Labour market flows from the LFS, what can we learn from registers?

Situation

In order to better understand labour market dynamics, we need flow statistics in addition to traditional stock statistics. The panel design of most LFS makes flow statistics possible, at least in theory. By linking data at the micro level, we can compare labour market information about the same person from several points in time. From these premises, users may wonder why Statistics Norway does not publish flow statistics based on the LFS on a regular basis. In addition to the resource situation, we would like to point out some quality problems.

Usually, stock figures are made from the complete sample at a point in time. At the next point in time, new stock figures are made from another complete sample. Change figures are usually the difference between these two stock figures. The two samples can be more or less overlapping, depending on several factors. You do not have the time to wait and see how well the data overlap, before you publish the first stock figure!

Linking data from two partially overlapping samples result in a panel data set, of smaller sample size. Rotation, attrition and demographic changes result in more or less comparability between the panel and the full sample. Nonresponse and measurement errors introduce additional problems. It is no wonder then, that inconsistencies can arise between stock- and flow estimates. We have not resolved these inconsistencies in a satisfactory manner.

In addition, we have found quality analysis of flow statistics to be more complex than the case for stock statistics. For instance, from survey-data at one point in time you can estimate non-response rate and proxy-response rate. Panel data from two points in time containing responses, proxy-responses and non-responses, result in nine different combinations for each person. A straightforward quality indicator for that situation is not obvious.

The idea

Statistics Norway uses register-based data extensively, in all steps in the LFS production process: sampling, interviewing, classification, estimation, quality analysis. So it is only natural that we again turn to registers for help. The reason we propose to start out with register-based data is to simplify the picture. As we have pointed out, flows statistics face more complex challenges than stock statistics. If we first look at simplified data we can perhaps better discern between the many challenges, and in time make better quality assessment.

The method assumes that the register is an error-free full-count. In practice, registers are not perfect. However, registers don't face typical survey problems such as design effect, sample error, nonresponse error, measurement error, mode effect etc.

Register data can be linked at the micro level by using universal codes for individual identification. The result from linking overlapping registers from two points in time can be expressed as 3 data sets, as was the case with survey panel. We use presumably complete and

correct data containing the LFS target population, namely 15–74 years old residents. Any differences between the overlapping and non-overlapping data should represent real population changes. These changes include additions: immigration, becoming 15 years old; and subtractions: emigration, deaths, becoming 75 years old.

So the idea is to assess quality by comparing the differences between stock and flow figures, between the survey and register data. Specifically, we first determine the differences that are caused by actual population changes. Resulting divergence must be caused by sampling, nonresponse, measurement errors, etc.

Some labour market flows are more interesting than others, for instance how many unemployed people get a job. However, we chose workforce status as the main variable of interest for this preliminary study. This was because we wanted a similar variable available from both register- and survey data, and because it illustrated an important point about population change.

Example 1: REGISTER

Table 1 shows register figures for workforce stocks and flows. From that we can calculate two figures for workforce change: "Stock change" is the difference in absolute figures from 2010 to quarter to 2011, while "Net flow" is the difference between inflow and outflow for the same period. This net flow is based on the panel only, i.e. people that are in the target population both in 2010 and 2011.

The stock data show an **increase** of 28.000 more people outside the workforce, the flow data show a **decrease** of 11.000. For workforce figures, stock data show over three times the increase of panel data. Studying relatively long time series from registers (Diagram 1), we have found these inconsistencies to be relatively small and rather stable. We conclude that this is a structural phenomenon caused by the difference in workforce rate among people moving in and out of the target population. This means that the inconsistency is a result of actual population change, and not a sign of low quality of the input data.

Table 1: Workforce stocks and flows 4th Quarter 2010 - 2011. Register data. Residents 15-74 years old.

