	-3.65	-9.48	0.24	6.49	42.39
PZU	-2.79	-6.51	-1.66	6.31	28.25
PEKAO	-0.69	-4.4	0.49	4.17	3.7
KGHM	-3.8	-9.79	-5.22	2.11	147.7
PKNORLEN	-4.7	-6.26	-0.31	6.28	11.9
PGE	-2.49	-7	-0.85	7.41	14.18
LPP	-0.7	-0.71	0.13	0.49	2.05
BZWBK	-1.08	-1.71	-0.51	2.4	0.37
PGNIG	-0.2	-4	-1.74	1.57	48.86
MBANK	-2.31	-2.4	0.47	1.9	3.82
TAURONPE	-2.75	-3.44	-2.18	3.62	25.33
ASSECOPOL	-1.02	-2.28	-0.17	2	7.85
BOGDANKA	-1.08	-1.08	0.18	0.71	3.49
ALIOR	-1.27	-1.79	-0.49	1.66	8.91
EUROCASH	-2.3	-2.39	-0.41	2.29	8.58
SYNTHOS	-2.53	-2.51	-1.46	2.55	18.01
LOTOS	-2.26	-2.5	-0.35	2.01	12.28
JSW	-2.46	-3.48	-2.38	1.96	45.16
KERNEL	-3.52	-4.04	-1.27	4.42	14.76

Table 5 summarizes the time pattern in the model. The coefficients are interpreted as the difference between the number of trades during 5-minute intervals in the corresponding interval in comparison with the first interval, 9:00-9:05. It can be seen that for all stocks trade intensity is lower during 9:05-11:00 than during 9:00-9:05. In the interval 14:50-15:50 trade intensity is comparable to the first 5 minutes and increases between15:50 and 16:45. In the last 5 minutes it spikes for every stock. It is important to note that the coefficients during the periods 9:05-11:00 and 14:50-15:50, and partially in 14:50-15:50, are negative. It means that intercepts for those periods are lower than the ones for 9:00-9:05. This is surprising because the average number of trades is higher in the first interval. The interpretation is that more information is stored in other variables than in the intercept.

All estimates for this model are presented in the Appendix, here only summary and interpretations are presented. There is a strong autoregression observed in the data, all estimates for the number of the corresponding stock are positive and relatively very high. The cross-sectional autoregression is also visible and mostly positive, meaning that high intensity in one period is observed before high intensity in the next period for other stocks on average. This property is mostly seen for big and liquid companies, and for some less liquid ones like BOGDANKA or ASSECO we see a negative coefficient.

The plus variable (the number of trades with a higher price than the preceding one) has a positive coefficient for the corresponding stocks (or zero in 3 cases). It means that plus-trades precede a higher intensity in trade, although the coefficients are much smaller than in the case of the number variable. The cross-sectional dependence between the plus variable and the number of trades is not clear, some coefficient are negative, some positive and a large number equal zero.

The minus variable (the number of trades with a lower price than the preceding one) seems less influential than the others, and there is no clear pattern recognized. There are many zero-coefficient ones and the majority of non-zero ones are positive. This, however, might be caused by the correlation between minus, plus and number variables.

There is a very interesting relationship observed for the volume variable. All except one (which is zero) coefficient are negative for the volume in the model for the corresponding stock. The immediate interpretation might be that a higher volume in one period is observed with a lower intensity in the next period. However, the number and the volume variables are highly correlated, thus the volume variable works as an off-set. After a period with many trades, there is a period which also has many trades (positive autocorrelation), although if those trades are relatively large (many stocks traded per trade) this relationship is lower. The cross-sectional dependence between the volume variable and the number of trades is low; most of the coefficients are zero and there is no visible pattern for the rest.

6. Conclusions

The forecasting of the number of trades in a given period is an alternative to the ACD model. The results obtained show a significant forecasting ability of GLM. Shrinkage methods such as lasso and ridge penalty functions are valuable tools in the model selection problem. For most of the stocks analyzed, the Poisson regression with the ridge penalty function is shown to be the best model in terms of MSE and MAE for the out-of-sample forecast. The Poisson regression with lasso penalty function gave almost the same results for all the stocks analyzed. The forecasting ability of the models proposed seems to work better for more liquid assets, with a higher average number of trades.

