The predictive power of the business and bank sentiment of firms: A high-dimensional Granger Causality approach

Wilms I, Gelper S, Croux C.

KBI_1520

The predictive power of the business and bank sentiment of firms: A high-dimensional Granger Causality approach

Ines Wilms[a]*, Sarah Gelper[b] and Christophe Croux[a]

[a] Faculty of Economics and Business, KU Leuven

[b] Innovation, Technology Entrepreneurship & Marketing Group, Eindhoven University of Technology

Abstract. We study the predictive power of industry-specific economic sentiment indicators for future macro-economic developments. In addition to the sentiment of firms towards their own business situation, we study their sentiment with respect to the banking sector - their main credit providers. The use of industry-specific sentiment indicators results in a high-dimensional forecasting problem. To identify the most predictive industries, we present a bootstrap Granger Causality test based on the Adaptive Lasso. This test is more powerful than the standard Wald test in such high-dimensional settings. Forecast accuracy is improved by using only the most predictive industries rather than all industries.

Keywords. Bootstrap; Granger Causality; Lasso; Sentiment surveys; Time series forecasting

*Corresponding author: ines.wilms@kuleuven.be

1 Introduction

Sentiment indicators are often considered to be among the most important leading indicators of the real economy (Dreger and Kholodilin, 2013) and are therefore closely followed by business cycle analysts, central banks and business owners (Vuchelen, 2004, Claveria et al., 2007, Martinsen et al., 2014). However, studies on the predictive power of sentiment indicators find mixed results. While many studies find that sentiment indicators have predictive power for future economic developments (Kumar et al., 1995, Hansson et al., 2005, Lemmens et al., 2005, Abberger, 2007, Klein and Oezmucur, 2010, Christiansen et al., 2014), others conclude that sentiment indicators provide only limited information for predicting economic variables (Cotsomitis and Kwan, 2006, Claveria et al., 2007, Dreger and Kholodilin, 2013 and Bruno, 2014).

An important communality between these studies is the use of aggregate sentiment indicators. This paper, instead, examines the predictive power of disaggregate sentiment indicators. Especially in the context of business sentiment – as is the topic of this paper – some segments have more predictive power than others. Here, we segment firms according to their industry. Our methodology takes into account that the different industry segments might contain predictive power for different macro-economic indicators.

To study the predictive power, we use a Granger Causality approach. A (set of) time series is said to Granger Cause another time series if the former has incremental predictive power for predicting the latter. Granger Causality tests in *low-dimensional* time series settings have a long history. They are used, among others, in macro-economics to study the predictive power of monetary aggregates for output and price variables (Sahoo and Acharya, 2010), in operational research to study the predictive power of academic literature for practitioner literature (Ghosh et al., 2010), or in finance to study the predictive power of volume for stock prices (Blasco et al., 2005). Because predictive analysis based on disaggregate sentiment indicators requires handling a large number of such indicators, we introduce a Granger Causality testing procedure applicable to *high-dimensional* time series.

Recently, a small but growing literature on inference in penalized regression models for *cross-sectional* data has arisen, such as Wasserman and Roeder (2009), Meinshausen et al. (2009) and Chatterjee and Lahiri (2011). We extend the residual bootstrap procedure of Chatterjee and Lahiri (2011) to high-dimensional *time series* data. The bootstrap test statistic, based on the Adaptive Lasso (Zou, 2006), identifies those industry segments whose predictive power is statistically significant. Our simulation study shows that this test statistic is more powerful than the standard Wald test statistic in a high-dimensional setting. Furthermore, important gains in forecast accuracy are obtained by not using all industry segments but by first selecting the most predictive ones using the bootstrap test statistic.

We use a unique data set that not only measures the sentiment of firms towards their own situation ("*business* sentiment") – as is classical for sentiment indicators – but also measures the sentiment of firms towards the banking industry ("*bank* sentiment"). For the economy to be able to grow, it is essential that firms have access to credit, typically provided by banks. Especially in the aftermath of the recent economic downturn and banking crises, distressed banks can constrain the economy (Kroszner et al., 2007, Dell'Ariccia et al., 2008, Fernandez et al., 2013). To the best of our knowledge, we are the first to study the importance of sentiment towards the banking industry.