Stocks	2010	2011		Stock change	Net flow
Total	3 643 182	3 705 934		62 752	0
Workforce	2 575 659	2 610 127		34 468	10 598
Inactive	1 067 523	1 095 807		28 284	-10 598

Flows		2010-2011	
		Total	Panel
Total	Total	3 770 261	3 578 855
	Workforce	2 610 127	2 567 939
	Inactive	1 095 807	1 010 916
	Not panel	64 327	
Workforce	Total	2 575 659	2 557 341
	Workforce	2 375 765	2 375 765
	Inactive	181 576	181 576
	Not panel	18 318	
Inactive	Total	1 067 523	1 021 514
	Workforce	192 174	192 174
	Inactive	829 340	829 340
	Not panel	46 009	
Not panel	Total	127 079	
	Workforce	42 188	
	Inactive	84 891	

Diagram 1: Workforce change. 4th Quarter – 4th Quarter. Register data 2000–2012.
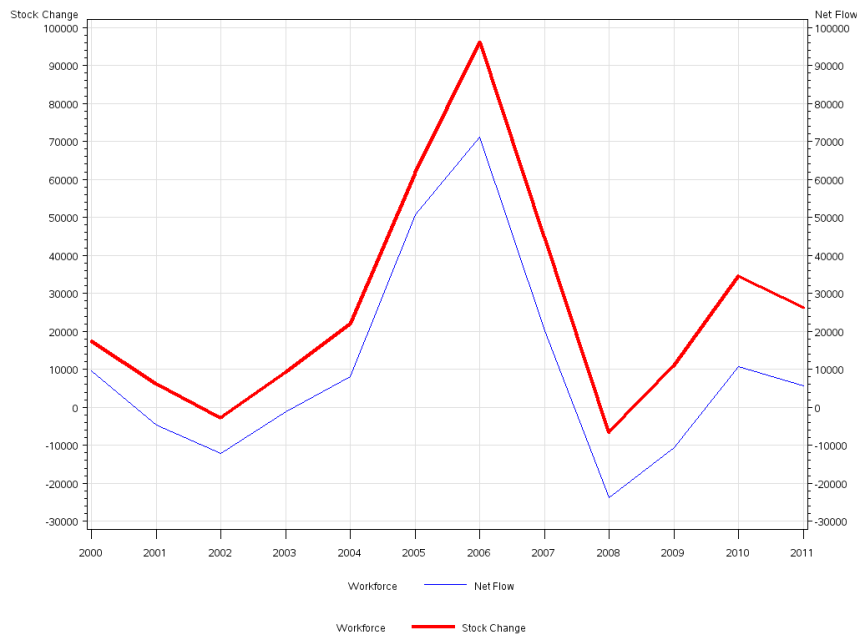
Example 2: SURVEY

Diagram 2 present time series of workforce change figures based on LFS survey panel data. "Stock change" is the difference in absolute figures from one quarter to 4 quarters later, while "Net flow" is the difference between inflow and outflow for the same period. Over a period of 4 quarters, the rotation plan result in a panel size maximum 50% of the full sample size. Diagram 3 shows the result of using regular weights for stock change and mean weights for flow change. In both diagrams we have adjusted the scales instead of reweighting at the micro level.

The survey sample figures show a more irregular and diverging picture, than comparable time series from register data. The results seem to indicate different trends at times. Possible reasons for this include, in addition to population changes, sampling- and nonsampling errors, rotation and attrition. The weighted series show more promising consistency, but are not very accurate. For instance unemployment-to-employment flow would require a better estimation method.

The point of using longitudinal data from registers as well is that we can begin to entangle the different factors behind the irregular and diverging trends. For instance, we have the opportunity to link register-based workforce status to the full sample, including the nonrespondents. From that kind of linked data we can discern between rotation effect and nonresponse effect.