The analysis of the daily pattern of trade intensity showed 6 periods during a trading day. During each period, trade intensity is statistically similar. For a 5-minute interval during different periods, the expected number of trades is different. Those periods are 09:00-9:05, 9:05-11:00, 11:00-14:50, 14:50-15:50, 15:50-16:45 and 16:45-50. The trade pattern seems marginally different for more liquid stocks than for less illiquid ones. Liquid assets show the expected U-shape in the daily pattern. Trade is relatively intense in the first period (09:00-9:05), in the second (9:05-11:00) it is lower and it is lowest in the third period (11:00-14:50). The intensity in the fourth (14:50-15:50) period is comparable to the intensity in the second one (9:05-11:00). The intensity in the fifth period (15:50-16:45) is higher than in preceding ones, relatively very high values (outliers) are observed more often in this period. Trade intensity is absolutely highest in the last 5-minute interval. However, for relatively illiquid assets, trade during 9:05-15:50 is similarly intense and becomes more intense only in the last hour of trade.

There is a significant positive serial correlation observed in the high frequency data. Most of the data shows positive cross-sectional autoregression. This means that in most cases high intensity in one period is observed before high intensity in the next period for other stocks. This property is mostly detected for big and liquid companies. For some less liquid firms like BOGDANKA or ASSECO, we see negative coefficients.

The cross-sectional dependence between the plus variable (the number of trades with a higher price than the preceding one) and the number of trades is not clear, some coefficients are negative, some positive and a large number equal zero. In the case of the minus variable (the number of trades with a lower price than the preceding one) no clear pattern can be recognized. There are many zero-coefficients and the majority of non-zero ones are positive. This, however, might be caused by correlations among regressors.

The volume variable (the value of trades for a corresponding stock in a given interval) shows an interesting feature. All coefficients for the volume except one (which is zero) are negative for the corresponding stock. A possible interpretation might be that a higher volume in one period is associated with a lower intensity in the next period. However, the number and the volume variables are highly correlated, thus the volume variable works as an off-set. After a period with many trades, there is a period which also has many trades (positive autocorrelation), although if the volume of transactions is high (many stocks traded per trade) this relationship becomes weaker. The cross-sectional dependence between the volume variable and the number of trades is low, most of the coefficients are zero and there is no visible pattern in the rest.

There are at least three potential directions for further studies. Firstly, the forecast abilities of other models are needed. Machine-learning techniques like neural network often possess better forecasting abilities than linear or GLM. Other approaches based on linear models are also possible, for instance, the time variable or input to the model can be defined in a different way.

Secondly, different regressors might be used in the regression. The four (except the time) variables chosen are relatively easy to interpret. However, they may not possess the best prediction power. Moreover, they are all strongly correlated, thus the results might not be stable. One of the possible regressors which might be used in the further analysis is the percentage of trades with a higher price. In addition, more lags may be taken into account. In this paper, the authors have used only variables obtained from the trades in one preceding interval.

Thirdly, the analysis may be repeated for longer and shorter intervals. In this paper, the authors show the analysis for the number of trades during 5-minute intervals. It is possible that some conclusions would be different for shorter or longer ones.

REFERENCES

BAUWENS, L.,GIOT., P., (2000). The logarithmic ACD model: An application to the bid–ask quote process of three NYSE stocks, “Annales D’economie Et De Statistique”, 60, pp. 117–149.

BAUWENS, L.,HAUTSCH, N., (2009). Modelling financial high frequency data using point processes, pp. 953–979, Berlin: Springer.

CARTEA, Á., JAIMUNGAL, S., (2013). Modelling Asset Prices for Algorithmic and High-Frequency Trading, “Applied Mathematical Finance”, Vol. 20, No. 6, pp. 512–547.

CARTEA, Á., MEYER-BRANDIS, T., (2010). How duration between trades of underlying securities affects option prices, “Review of Finance”, 14 (4), pp. 749–785.

de JONG, F., RINDI, B., (2009). The microstructure of financial markets (1st ed.), Cambridge: Cambridge University Press.