The remainder of this article is structured as follows. Section 2 describes the data on the business and bank sentiment, as well as the macro-economic indicators. Section 3 introduces Granger Causality Testing in high-dimensional time series models. In Section 4, a simulation study shows the good performance of our methodology in terms of size and power of the test statistic and forecast accuracy. In Section 5, we apply the proposed methodology to identify the most predictive industry segments for several future macro-economic indicators. In Section 6, we show that forecast accuracy can be improved by using only the most predictive industry segments instead of all industry segments. Finally, Section 7 concludes.

Table 1: Industry Segments. Businesses are divided into 10 industry segments.

Industry	Description	Sector
Industry 1	Agriculture, forestry, fishing, mining and quarrying and other industry	Primary
Industry 2	Manufacturing	Secondary
Industry 3	Construction	Secondary
Industry 4	Wholesale and retail trade, transportation and storage accomodation and food and service activities	Tertiary
Industry 5	Information and communication	Quaternary
Industry 6	Financial and insurance activities	Quaternary
Industry 7	Real estate activities	Quaternary
Industry 8	Professional, scientific, technical administration and support service activities	Quaternary
Industry 9	Public administration, defence, education,	Quaternary
Industry 10	Other services	Quaternary

2 Data

We use a unique data set provided to us by EUWIFO, the European Economic Research Institute. EUWIFO is an owner-managed business that conducts business climate interviews. By conducting interviews with firms spread over Germany, EUWIFO gathers information on the confidence these firms have in their own economic situation and in the banking sector. Firms are divided into segments according to the industry in which they are active based on their NACE code. These 10 industry segments are listed in Table 1.

The interviews consist of two parts. In the first part, the Business Survey, firms are asked to assess their own situation. In the second part, the Bank Survey, firms are asked to assess the German bank sector.

Business Survey Each firm receives 9 questions to assess their own economic situation. They are asked to assess changes (this year compared to last year) in (1) turnover, (2) earnings, (3) number of employees, (4) investments, (5) incoming domestic orders, (6) incoming foreign orders, (7) utility and maintenance costs, (8) tax burden, and (9) cost through government red tape. For each question, answers are favorable, neutral or unfavorable. For all the firms within an industry segment, a balance of opinion indicator is calculated for each question, being the percentage of favorable answers minus the percentage of unfavorable answers. As we construct 9 sentiment indicators for each of the 10 industries, this amounts

Table 2: Macro-economic indicators. All time series are seasonally adjusted (Eurostat).

Indicator	Description
IP-A1	Production in industry: Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply
IP-A2	Production in industry: Construction, Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply
IP-M	Production in industry: Manufacturing
IP-E	Production in industry: Energy
IP-CaGo	Production in industry: Capital goods
IP-CoGo	Production in industry: Consumer goods
RT	Retail Trade, except of motor vehicles and motorcycles
WS	Wholesale Trade, except of motor vehicles and motorcycles

to 90 business sentiment indicators.

Bank Survey Each firm is asked to assess the German bank sector. In total, 243 German banks are included in the Bank Survey. Each firm first has to indicate which of these 243 German banks they know. For the banks they know, they are asked to assess their *consideration* towards that specific bank and the *reputation* of that specific bank. Answers are either favorable or unfavorable and a balance of opinion indicator is calculated for each question. We include three indicators: the average consideration indicator, averaged over all German banks, the consideration indicator towards the Sparkassen, and the consideration indicator towards the Volksbanken. The latter two are the most well known banks in Germany. We also construct three reputation indicators per industry segment following an analogous approach. As we construct three bank consideration and three bank reputation indicators for each of the 10 industries, this amounts to 60 bank sentiment indicators.

Joining the 90 business sentiment indicators and the 60 bank sentiment indicators results in a total of 150 time series. We combine all 150 sentiment indicators in one high-dimensional data set. All time series are observed over $T = 40$ months (January 2012-April 2015). We study the predictive power of these sentiment indicators for 8 German macro-economic indicators (Table 2).

The 150 time series are grouped into blocks by industry segment (cfr. Table 1). For each industry segment, we have one block of 9 indicators from the Business Survey and one block of 6 indicators from the Bank Survey. Our methodology is such that we select either all 9 business sentiment indicators for an industry, or none. Similarly, we will select either all 6 bank sentiment indicators for an industry or none. This way, we can investigate the difference in predictive power between the business and bank sentiment indicators for the 10 industries. To identify the most predictive blocks, we perform joint hypothesis tests. We test if the set of indicators in a particular block Granger Causes a particular macro-economic indicator. This predictive analysis involves a large number of disaggregate sentiment indicators. In the next section, we introduce a Granger Causality testing procedure that can handle such a high-dimensional situation.

