Finnish experiences: Nowcasting Finnish Turnover Indexes Using Firm-Level Data

Henri Luomaranta, Statistics Finland
Sofia, ESS Big Data WP1 Dissemination seminar
Content

- Motivations.
- Methodological issues, what we do.
- Data and empirical application.
- Results.
- Conclusions and next steps.
Motivations

• Nowcasting is usually done by using monthly indicators to forecast quarterly ones.
• However, statistical offices have highly disaggregated datasets available that resemble Big Data and that are very timely
  • Continuously accumulating
  • Usage can be difficult as it accumulates in different registers
  • Size can be big
• A starting point for constructing early estimates: can we extract meaningful information from the firm level data that is already collected to construct aggregate indexes.
• This can be added value w.r.t. academia, central banks, etc. that do nowcasting, since such microdata is not publicly available.
Motivations

- With disaggregated information sets comes econometric difficulties. How to extract relevant information?
- Fortunately, the literature has proposed a number of large dimensional techniques which we can rely on.
- One modeling option: reduce the dimension in the data by factor analysis. They extract the common component in the data.
- Second modeling option: use models that can deal with many dimensions. The shrinkage models (in our work, we use Lasso, Rigde and Elastic Nets regressions). They are able to include some idiosyncratic firm level information.
Factor models

• In macroeconomics, they became very popular after Stock and Watson (2002). They show that principal components provide a consistent estimates of the (dynamic) factors underlying a panel of variables. The model can be expressed as

\[ X_t = \Lambda F_t + e_t \]

• The estimates of the factors correspond to the normalized eigenvectors associated to the largest eigenvalues of \( XX' \).

• The number of factors is usually selected by information criteria (see, e.g., Bai and Ng, 2002).
Shrinkage models

• Factor models have been shown to be really effective in terms of forecasting and nowcasting performance.
• However, they are limited at attempting to estimate the common component underlying our panel.
• In a granular economy such as the Finnish one, we might be interested in keeping some firm-level information (the idiosyncratic components).
• Shrinkage methods such as Lasso, can be extremely interesting in this setting. We also test Ridge, Elastic-Nets (these are well known in e.g. machine learning and in nowcasting literatures)
• The idea is to introduce some bias in the OLS estimation to make it more efficient when the problem is high dimensional.
What we do

• In "Nowcasting Finnish Turnover Indexes using Firm-Level Data (2016)" with Paolo Fornaro (the Research Institute of the Finnish Economy) and Lauri Saarinen.

• We compute faster estimates of Finnish turnover indexes for 4 main industries.

• Methodology to compute 118 turnover indexes faster and more accurately by imputing part of the data sources, using factor analysis.

• To nowcast these indicators we use firm-level sales that accumulates continuously from the web based survey.

• We explored other sources, like wages and salaries and taxes paid data from the tax authority but find that the survey data is the most useful.
Data and empirical application

- We want to make rapid estimates of the year-on-year growth rates of various turnover indexes.
- We make four estimates: after 5, 10, 20 and 26 days after the end of the month.
- And one in 28 days after (by forecasting part of the source data)
- In pseudo real time: we simulate the realistic data accumulation faced by Statistics Finland from January 2012 up to July 2016.
- First estimates t+5 based on ca. 60 firms, t+26 based on ca. 800 firms.
What we obtain: results relative to ARIMA benchmark

- Shrinkage methods and factor models consistently beat ARIMA, even at early estimation rounds.
- Shrinkage models seem to do usually better than factor ones.
- The improvements in nowcasting performance happen between 5 and 10 from 10 to 20 days. The information accumulated after 20 days does not seem too beneficial.
Results relative to the current first estimate (t+45) in the four main industries

• For services, we beat the current estimate (t+45) already at t+5. We would be able to obtain more accurate estimate 40 days before the current date.
• For construction, we beat the current estimates from 20 days estimates onward.
• We cannot beat the Statistics Finland estimates for manufacturing and trade.
• However our nowcasts for the trade sector are competitive.
Extension – nowcasting part of the source data

- Statistics Finland produces the turnover indexes based on two data sources: the sales inquiry and VAT data.
- The sales inquiry is fairly complete at t+28, while the VAT data is slow to obtain and is available t+52.
- Current first estimates at t+45 are subject to large revisions.
- Our strategy is to nowcast the VAT part by using the factors extracted from the turnover inquiry as external predictors in an ARIMA model.
- The nowcasted VAT component is then included in t+28 estimate

\[
\hat{y}_{t,K} = \hat{w}_{t,I,K} \left( \frac{I_{t,K}}{I_{t-12,K}} - 1 \right) + \hat{w}_{t,VAT,K} \left( \frac{VAT_{t,K}}{VAT_{t-12,K}} - 1 \right)
\]
## Results from the main industries, but we did it for 120 published series (t+28), mean absolute error

<table>
<thead>
<tr>
<th></th>
<th>Manufacturing</th>
<th>Services</th>
<th>Trade</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE of estimate (1 factor)</td>
<td>0.592</td>
<td>0.391</td>
<td>0.535</td>
<td>1.100</td>
</tr>
<tr>
<td>MAE of estimate (2 factors)</td>
<td>0.423</td>
<td>0.381</td>
<td>0.543</td>
<td>2.295</td>
</tr>
<tr>
<td>MAE of estimate no factors</td>
<td>0.844</td>
<td>0.917</td>
<td>0.751</td>
<td>1.914</td>
</tr>
<tr>
<td>MAE of $t</td>
<td>45$ (current method)</td>
<td>1.084</td>
<td>1.730</td>
<td>0.895</td>
</tr>
</tbody>
</table>
Results from nowcasting VAT component

- We always beat the current estimate of Statistics Finland, even by a large margin.
- In all cases, adding factors is beneficial for the estimate.
- Our approach provides better results for 118 series of the 120 considered and represents a significant methodological improvement.
- Reduces the dependency on the administrative sources.
Conclusions

• We manage to create rapid estimates of turnover indexes using micro-level sources available to NSIs
• Using large dimensional models, we are able to extract the relevant information.
• We are able to apply this methodology to a wide range of indicators, in a computationally tractable way.
• Statistics Finland is foreseen to start regular production of faster turnover indexes based on this work, so it is indeed a quick win for the NSI.
• Even though some of the issues discussed are country specific, the methodology should be generalizable to other settings.
Next steps

• First, we are going to apply these methodologies to different series.
• Most importantly the Trend indicator of output, which is the monthly economic output.
• Nowcasts and forecasts of monthly indicators to form real time estimates of the quarterly GDP, initial tests suggests that this will yield very good results.
• Other data sources are available: like the CCI and the survey responses within.
• We are also exploring scanner data from the most important Finnish retail chain to nowcast retail trade index.