DIAMOND, D. W.,VERRECHIA, R. E., (1987). Constraints on short-selling and asset price adjustment to private information, “Journal of Financial Economics”, 18, pp. 277–311.

DUFOUR, A., ENGLE, R. F., (2000). Time and the price impact of a trade, “The Journal of Finance”, LV (6), pp. 2467–2498.

EASLEY, D., O’HARA, M., (1992). Time and the process of security price adjustment, “The Journal of Finance”, XLVII (2), pp. 577–605.

ENGLE, R. F., (2000). The econometrics of ultra-high-frequency data, ”Econometrica”, 68 (1), pp. 1–22.

ENGLE, R. F., RUSSELL, J. R., (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data,” Econometrica”, 66 (5), pp. 1127–1162.

FERNANDES, M., GRAMMIG, J., (2005). Nonparametric specification tests for conditional duration models, “Journal of Econometrics”, 127 (1), pp. 35–68.

FRIEDMAN, J., TIBSHIRANI, R., HASTIE, T., (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent, “Journal of Statistical Software”, Vol. 33, No. 1. http://www.jstatsoft.org/v33/i01.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., (2008).The Elements of Statistical Learning, 2nd edition, Springer, New York.

HUJER, R., VULETIC, S., KOKOT, S., (2002). The Markov switching ACD model, SSRN eLibrary, Retrieved from http://ssrn.com/abstract=332381.

MAHEU, J. M., MCCURDY, T. H., (2000). Volatility dynamics under duration-dependent mixing, "Journal of Empirical Finance", 7 (3–4), pp. 345–372.

MANGANELLI, S., (2005). Duration, volume and volatility impact of trades, "Journal of Financial Markets", 8 (4), pp. 377–399.

MEITZ, M., TERASVIRTA, T., (2006). Evaluating models of autoregressive conditional duration, "Journal of Business & Economic Statistics", 24, pp. 104–124.

RENAULT, E., VAN DER HEIJDEN, T.,WERKER, B. J. M., (2012). The dynamic mixed hitting-time model for multiple transaction prices and times, Working Paper, Retrieved fromhttp://dx.doi.org/10.2139/ssrn.2146220.

TIBSHIRANI, R., (1996). Regression Shrinkage and Selection via the Lasso, "Journal of the Royal Statistical Society, Series B", Vol. 58, No. 1, pp. 267–288.

ZHANG, M. Y., RUSSELL, J. R., TSAY, R. S., (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data, "Journal of Econometrics", 104 (1), pp. 179–207.

ZOU, H., HASTIE, T., (2005). Regularization and Variable Selection via the Elastic Net, "Journal of the Royal Statistical Society, Series B", Vol. 67, No. 2, pp. 301–320.