3 High-dimensional Granger Causality Testing

Performing Granger Causality tests on a data set with many time series relative to the length of the series is challenging. In these high-dimensional settings, estimation by standard procedures becomes inaccurate. In our sentiment application, the number of time series (i.e. $k = 150$) even exceeds the length of the time series (i.e. 40), making it impossible to use standard estimation procedures. Penalized estimation brings an outcome.

3.1 Penalized Maximum Likelihood estimation

Let y_t be a one-dimensional stationary time series. We assume that y_t follows a ARX(p) model, i.e. an autoregressive model of order p with k predictor time series collected in the $(k \times 1)$ vector $\mathbf{x}_t$:

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \mathbf{a}_1 \mathbf{x}_{t-1} + \mathbf{a}_2 \mathbf{x}_{t-2} + \ldots + \mathbf{a}_p \mathbf{x}_{t-p} + e_t \,, \qquad (1)$$

where b_1 to b_p are the autoregressive parameters, the parameters $\mathbf{a}_1$ to $\mathbf{a}_p$ are $(1 \times k)$ vectors and the error term e_t is assumed to follow a $N(0, \sigma)$ distribution. We assume, without loss of generality, that all time series are mean centered such that no intercept is included.

If the number of components in $\mathbf{x}_t$ is large, the number of unknown parameters in equation (1) explodes. To ensure accurate estimation, we use Penalized Maximum Likelihood estimation (e.g. Zou, 2006 in a regression context, or Gelper et al., 2015 in a time series context). Write model (1) in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\,, \tag{2}$$

where $\mathbf{y}$ is the column vector $(y_1, \ldots, y_T)$, and the matrix $\mathbf{X} = (\underline{\mathbf{Y}}_1, \ldots, \underline{\mathbf{Y}}_p, \underline{\mathbf{X}}_1, \ldots, \underline{\mathbf{X}}_p)$. Here $\underline{\mathbf{Y}}_j$ is $(T \times 1)$, containing the values of the time series at lag j in its column; and $\underline{\mathbf{X}}_j$ is an $(T \times k)$ matrix, containing the values of the k predictor time series at lag j in its columns, for $1 \leq j \leq p$. The vector $\boldsymbol{\beta}$ contains the parameters values $b_1, \ldots, b_p, \mathbf{a}_1, \ldots, \mathbf{a}_p$, and has length $p(1+k)$. In case $p(1+k) > T$, the Maximum Likelihood estimator does not exist. The Penalized Maximum Likelihood estimator is, however, still computable.

The penalized estimator of the regression parameter $\boldsymbol{\beta}$ is obtained by minimizing the negative log likelihood with a penalization on the elements of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_\lambda = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{i=1}^{p(1+k)} \hat{w}_i |\beta_i|\,, \tag{3}$$

where $\hat{w}_i$ are weights and $\lambda > 0$ is a sparsity parameter. This estimator is the Adaptive Lasso (Zou, 2006). It generalizes the popular Lasso (e.g. Hastie et al., 2009, Chapter 3) which shows good performance in operational research (e.g. Ballings and Van den Poel, 2015, Huang et al., 2014). The Adaptive Lasso ensures that the bootstrap (Section 3.3) is consistent (Chatterjee and Lahiri, 2011). We take the weights of the Adaptive Lasso $\hat{w}_i = 1/|\hat{\beta}_i^{\text{ridge}}|$, where the Ridge estimator (Hastie et al., 2009, Chapter 3) is

$$\hat{\boldsymbol{\beta}}_\lambda^{\text{ridge}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_{\text{ridge}} \sum_{i=1}^{p(1+k)} \beta_i^2.$$

The sparsity parameter λ and the order of the ARX, p, are selected using the Bayesian Information Criterion (BIC) (e.g. Abegaz and Wit, 2013 and references therein):

$$\text{BIC}_\lambda = T \cdot \log\left(\frac{1}{T}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\lambda)'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\lambda)\right) + df_\lambda \cdot \log(T),$$

where df_λ equals the number of non-zero estimated regression coefficients. We solve (3) over a range of values for λ and select the one with lowest value of the BIC. To select the order of the ARX model, we estimate the ARX model for different values of p, each time using the optimal value of λ for that value of p. We then select the order p of the ARX model again by minimizing the BIC.

