10 September 2018

Methodological report on creating early estimates of turnover indices

The data for compiling the turnover index for the most recent months is still incomplete. To minimize the error caused by this, new estimating models have been introduced. Later in the report they are referred to as nowcasting methods.

1. Methods

Three different main methods have primarily been used in nowcasting:

1) **Arima modelling**, the time series were forecasted with Arima-models which were augmented with common factors estimated from the turnover inquiry or wages and salaries micro-level data.

2) **Static models**, which use the factors estimated from the sales inquiry or wages and salaries micro-level data. These models do not model explicitly the time series structure of the indicators (we use simple OLS in this category of models).

3) **Regularization models**, where a linear model is supplemented with a regularization algorithm to solve the problem of overfitting, or curse of dimensionality.

Factors are estimate with principal component analysis (PCA). They are estimated from different subsets of data. Both the sales inquiry and wages and salaries data are classified by industry, and the corresponding factors are obtained for each of these subsets of data. In PCA, the variance-covariance matrices of firm-level year-on-year changes are transformed to Eigen matrices and the eigenvalues related to each eigenvector describes the significance of the vector. 35 vectors that explain the most variation in the data sets are selected at this stage. These factors are subsequently used in the models as explanatory variables.

The most significant vectors are selected as factors explaining the time series. Dropping of factors has not been found to be a problem, because most of the changes are explained by just a few significant factors. (see e.g. Stock and Watson 2002).

Source data is validated, so that factors are estimated from data that does not contain clear errors or outliers, and firm level series do not have missing observations and are long lasting. New enterprises are excluded from the source data. This selection is based on quality and practical considerations, for example there is no need to carry out imputation on missing observations.

1.1. Predicting the VAT group as a separate series

It is possible that the sales inquiry enterprises (in practice, large enterprises) behave in a different way than (small) enterprises included only with the VAT data of the value added tax data. Therefore, predicting the sales growth of enterprises not included in the sales inquiry as a separate series seems to give better results than predicting the entire time series in many cases. For this reason, enterprises are divided into two groups in some of the models, those providing sales data (I_t) and those included only in the Tax Administration's data (VAT_t). The sales inquiry data are available at the time of the prediction so the sales inquiry series is formed from directly available annual changes. The VAT series is in turn forecasted for one month ahead. Enterprises on which sales inquiry data can be found for some month are included in the sales inquiry enterprises in the whole prediction period even though sales data for some individual month were missing. Missing sales data are replaced with VAT data when available.

1.2. ARIMA models with regressors

Basic ARIMA model factors as additional explanatory variables (max 5). The growth of the VAT group is estimated with the ARIMA model to which factors are added as explanatory variables.

$$\hat{Y}_t - \phi_1 Y_{t-1} = \mu - \theta_1 e_{t-1} + \beta(X_t - \phi_1 X_{t-1}) \tag{1}$$

Y here stands for predicted VAT growth in other words.

$$\left(\frac{\widehat{VAT_{t,K}}}{VAT_{t-12,K}} - 1\right) \tag{2}$$

When the growth of the VAT group is known, weights are formed for both groups based on their turnover shares.

$$\hat{y}_{t,K} = \hat{w}_{t,I,K}\left(\frac{I_{t,K}}{I_{t-12,K}} - 1\right) + \hat{w}_{t,VAT,K}\left(\frac{\widehat{VAT_{t,K}}}{VAT_{t-12,K}} - 1\right) \tag{3}$$

After this, all the components of the formula above are known, because the turnover of respondents to the sales inquiry is known. This is assumed to be the best functioning model when the behaviour of the time series can be modelled well.

For this there are various variations where a different number of principal components (or factors) are taken as explanatory variables, or the whole series is estimated without division into VAT components and the inquired part. Thus there is a total of 42 model variations based on the ARIMA model.