APPENDIX

Table 6. Estimates for number variable

	PKOBP Number	PZU Number	PEKAO Number	KGHM Number	PKNORLEN Number	PGE Number	LPP Number	BZWBK Number	PGNIG Number	MBANK Number	TAURONPE Number	ASSECOPOL Number	BOGDANKA Number	ALIOR Number	EUROCASH Number	SYNTHOS Number	LOTOS Number	JSW Number	KERNEL Number
PKOBP	**15.92**	**1.17**	**2.47**	0	0	*-0.1*	0	0	**2.55**	0	0	0	*-0.74*	*-0.37*	0	0	0	**0.27**	0
PZU	**1.87**	**11.1**	**0.34**	0	**0.41**	**0.6**	0	0	**0.43**	0	0	*-0.59*	0	0	**0.74**	0	0	**0.34**	*-0.57*
PEKAO	**1.48**	0	**9.01**	**0.06**	**0.28**	**0.26**	0	**0.3**	**0.01**	0	**0.09**	*-0.07*	*-0.06*	*-0.18*	0	0	0	0	0
KGHM	*-0.37*	**1.2**	0	**25.61**	0	0	**1.62**	0	**1.02**	*-0.72*	**0.08**	*-0.31*	0	0	0	*-0.09*	0	0	0
PKN.	**0.2**	**0.12**	0	**1.04**	**8.31**	0	*-0.38*	0	**0.64**	0	**0.7**	0	*-0.37*	0	**0.33**	**1.09**	*-0.03*	0	*-0.81*
PGE	**0.28**	0	**0.17**	**0.24**	0	**16**	**0.46**	**0.2**	**0.63**	0	**0.73**	*-0.1*	*-0.16*	**0.01**	0	0	0	*-0.18*	*-0.11*
LPP	0	0	0	0	0	**0.24**	**1.11**	**0.13**	0	0	0	0	0	**0.24**	0	0	0	0	**0.08**
BZWBK	**0.05**	**0.81**	**0.11**	**0.01**	**0.42**	**0.39**	**0.24**	**6.06**	**0.08**	0	0	0	0	**0.24**	0	0	**0.16**	**0.2**	*-0.06*
PGNIG	0	**0.04**	0	**1.46**	**0.07**	**0.31**	**0.5**	**0.19**	**9**	0	**0.33**	*-0.14*	0	**0.07**	0	0	0	0	0
MBANK	**0.55**	**0.05**	**0.2**	**0.24**	**0.03**	**0.25**	**0.43**	0	*-0.28*	**2.83**	**0.02**	0	0	**0.03**	0	**0.19**	0	**0.33**	0
TAU.	0	**0.22**	**0.12**	**0.93**	**0.31**	**1.53**	0	0	**0.42**	0	**6.37**	0	*-0.53*	0	*-0.29*	**0.67**	*-0.23*	*-0.28*	*-0.35*
ASSEC.	**0.27**	0	**0.4**	**0.1**	**0.15**	**0.12**	**0.16**	0	0	0	**0.07**	**3.75**	0	*-0.02*	**0.01**	0	0	**0.54**	**0.02**
BOGD.	0	**0.11**	*-0.08*	0	0	**0.05**	**0.15**	**0.15**	**0.08**	0	0	**0.08**	**2.32**	**0.06**	0	0	0	0	0
ALIOR	0	0	**0.1**	0	0	0	**0.09**	**0.07**	*-0.07*	0	0	0	0	**1.95**	0	**0.19**	0	0	0
EURO.	0	**0.07**	0	0	**0.32**	**0.27**	0	0	**0.01**	0	0	*-0.61*	0	0	**5.84**	*-0.22*	0	0	0
SYNT.	0	0	0	**0.04**	**0.4**	**0.2**	0	0	**0.25**	0	**0.45**	0	*-0.23*	0	*-0.11*	**3.99**	0	**0.05**	0
LOTOS	0	0	*-0.05*	0	0	**0.18**	*-0.01*	0	**0.1**	0	**0.07**	0	*-0.04*	0	0	0	*-0.58*	0	0
JSW	**0.07**	0	**0.05**	0	0	**0.13**	0	0	0	0	0	0	0	0	0	0	0	*-1.29*	0
KERNEL	*-0.14*	*-0.17*	**0.02**	**0.3**	*-0.17*	*-0.46*	*-0.08*	**0.06**	0	*-0.04*	**0.05**	**0.09**	**0.31**	*-0.22*	0	0	**0.1**	**0.29**	*-2.12*