3.2 Granger Causality in the ARX framework

We partition the vector $\mathbf{x}_t$ in different blocks, and denote the j^{th} block of $\mathbf{x}_t$ by $\mathbf{x}_{t,j}$, consisting of k_j time series. In the ARX model (1), denote the j^{th} block of coefficients at lag i corresponding to $\mathbf{x}_{t,j}$ by $\mathbf{a}_{i,j}$. The multivariate time series $\mathbf{x}_{t,j}$ is said to Granger Cause y_t if the former has incremental predictive power for the latter. We say that $\mathbf{x}_{t,j}$ does not Granger Cause y_t if the coefficients on all lags of $\mathbf{x}_{t,j}$ are equal to zero, i.e. $\mathbf{a}_{1,j} = \ldots = \mathbf{a}_{p,j} = \mathbf{0}$.

The Adaptive Lasso estimator in (3) is sparse, meaning that some of its elements are exactly zero. The larger the value of λ, the sparser the estimator. The "Granger Lasso Selection" method (e.g. Fujita et al., 2007, Bahadori and Liu, 2013) says that a time series $\mathbf{x}_{t,j}$ Granger Causes y_t if at least one of the corresponding parameters $\mathbf{a}_{1,j}, \ldots, \mathbf{a}_{p,j}$ is estimated as non-zero. Our approach is different, we infer Granger Causality relations from a bootstrap testing procedure.

3.3 Granger Lasso test

The null hypothesis that a block of time series $\mathbf{x}_{t,j}$ is not Granger Causing y_t can be stated as

$$H_0: \ \mathbf{R}_j\boldsymbol{\beta} = \mathbf{0}, \tag{4}$$

where $\mathbf{R}_j$ is a suitable $pk_j \times p(1+k)$ matrix. The elements of $\mathbf{R}_j$ are either zero or one. We assign the value one to the elements of $\mathbf{R}_j$ corresponding to the autoregressive parameters $\mathbf{a}_{1,j}, \ldots, \mathbf{a}_{p,j}$. The corresponding Wald test statistic is given by

$$Q = (\mathbf{R}_j\widehat{\boldsymbol{\beta}})'(\mathbf{R}_j\mathrm{Cov}(\widehat{\boldsymbol{\beta}})\mathbf{R}_j')^{-1}(\mathbf{R}_j\widehat{\boldsymbol{\beta}}). \tag{5}$$

To bootstrap this test statistic, we use the following residual bootstrap procedure (Kreiss and Lahiri, 2012):

1. Estimate the model under the null hypothesis, i.e. model (1) with the block $\mathbf{x}_{t,j}$ removed at the right-hand-side. Compute the centered residuals $\widehat{\varepsilon}_t$, for $t = 1, \ldots, T$.

2. Let $B = 500$ be the number of bootstraps. For $b = 1, \ldots, B$:

   (a) Construct the bootstrap time series y_t^* from model (1) with the parameter estimates from step 1 and with bootstrap errors $\varepsilon_t^* = \widehat{\varepsilon}_{\mathcal{U}_t}$ with $\mathcal{U}_t, t = 1, \ldots, T$ an i.i.d. sequence of discrete random variables uniformly distributed on $\{1, \ldots, T\}$. The predictor time series are kept fixed.

   (b) Apply the Penalized Maximum Likelihood estimator of equation (3) to the bootstrap sample. Denote the bootstrap estimate by $\widehat{\boldsymbol{\beta}}_b^*$.

   (c) Compute the bootstrap statistic $Q_b^* = (\mathbf{R}_j \widehat{\boldsymbol{\beta}}_b^*)'(\mathbf{R}_j \text{Cov}(\widehat{\boldsymbol{\beta}})\mathbf{R}_j')^{-1}(\mathbf{R}_j \widehat{\boldsymbol{\beta}}_b^*)$.

3. Compute

$$\text{mid } p\text{-value } = \frac{1}{B} \sum_{b=1}^{B} \Big(I(Q_b^* > Q) + \frac{1}{2} I(Q_b^* = Q) \Big),$$

   with Q_b^* (for $b = 1, \ldots, B$) B independent bootstrap statistics. $I(\cdot)$ is an indicator function that takes on the value one if its argument is true and equals zero otherwise. We use the mid p-value (Lancaster, 1949) since it may occur that the value of the test statistic and the bootstrap test statistic are both equal to zero.