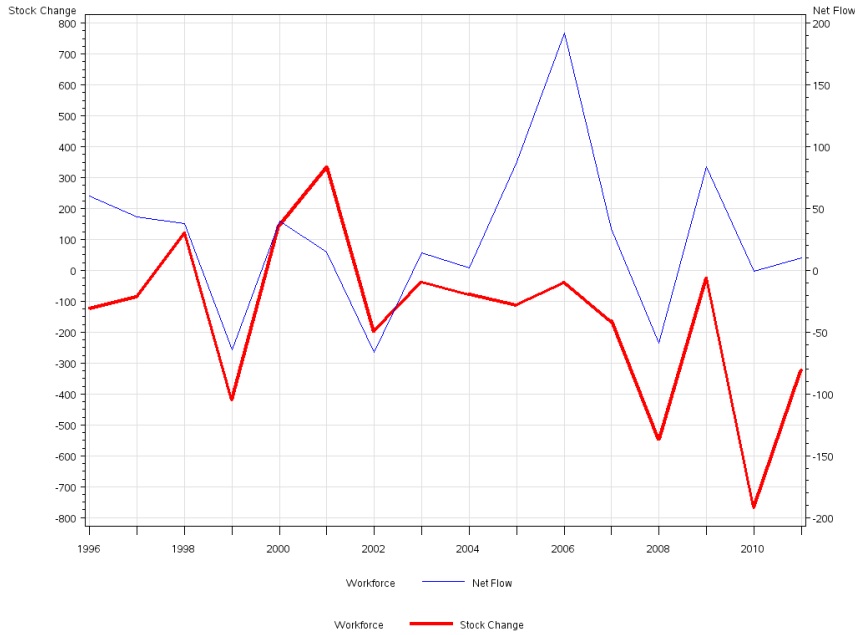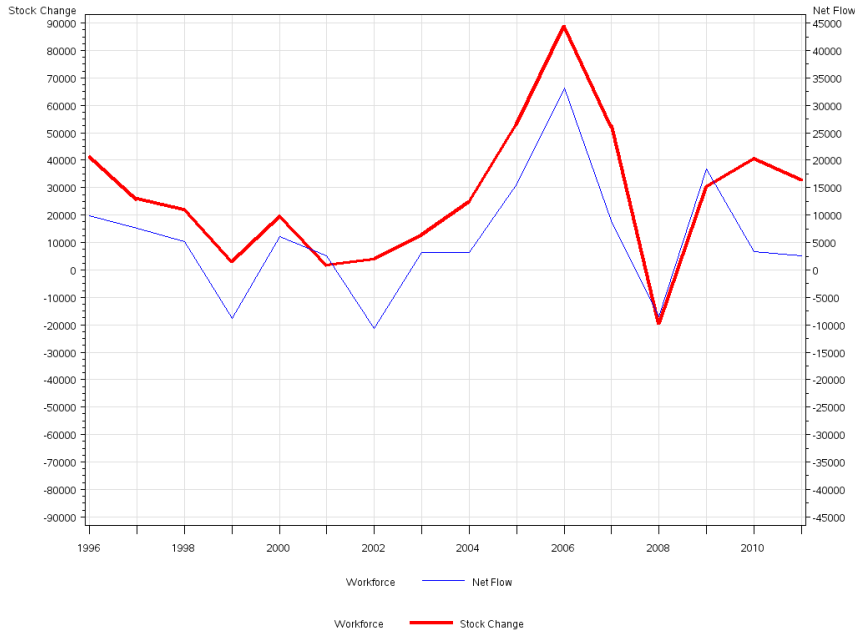Diagram 2: Workforce change. 4th Quarter – 4th Quarter. LFS 1996–2012. Sample figures.

Diagram 3: Workforce change. 4th Quarter – 4th Quarter. LFS 1996–2012. Estimates.

Conclusion

Can this help us, and other countries, to present better output?

By using registers, we have identified that actual population changes causes some slight divergence between stock and flow figures for workforce trends. Different workforce rate between people moving in and out of the population causes different change figures between stock- and flow-data. This can be documented by making some tables that include "non-panel" flows. If you have register-based longitudinal data with at least some kind of labour market information, this could be used to complement the LFS output.

4

In cases with more rapid demographic changes, this is even more relevant. For instance, due to the increasing influx of immigrant workers from new EU-countries, the population changes pose more challenges to survey-based statistics in Norway than before.

We have observed that LFS panel data show larger irregularities when comparing stock and flow figures for workforce trends. We plan to link register data containing labour market information to the sample data, in order to assess the impact of measurement errors and nonresponse errors to flow estimates. At this stage, we are very interested in experiences from countries that uses register data for similar purposes.

The regular stock statistics from the Norwegian Labour Force Survey are based on an estimation procedure involving post-stratification weights. Through analysis of relatively long time series, we have established that although the weighting adjusts the workforce rate level, it doesn't affect the change figures noticeably. With this in mind, and for the sake of simplification, we have disregarded the regular estimation procedure in this preliminary research. The next step is to develop a workable production system for estimation and dissemination of flow statistics.