Table 7. Estimates for plus variable

	PKOBP Number	PZU Number	PEKAO Number	KGHM Number	PKNORLEN Number	PGE Number	LPP Number	BZWBK Number	PGNIG Number	MBANK Number	TAURONPE Number	ASSECOPOL Number	BOGDANKA Number	ALIOR Number	EUROCASH Number	SYNTHOS Number	LOTOS Number	JSW Number	KERNEL Number
PKOBP	**0.51**	0	**0.21**	**0.11**	**0.63**	**0.62**	*-0.17*	**1.44**	**0.32**	**1.51**	*-0.35*	0	0	0	*-0.41*	*-0.39*	*-0.99*	*-0.07*	0
PZU	**0.34**	**0.11**	**0.06**	**0.43**	**0.28**	**0.07**	0	**1.14**	0	**0.05**	**0.64**	*-0.37*	0	0	0	**0.82**	0	**0.86**	0
PEKAO	**0.37**	**0.3**	**0.98**	**1.29**	0	**0.51**	0	0	0	0	0	0	0	0	*-0.2*	0	*-0.32*	**0.61**	0
KGHM	0	0	**0.43**	**2.22**	*-0.11*	*-0.8*	*-1.07*	0	**0.43**	0	**0.59**	*-1.16*	*-2.19*	0	**0.78**	**1.49**	**0.94**	**0.92**	**1.28**
PKN.	**0.24**	**0.23**	0	*-0.43*	0	**0.17**	**0.24**	**0.49**	0	0	**0.56**	0	**0.3**	**0.04**	**0.22**	**0.43**	0	**0.28**	0
PGE	0	**0.63**	**1**	0	0	**1.01**	**0.15**	0	**0.1**	*-0.41*	**0.69**	0	*-0.22*	0	0	0	*-0.71*	0	*-0.16*
LPP	*-0.03*	0	0	**0.04**	*-0.05*	0	**0.08**	0	*-0.13*	0	0	*-0.09*	**0.08**	0	**0.06**	**0.15**	**0.02**	**0.01**	**0.01**
BZWBK	0	0	**0.15**	0	*-0.27*	0	**0.12**	**0.22**	0	**0.22**	0	0	0	**0.31**	0	**0.26**	0	0	*-0.22*
PGNIG	0	0	**0.08**	0	0	0	0	0	**0.71**	0	0	*-0.38*	0	**0.01**	**0.26**	0	**0.06**	0	**0.09**
MBANK	0	**0.17**	0	*-0.23*	**0.04**	**0.09**	**0.07**	**0.16**	**0.09**	**0.25**	0	0	**0.17**	**0.33**	**0.22**	*-0.08*	**0.18**	0	0
TAU.	0	*-0.11*	*-0.17*	*-0.14*	*-0.26*	*-0.15*	*-0.09*	*-0.09*	*-0.03*	0	*-0.19*	**0.27**	**0.35**	**0.02**	**0.28**	0	0	**1.15**	**0.11**
ASSEC.	0	*-0.17*	*-0.21*	*-0.33*	0	**0.06**	*-0.49*	0	**0.03**	**0.05**	**0.16**	0	0	**0.1**	*-0.13*	**0.1**	*-0.13*	0	0
BOGD.	0	0	0	0	0	0	0	0	*-0.07*	*-0.01*	0	0	**0.65**	**0.07**	**0.12**	**0.17**	0	**0.17**	**0.03**
ALIOR	0	0	0	**0.04**	*-0.12*	*-0.33*	**0.44**	0	*-0.21*	**0.37**	0	0	**0.52**	**1.16**	**0.28**	**0.1**	**0.03**	*-0.19*	**0.06**
EURO.	0	0	**0.16**	**0.05**	0	0	0	0	0	**0.32**	*-0.29*	0	**0.19**	*-0.11*	0	0	0	**0.53**	**0.29**
SYNT.	**0.02**	0	0	0	0	0	0	0	0	0	0	**0.04**	0	0	0	**0.06**	**0.15**	0	0
LOTOS	0	0	0	0	0	0	0	0	0	*-0.25*	*-0.18*	0	**0.14**	0	**0.19**	**0.28**	**0.68**	*-0.14*	**0.08**
JSW	**0.09**	0	*-0.15*	0	0	0	**0.12**	**0.29**	**0.07**	0	0	**0.11**	*-0.18*	**0.07**	0	**0.62**	0	**1.2**	*-0.13*
KERNEL	*-0.47*	**0.77**	*-0.63*	0	**0.03**	*-0.61*	**0.51**	*-0.03*	**0.39**	0	**0.65**	**0.54**	**0.65**	**0.64**	**0.07**	**1.8**	**0.52**	*-0.57*	**1.31**