1.3. Static models

The time series structure is not included in static models, and explanatory variables (the factors) are added linearly to the model until the Bayes information criterion (BIC) is minimized. In some of the models the VAT component is predicted, which is combined with the sales inquiry component (that is known).

1.4. Shrinkage Models

While the factor model described above solves the curse of dimensionality by extracting a relatively small number of variables from our large dimensional dataset, resulting in a two-step procedure, shrinkage methodologies regularize the coefficients of the original predictors. Next, three regularized regressions approaches is examined, namely the ridge regression, the lasso and the elastic-net. One similarity among these models is that the predictors are included linearly. Later on, approaches that augment the set of predictors with a number of nonlinear transformations are described.

Ridge Regression

The basic idea of the ridge regression methodology is to penalise the size of the regression coefficients and shrink them toward 0. In practice, this is obtained by minimizing

$$(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)+\lambda\sum_{j=1}^{K}\beta_j^2, \tag{4}$$

where y is the variable we want to predict and X is the matrix of K predictors. λ determines the degree of shrinkage (i.e. how much we are forcing the parameters to be near 0). In a Bayesian framework this can be interpreted as imposing a prior following a normal distribution with mean 0 and variance proportional to λ. The solution of the minimization problem of gives us:

$$\widehat{\beta}_{ridge}=(\mathbf{X}'\mathbf{X}+\lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \tag{5}$$

where I is K ×K identity matrix. Notice that the ridge regression does not attempt to isolate the variables with good predictive power, instead it is aimed at regularizing the large dimensional regression solution.

Lasso

This shrinkage estimator was introduced in Tibshirani (1996). The main idea of the methodology is to produce models where the parameters of irrelevant variables are estimated to be exactly zero, leading to a variable selection setting. The minimization problem behind the lasso can be specified as

$$(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)+\lambda\sum_{j=1}^{K}|\beta_j|. \tag{6}$$

Even though lasso has many benefits, it does have some drawbacks. For example, if there are many multicollinear predictors, lasso estimation will lead to select only one of these useful predictors, disregarding all others. The elastic-net of Zou and Hastie (2005) is helpful in this scenario.

Elastic-Net

Introduced in Zou and Hastie (2005), the elastic net combines ridge-regression and the lasso. It is based on the following minimization problem

$$(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)+\lambda_1\sum_{j=1}^{K}|\beta_j|+\lambda_2\sum_{j=1}^{K}\beta_j^2$$

One of the main benefits of the elastic-net is that it is better suited in a scenario where the predictors are strongly correlated, and it has been shown to work better when the number of predictors is larger than the number of observations. Given that the firm-level data is based on turnovers, their year-on-year growth rates are expected to be fairly cross-correlated, due to the impact of aggregate business conditions. Moreover, especially when looking at firm data accumulated many days after the end of the reference month, the number of firms in our predictors set is expected to be larger than the number of time series observations.

All models are estimated using the 'glmnet' package for R. The details of the computation algorithm are given in Friedman, Hastie, and Tibshirani (2010). The degree of shrinkage (i.e. the values of λ, $\lambda 1$ and $\lambda 2$ in (1)-(3)) is selected through 10-fold cross validation. For the Elastic Net, Lasso and Ridge regression-models there are lagged and fitted versions applied besides the basic version.

2. Data sources

The main predictors in our nowcasting application are firm-level sales extracted from the sales inquiry, a monthly survey conducted by Statistics Finland for the purposes of obtaining turnovers from the most important firms in the economy. This dataset covers around 2,000 enterprises and encompasses different industries (services, trade, construction, manufacturing), representing ca. 70 per cent of total turnovers. The data are available soon after the end of the month of interest and a considerable share of the final data are accumulated around 15 to 20 days after the end of the reference month. Formally, Statistics Finland imposes a deadline to the firms, which are supposed to send their data by the end of the 15th day of the month. However, the set of firms' sales does not cover always the entire sample.