4 Simulation study

By means of a simulation experiment, we (i) evaluate the size and power of the Granger Lasso test and (ii) conduct a forecast exercise. We generate y_t according to the following ARX(1) model

$$y_t = 0.5 y_{t-1} + \mathbf{a}_1 \mathbf{x}_{t-1} + e_t, \tag{6}$$

where $e_t \sim N(0, 0.1)$. The predictors are generated as autoregressive processes $\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \mathbf{u}_t$, with $\mathbf{u}_t \sim N_k(\mathbf{0}, 0.1\mathbf{I})$, $\mathbf{C} = 0.5\mathbf{I}$ and $\mathbf{I}$ the k-dimensional identity matrix. The model parameters are chosen according to the four designs detailed in Table 3. The first three

Table 3: Simulation designs.

Design	under H_0	under H_A
$T = 100, k = 25$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$
$T = 100, k = 50$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$
$T = 100, k = 75$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.2}_{1\times5} & \mathbf{0.2}_{1\times5} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times(k-15)} \end{bmatrix}$
$T = 40, k = 150$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.4}_{1\times9} & \mathbf{0}_{1\times9} & \dots & \mathbf{0}_{1\times9} & \mathbf{0}_{1\times6} & \dots & \mathbf{0}_{1\times6} \end{bmatrix}$	$\mathbf{a}_1 = \begin{bmatrix} \mathbf{0.4}_{1\times9} & \mathbf{0.4}_{1\times9} & \mathbf{0}_{1\times9} & \dots & \mathbf{0}_{1\times9} & \mathbf{0}_{1\times6} & \dots & \mathbf{0}_{1\times6} \end{bmatrix}$

designs are the same except for the number of time series k. In design two and three, we add more non-informative time series to the model, i.e. time series with a coefficient equal to zero. The standard Maximum Likelihood estimator is computable in these three designs. The last design corresponds to the design of our sentiment application, with $k = 150$ predictor time series and $T = 40$. Here, only the Penalized Maximum Likelihood estimator is computable.

For each design, we consider a data generating process under the null hypothesis H_0 and under the alternative hypothesis H_A. We divide the time series $\mathbf{x}_t$ and the corresponding coefficient vector $\mathbf{a}_1$ into several blocks, as can be seen from Table 3. The first block of time series Granger Cause the response both under H_0 and under H_A. The second block of time series Granger Cause the response only under H_A. The remaining blocks of time series never Granger Cause the response. In the first three designs, block one to three each contain five time series, the fourth block contains the remaining ones. In the last design, there are 20 blocks, similar to our sentiment application.

4.1 Size and power of the test statistic

We test the null hypothesis that the second block of time series does not Granger Cause the response. We compare the performance of Granger Lasso test to the standard Wald test computed from the standard Maximum Likelihood (ML) estimator.

To study the *size* of the test statistic, we simulate $N = 1000$ time series under the null hypothesis and compute the simulated size, i.e. the proportion of simulation runs were the

Table 4: Simulated sizes for the Wald test and Granger Lasso test.

Simulation design	Wald test		Granger Lasso test	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
$T = 100, k = 25$	0.017	0.064	0.013	0.058
$T = 100, k = 50$	0.025	0.079	0.010	0.052
$T = 100, k = 75$	0.035	0.082	0.015	0.051
$T = 40, k = 150$	NA	NA	0.007	0.051

null hypothesis is rejected:

$$\text{Simulated size} = \frac{1}{N}\sum_{j=1}^{N} I(p_j^{H_0} < \alpha), \tag{7}$$

where $p_j^{H_0}$ is the mid p-value obtained in simulation run $j = 1, \ldots, N$, and α is the pre-specified significance level. We consider $\alpha = 0.01$ and $\alpha = 0.05$.

Results. Table 4 shows the simulated sizes for the standard Wald test and the Granger Lasso test. The simulated sizes of the Granger Lasso test and the standard Wald test are both close to the nominal size α in the design with $T = 100, k = 25$. When the number of time series increases relative to the length of the time series (i.e. second and third design), the Granger Lasso test remains accurately sized whereas the standard Wald test statistic gets distorted: its simulated size deviates strongly from the nominal size. In the last design, only the Granger Lasso test is available. For both $\alpha = 0.01$ and $\alpha = 0.05$, the Granger Lasso test is reasonably accurately sized.