As we have pointed out, official statistics for labour market changes could reveal seemingly inconsistent stock- and flow- figures. We should provide an explanation alongside the statistics if this inconsistency becomes apparent to the users. Other ideas include "hiding" the inconsistency, for instance by publishing only relative flow figures or adjusting the figures by some kind of weighting or calibration.

However, we believe that the impact of measurement errors and nonresponse errors are a much more pressing problem regarding flow estimates. We also believe that register data will prove useful in the further development, both for quality analysis and estimation method.

ANNEX I: TABLES

Table A: Workforce stocks. Register-based data 4th Quarter 2000–2012 [1]

	Total	Workforce	Inactive
'2000	3 199 439	2 307 083	892 356
'2001	3 211 032	2 326 929	884 103
'2002	3 234 083	2 334 681	899 402
'2003	3 256 107	2 323 697	932 410
'2004	3 282 342	2 334 885	947 457
'2005	3 371 778	2 360 932	1 010 846
'2006	3 413 695	2 423 224	990 471
'2007	3 469 345	2 517 275	952 070
'2008	3 528 773	2 562 265	966 508
'2009	3 582 114	2 550 765	1 031 349
'2010	3 643 182	2 575 659	1 067 523
'2011	3 705 934	2 610 127	1 095 807
'2012	3 768 005	2 636 483	1 131 522

Table B: Workforce flows. Register-based data 4th Quarter 2000–2012

	Total				Workforce			
	Total	Workforce	Inactive	Not panel	Total	Workforce	Inactive	Not panel
2000-2001	3 275 511	2 333 659	877 373	64 479	2 316 241	2 141 312	159 804	15 125
2001-2002	3 295 077	2 339 900	894 183	60 994	2 333 659	2 150 944	168 109	14 606
2002-2003	3 317 219	2 337 232	918 875	61 112	2 339 900	2 142 536	182 565	14 799
2003-2004	3 341 976	2 346 478	935 864	59 634	2 337 232	2 147 883	175 264	14 085
2004-2005	3 426 876	2 368 594	1 003 184	55 098	2 346 478	2 165 346	167 680	13 452
2005-2006	3 471 702	2 430 267	983 428	58 007	2 368 594	2 199 951	155 231	13 412
2006-2007	3 528 645	2 526 416	942 929	59 300	2 430 267	2 256 433	159 456	14 378
2007-2008	3 584 826	2 571 237	957 536	56 053	2 526 416	2 337 428	173 718	15 270
2008-2009	3 643 155	2 564 759	1 017 355	61 041	2 571 237	2 365 627	188 841	16 769
2009-2010	3 706 022	2 575 659	1 067 523	62 840	2 564 759	2 354 888	192 256	17 615
2010-2011	3 770 261	2 610 127	1 095 807	64 327	2 575 659	2 375 765	181 576	18 318
2011-2012	3 833 606	2 636 483	1 131 522	65 601	2 610 127	2 403 873	187 799	18 455

	Inactive				Not panel		
	Total	Workforce	Inactive	Not panel	Total	Workforce	Inactive
2000-2001	883 198	169 482	664 362	49 354	76 072	22 865	53 207
2001-2002	877 373	163 428	667 557	46 388	84 045	25 528	58 517
2002-2003	894 183	170 520	677 350	46 313	83 136	24 176	58 960
2003-2004	918 875	174 000	699 326	45 549	85 869	24 595	61 274
2004-2005	935 864	175 821	718 397	41 646	144 534	27 427	117 107
2005-2006	1 003 184	205 881	752 708	44 595	99 924	24 435	75 489
2006-2007	983 428	230 686	707 820	44 922	114 950	39 297	75 653
2007-2008	942 929	193 915	708 231	40 783	115 481	39 894	75 587
2008-2009	957 536	164 955	748 309	44 272	114 382	34 177	80 205
2009-2010	1 017 355	181 733	790 397	45 225	123 908	39 038	84 870
2010-2011	1 067 523	192 174	829 340	46 009	127 079	42 188	84 891
2011-2012	1 095 807	193 439	855 222	47 146	127 672	39 171	88 501

---

[1] Break in time series: 2005–2006 target population changed from 16–74 to 15–74 years old residents.