Table 8. Estimates for minus variable

	PKOBP Number	PZU Number	PEKAO Number	KGHM Number	PKNORLEN Number	PGE Number	LPP Number	BZWBK Number	PGNIG Number	MBANK Number	TAURONPE Number	ASSECOPOL Number	BOGDANKA Number	ALIOR Number	EUROCASH Number	SYNTHOS Number	LOTOS Number	JSW Number	KERNEL Number
PKOBP	0	**0.64**	**0.59**	**0.03**	**0.17**	0	*-0.73*	0	0	0	0	**0.81**	*-0.33*	0	0	0	**0.78**	**0.71**	0
PZU	0	**0.5**	**0.18**	**0.45**	**0.59**	0	*-0.9*	0	*-0.34*	0	0	0	0	*-0.28*	0	*-0.06*	**0.74**	**0.31**	**0.5**
PEKAO	0	**0.4**	0	**0.04**	0	0	0	**0.03**	**0.33**	0	**0.67**	0	*-0.74*	0	0	0	0	0	*-0.09*
KGHM	0	**0.16**	0	**3.78**	*-0.07*	0	*-0.69*	**0.17**	*-0.06*	**0.38**	0	*-1.15*	**0.49**	0	**0.21**	**1.72**	*-0.21*	**0.38**	**0.74**
PKN.	**0.12**	**0.77**	**0.58**	0	0	**0.46**	*-0.97*	**0.04**	**0.45**	**0.1**	**0.23**	*-1.26*	*-0.32*	0	*-0.67*	**0.49**	**0.46**	**0.86**	**1.03**
PGE	0	0	0	**0.6**	0	*-3.72*	0	0	0	0	**0.21**	0	*-0.05*	**0.19**	0	0	**0.16**	0	**0.44**
LPP	0	*-0.11*	0	0	0	*-0.07*	**0.39**	*-0.11*	*-0.01*	**0.09**	0	0	**0.06**	**0.11**	0	*-0.08*	**0.02**	0	**0.04**
BZWBK	0	0	*-0.07*	**0.03**	0	0	0	**0.05**	0	*-0.35*	*-0.09*	0	*-0.18*	0	**0.29**	**0.01**	0	**0.13**	*-0.08*
PGNIG	**0.23**	**0.05**	0	0	**0.41**	0	0	0	**1.08**	0	0	0	0	0	0	**0.15**	**0.41**	**0.3**	**1.14**
MBANK	*-0.02*	0	*-0.32*	0	*-0.02*	0	*-0.66*	**0.18**	0	**1.21**	**0.48**	*-0.14*	**0.09**	**0.17**	*-0.04*	0	**0.39**	*-0.13*	*-0.07*
TAU.	**0.12**	0	0	*-0.05*	**0.26**	*-0.87*	*-0.45*	0	0	0	**0.11**	**0.17**	0	**0.13**	*-0.06*	**0.08**	**0.21**	0	**0.21**
ASSEC.	0	0	*-0.29*	0	*-0.02*	*-0.01*	*-0.34*	*-0.03*	*-0.09*	**0.35**	**0.21**	*-0.27*	**0.12**	**0.21**	**0.02**	0	0	*-0.47*	**0.04**
BOGD.	0	**0.04**	0	**0.16**	*-0.01*	0	*-0.14*	0	0	0	**0.13**	0	**0.68**	0	**0.11**	0	**0.23**	0	0
ALIOR	*-0.07*	*-0.03*	*-0.32*	**0.05**	*-0.02*	*-0.11*	0	0	0	0	0	**0.05**	0	**1.39**	0	0	**0.27**	0	0
EURO.	**0.46**	0	0	**0.79**	**0.08**	**0.06**	*-0.11*	*-0.15*	0	*-0.18*	0	0	**0.09**	**0.13**	**0.86**	*-0.28*	**0.25**	**0.15**	**0.22**
SYNT.	0	**0.04**	0	0	0	*-0.1*	0	**0.17**	0	0	0	*-0.26*	0	**0.24**	0	0	0	**0.16**	**0.22**
LOTOS	0	**0.06**	0	**0.25**	0	*-0.1*	0	0	0	0	0	0	**0.08**	**0.28**	**0.1**	**0.09**	**1.38**	0	**0.18**
JSW	0	**0.36**	**0.07**	**0.65**	0	0	*-0.34*	0	0	*-0.03*	**0.08**	0	0	**0.48**	0	**0.36**	**0.03**	**0.18**	0
KERNEL	**0.1**	*-0.35*	*-0.42*	*-0.37*	**0.39**	*-1.16*	0	**0.72**	**0.66**	**0.45**	**0.64**	*-0.01*	*-0.25*	**0.31**	0	**0.52**	**0.83**	0	**1.06**