To study the *power* of the test statistic, we use size-power curves (see Davidson and MacKinnon, 1998). Size-power curves are constructed using two empirical distribution functions. We carry out the following steps:

1. Simulate $N = 1000$ time series under the null hypothesis. Compute for each simulation run $j = 1, \ldots, N$ the mid p-value $p_j^{H_0}$. Calculate the empirical distribution function of the p-values:

$$\widehat{F}^{H_0}(x_i) = \frac{1}{N}\sum_{j=1}^{N} I(p_j^{H_0} \leq x_i),$$

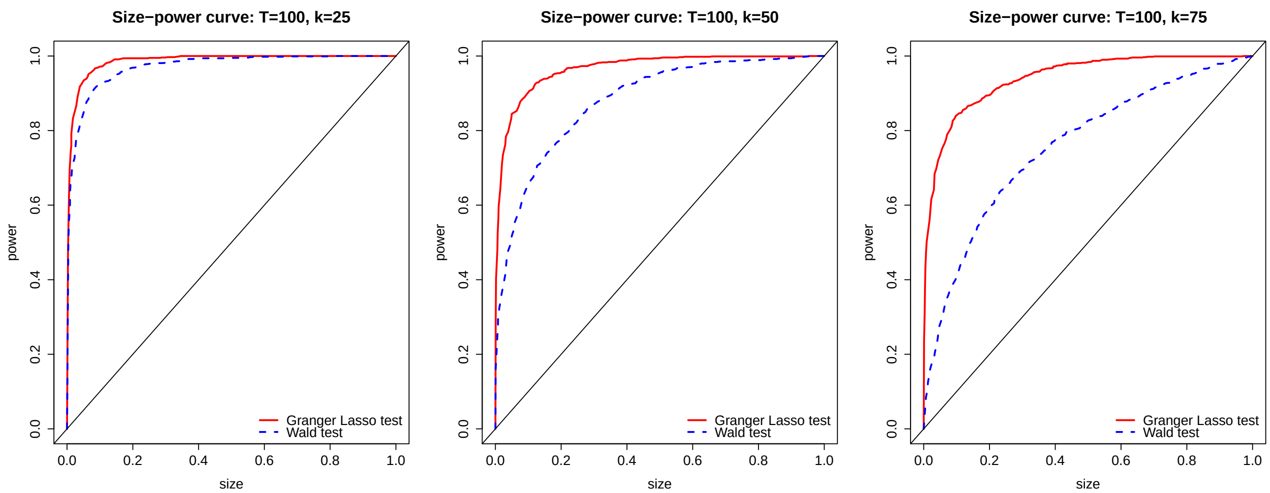

Figure 1: Size-power curve of the Granger Lasso test (solid gray line) and the standard Wald test (dashed line), for increasing number of time series $k = 25$ (left), $k = 50$ (middle) and $k = 75$ (right) with time series length $T = 100$. The 45°line is indicated as well.

for a grid of values $x_i, i = 1, \ldots, m$ between zero and one.

2. Simulate $N = 1000$ time series under the alternative hypothesis. Compute for each simulation run $j = 1, \ldots, N$ the mid p-value $p_j^{H_A}$. Calculate

$$\widehat{F}^{H_A}(x_i) = \frac{1}{N} \sum_{j=1}^{N} I(p_j^{H_A} \leq x_i).$$

3. Plot $\widehat{F}^{H_0}(x_i)$ against $\widehat{F}^{H_A}(x_i)$, for $x_i, i = 1, \ldots, m$.

Results. Size-power curves of the Granger Lasso test and standard Wald test are shown in Figure 1 (first three designs). The larger the difference between the size-power curve and the 45°line, the more power the test has. For $k = 25$ (i.e. left panel) both curves are rapidly increasing and very similar. When the number of time series increases (i.e. middle and right panel), the size-power curve of the Granger Lasso test is hardly affected, and achieves a much larger power than the standard Wald test.

4.2 Forecast exercise

For forecasting the time series y_t, we use a two-step procedure. First, we select predictor time series. Second, we estimate the model with only the selected predictor time series. We consider four selection and four estimation techniques, yielding 16 selection-estimation combinations. We investigate the performance of each combination in forecasting the response.

As selection techniques we consider: (1) use all time series, (2) use the standard Wald test to discard blocks of time series that are not Granger Causing the response, (3) use Granger Lasso Selection (cfr. Section 3.1) to discard blocks of time series that are not Granger Causing the response, (4) use the Granger Lasso test to discard blocks of time series that are not Granger Causing the response. Selection technique (4) is our proposed selection technique. The tests are carried out at a 1% significance level.