The firms are required to have long time series (starting in 2000), leading to predictors' set of 500 firms on average. The sales growth rates are computed for all the months from 2000 until the nowcast month of interest. If the firm has reported sales by the time information extracted, but has missing values during the time span (i.e. the firm did not reply at some earlier date, or the firm was not included in the turnover inquiry at that time), the missing growth rates are tried to be obtained from the earlier reported VAT data, which should include all the firms in the economy. Notice that our resulting data do not contain missing values.

Wages and salaries from the employer's contributions data from the tax administration are also available T+20 and are also used. Both sales data and wages and salaries are used to form year-on-year changes in these variables. All the input variables are stationary.

For the model selection we use retrospective test data for 2015-2016 and real time test data for 2017-2018. The real time data accumulates during the production process. For retrospective predictions the data accumulation is realistically simulated by using the time stamp of the reported sales, which allows us to track what data were available by each date of a month. Further, the more recent data points, starting from February 2017, are based on real time data collection.

3. Selection of the model

The selection of the model is made automatically in the production system based on the test data and on the production history. The appositeness of predictions is evaluated by using the Mean error (ME) and Mean absolute error (MAE) measures. The best models are selected for the month to be predicted. The "best" model cannot be known and in practice, can be seen that diversification benefits are attained with model combinations. In particular, model combinations protect against structural breaks.

There is different criteria tried in order to trim the original nowcasting models and found that keeping the models with the lowest mean error (i.e. the ones producing unbiased nowcasts of turnover indicators) tend to produce the best estimates, once combined. Once the fast estimate is produced of the indicator of interest, re-evaluate the whole set of models is re-evaluated to make sure that the performance with respect to the latest months does not alter the best set of models. This implies that, in principle, the models which are going to be included in the estimate can change over time (Stock and Watson, 2004).

The primary aim has been to attain prediction errors that would be below the revisions of current T+45 produced turnover indices.

The final method is a combination of models to which a certain number of the 70 model variations described in Section 3 are selected, based on errors obtained in testing.

Based on the above-mentioned key figures, different criteria were tested for forming combinations and the final method is left to be selected by the statistician responsible for the indicators in question, but errors based on historical performance are offered to help in the selection.

Predictions are selected in the model combinations industry-specifically based on the test period, so clearly non-functioning models could be dropped. In this way combinations are formed which include the best 25 per cent, the best ten, the best five and only the best prediction. In addition, there is a combination formed including the predictions whose revision is smaller than that of the T+45 index produced with the current method. In practice, the precondition for using the predictions has been that models are found that can go below the error of the currently produced T+45 index. Thus, a prediction could be generated for all mentioned combinations.

A monthly MAE is calculated for each model combination. On the basis of those MAEs the quality of the predictions for the model combinations have been assessed. The final prediction of each industry is a prediction produced by the selected model combination. Automatic selection was built so that it always suggests a prediction that had the lowest average MAE in the test period. However, the possibility to change this selection is left to the expert of the index in question. The selection of both the model combination and the models included in the combination is made separately for each month and the series to be predicted, so the predictions included in the model combination vary by time and industry.

References

- James H. Stock and Mark W. Watson. Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97:1167–1179, December 2002.
- James H. Stock and Mark W. Watson. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting, 23(6):405–430, 2004. ISSN 1099-131X. doi: 10.1002/for.928. URL http://dx.doi.org/10.1002/for.928.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2):301–320, 2005. ISSN 1369-7412.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference and prediction. Springer, 2 edition, 2009.
- Paolo Fornaro. Predicting Finnish economic activity using firm-level data. International Journal of Forecasting, 32(1):10–19, 2016.
- Fornaro, Paolo, Luomaranta, Henri & Saarinen, Lauri (10.1.2017). “Nowcasting Finnish Turnover Indexes Using Firm-Level Data”. ETLA Working Papers No 46. http://pub.etla.fi/ETLA-Working-Papers-46.pdf