Table 9. Estimates for volume variable

	PKOBP Volume	PZU Volume	PEKAO Volume	KGHM Volume	PKNORLEN Volume	PGE Volume	LPP Volume	BZWBK Volume	PGNIG Volume	MBANK Volume	TAURONPE Volume	ASSECOPOL Volume	BOGDANKA Volume	ALIOR Volume	EUROCASH Volume	SYNTHOS Volume	LOTOS Volume	JSW Volume	KERNEL Volume
PKOBP	-1.21	0	-0.64	**0.24**	0	0	0	-0.24	-0.68	0	-0.11	-0.04	0	-0.11	0	-0.07	0	0	-0.12
PZU	**0.09**	-2.29	**0.05**	**0.44**	0	0	-0.3	-0.08	0	**0.09**	0	0	0	0	0	**0.03**	0	0	0
PEKAO	0	**0.02**	-1.88	0	**0.01**	0	0	0	**0.06**	0	0	0	0	-0.06	**0.06**	0	0	0	0
KGHM	**0.06**	-0.53	0	-5.22	-0.04	**0.17**	-0.97	**0.14**	**0.19**	0	0	0	-0.02	**0.13**	-0.01	-0.15	**0.1**	**0.27**	-0.18
PKN.	**0.36**	-0.14	0	-0.11	-1.58	**0.09**	0	0	-0.12	0	-0.16	0	-0.09	**0.09**	0	-0.2	0	0	0
PGE	0	0	0	**0.45**	0	-2.98	0	**0.08**	**0.13**	0	0	0	0	0	**0.04**	-0.1	**0.03**	0	-0.16
LPP	0	0	**0.01**	0	0	**0.02**	0	**0.02**	**0.06**	0	0	-0.03	0	0	**0.01**	0	-0.01	0	**0.04**
BZWBK	**0.34**	0	0	**0.1**	**0.05**	**0.09**	0	-1.41	**0.08**	**0.15**	**0.01**	-0.02	0	0	**0.04**	-0.07	-0.02	0	0
PGNIG	0	0	0	0	0	0	0	0	-1.27	0	0	0	0	0	0	0	**0.01**	**0.17**	0
MBANK	-0.13	**0.12**	-0.02	**0.01**	**0.16**	0	0	**0.02**	**0.16**	-0.82	**0.01**	-0.05	0	0	-0.07	0	-0.05	0	0
TAU.	0	0	0	-0.04	**0.01**	0	0	**0.04**	**0.02**	0	-1.13	-0.08	**0.1**	**0.01**	**0.12**	-0.18	0	0	**0.03**
ASSEC.	0	**0.06**	-0.02	**0.11**	-0.04	0	0	0	**0.02**	-0.05	-0.02	-0.69	-0.02	-0.02	0	0	0	-0.06	0
BOGD.	0	0	0	0	0	0	0	0	**0.04**	**0.02**	**0.04**	0	-0.44	0	0	-0.04	-0.01	0	**0.01**
ALIOR	0	0	0	0	0	**0.38**	**0.12**	0	**0.18**	-0.06	0	-0.01	-0.05	-0.26	-0.05	0	-0.02	**0.05**	**0.03**
EURO.	0	0	-0.13	**0.22**	0	**0.08**	0	**0.07**	0	-0.06	**0.01**	0	0	0	-0.92	0	0	0	0
SYNT.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.63	0	0	0
LOTOS	0	0	-0.05	0	0	**0.18**	-0.01	0	**0.1**	0	**0.07**	0	-0.04	0	0	0	-0.58	0	0
JSW	**0.07**	0	**0.05**	0	0	**0.13**	0	0	0	0	0	0	0	0	0	0	0	-1.29	0
KERNEL	-0.14	-0.17	**0.02**	**0.3**	-0.17	-0.46	-0.08	**0.06**	0	-0.04	**0.05**	**0.09**	**0.31**	-0.22	0	0	**0.1**	**0.29**	-2.12