After selecting the predictor time series, we forecast the response using either (1) Maximum Likelihood, (2) the Adaptive Lasso estimator, (3) Bayesian shrinkage with the Minnesota prior (Litterman, 1986), (4) the Factor Model of Stock and Watson (2002). These are all leading methods for macro-economic forecasting (Inoue and Kilian, 2008). Methods (2) and (3) perform shrinkage. Where the Adaptive Lasso puts some of the estimated coefficients exactly to zero, the Bayesian estimator only shrinks the estimated coefficients towards zero. Factor Models reduce the dimension of the predictor time series by extracting a small number of common factors using principal component analysis.[1]

To evaluate forecast accuracy, we conduct a rolling window forecast exercise. We use a window of size $S = \lfloor 0.90 \cdot T \rfloor$. At each point $t = S, \ldots, T-1$, the models are re-estimated and one-step-ahead forecasts are calculated. We evaluate the forecast accuracy of each selection-estimation technique combination by calculating the Mean Absolute Forecast Error[2]

$$\text{MAFE} = \frac{1}{T-S} \sum_{t=S}^{T-1} \left| \hat{y}_{t+1} - y_{t+1} \right|, \tag{8}$$

[1]The number of factors r is determined by calculating the maximum eigenvalue ratio criterion $\hat{r}_j = \hat{\lambda}_j / \hat{\lambda}_{j+1}$ for $j = 1, \ldots, \mathrm{k}-1$ from the eigenvalues $\hat{\lambda}_j, \ldots, \hat{\lambda}_\mathrm{k}$ and selecting $r = \text{argmax}_j \hat{r}_j$.

[2]Similar conclusions can be drawn by looking at the Mean Squared Forecast Error.

Table 5: Average MAFE for the four selection techniques (rows) and four estimation techniques (columns).

Simulation design	Selection technique	Estimation technique			
		ML	Adaptive Lasso	Bayesian	Factor Model
$T = 100, k = 25$	All	0.093	0.089	0.116	0.129
	Wald test	0.082	0.082	0.121	0.086
	Granger Lasso Selection	0.089	0.085	0.118	0.121
	Granger Lasso test	0.082	0.082	0.120	0.086
$T = 100, k = 50$	All	0.126	0.092	0.122	0.138
	Wald test	0.087	0.084	0.124	0.089
	Granger Lasso Selection	0.119	0.092	0.122	0.137
	Granger Lasso test	0.084	0.083	0.124	0.086
$T = 100, k = 75$	All	0.208	0.089	0.123	0.141
	Wald test	0.117	0.088	0.121	0.107
	Granger Lasso Selection	0.170	0.091	0.123	0.140
	Granger Lasso test	0.083	0.080	0.119	0.085
$T = 40, k = 150$	All	NA	0.189	0.315	0.322
	Granger Lasso Selection	NA	0.181	0.305	0.300
	Granger Lasso test	NA	0.165	0.379	0.199

where $\hat{y}_{t+1}$ is the predicted response for time $t+1$. The MAFE is computed for each simulated time series, and their average over $N = 100$ simulation runs is reported in Table 5.

Results. Table 5 shows that selecting predictor time series is better than taking all series, for all estimation techniques (except the Bayesian shrinkage estimator). Among the selection techniques, improvements are larger with our Granger Lasso test compared to the Granger Lasso Selection approach. Granger Lasso Selection discards less blocks of time series compared to the Granger Lasso test, yielding less parsimonious models and reduced forecast performance. When the number of time series increases relative to the length of the time series, the Granger Lasso test also performs substantially better than the standard Wald test. Paired t-tests confirm that (in the majority of cases), the improvements of the Granger Lasso test compared to the other selection techniques are significant. More precisely, the good performance of the Granger Lasso test is most pronounced in the high-dimensional designs: it performs significantly best - among the four selection techniques - in 8 out of 12 cases (design $T = 100, k = 50$), 12 out of 12 cases (design $T = 100, k = 75$), and 6 out of 9 cases (design $T = 40, k = 150$).

For all simulation designs, the best forecast always involves the Granger Lasso test. Among the estimation techniques, the Adaptive Lasso performs best. After the first selection of predictive blocks of time series, the Adaptive Lasso can further reduce the number of predictor time series in the second step. This is most suited for settings with a few number of relevant predictor time series and a large number of irrelevant, noise predictor time series. Similar conclusions are obtained by Bühlmann and Hothorn (2010) who discuss a "Twin Boosting" procedure for improved feature selection and prediction.

5 The role of business and bank sentiment for macro-economic forecasting

We identify the most predictive industry segments for future macro-economic developments using the Granger Lasso test from Section 3.

5.1 Model

We estimate 8 ARX models, one for each macro-economic indicator to predict. The time series y_t entering model (1) is one of the 8 macro-economic indicators of Table 2 taken in first differences. The vector $\mathbf{x}_t$ contains the $k = 150$ business and bank sentiment indicators in first differences at time t. We use differences to ensure stationarity of the time series.[3] We

[3]Following standard practice, we first test for stationarity. A stationarity test of all individual time series using the Augmented Dickey-Fuller test indicates that most time series in levels are integrated of order 1.

estimate each ARX model using the Penalized Maximum Likelihood estimator from Section 3. Then, we perform Granger Causality tests, one for each of the 20 blocks of sentiment indicators (cfr. Section 2). As such, we test if the opinion of a particular industry segment - as measured through the Business Survey - has incremental predictive power for the German macro-economic indicators. We repeat this exercise for each industry segment using the Bank Survey.

5.2 Identifying the most predictive industries

For each industry, Table 6 reports the p-value of the test that the opinion of that particular industry does not Granger Cause a particular macro-economic indicator. Significant results at the 1% level are in bold. We discuss the results by building on the sectoral classification framework which distinguishes the primary, secondary, tertiary and quaternary sector.

Business Survey. The primary sector, unlike the other sectors, has almost no incremental predictive power. The primary sector's contribution to Germany's GDP is also the smallest. The secondary industry has most incremental predictive power for the macro-economic indicators to which these sectors contribute most (IP-A1, IP-A2, IP-M and IP-E). Firms active in the tertiary and especially the quaternary sector have incremental predictive power for several macro-economic indicators. This sector consists of the knowledge-based part of the economy, and accounts for roughly 65% of Germany's GDP. Firms active in these sectors are at the heart of the whole economy.

Bank Survey. The Bank Survey contains less incremental predictive power than the Business Survey. The predictive power of bank sentiment for predicting future macro-economic developments is limited. This is in line with Dell'Ariccia et al. (2008) who find that the real effects of a banking crisis are limited in developed countries, in countries that have more access to foreign financing, and countries where banking crises are less severe, which all apply to Germany.

Table 6: P-values of the Granger Causality test with null hypothesis that the opinion of a particular industry segment (rows) does not Granger Cause a particular macro-economic indicator (columns). Significant results at the 1% level are in bold.

			Macro-economic indicators							
	Industry segment	Sector	IP-A1	IP-A2	IP-M	IP-E	IP-CaG	IP-CoG	RT	WS
Business Survey	Agriculture, mining & other industry	Primary	0.03	0.04	0.03	0.99	**0.01**	0.01	0.01	0.84
	Manufacturing	Secondary	**0.01**	0.07	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	0.37
	Construction	Secondary	**0.01**	**0.00**	0.01	0.04	**0.00**	0.70	**0.00**	0.50
	Wholesale, retail trade, transportation, food & service	Tertiary	0.02	**0.00**	0.04	**0.01**	0.02	0.923	0.27	0.06
	Information & communication	Quaternary	0.92	0.02	0.90	**0.00**	0.02	0.50	0.04	0.04
	Finance	Quaternary	0.56	0.03	0.13	**0.00**	0.06	0.04	0.13	0.39
	Real estate	Quaternary	0.96	0.84	0.26	**0.01**	1.00	**0.00**	**0.00**	0.60
	Administration & support	Quaternary	**0.01**	0.03	**0.01**	**0.00**	**0.00**	**0.01**	0.21	**0.00**
	Public services	Quaternary	**0.00**	0.02	0.23	0.04	**0.00**	0.02	0.86	0.04
	Other services	Quaternary	0.05	**0.00**	**0.01**	**0.00**	**0.00**	0.07	0.66	0.12
Bank Survey	Agriculture, mining & other industry	Primary	1.00	1.00	1.00	0.59	1.00	0.92	0.86	0.90
	Manufacturing	Secondary	0.05	0.20	0.06	1.00	0.99	0.14	0.85	0.39
	Construction	Secondary	0.82	0.82	0.92	**0.01**	1.00	0.70	0.84	0.03
	Wholesale, retail trade, transportation, food & service	Tertiary	1.00	0.76	0.98	1.00	**0.00**	0.04	0.53	0.23
	Information & communication	Quaternary	0.72	0.02	0.09	1.00	0.04	0.53	0.05	0.79
	Finance	Quaternary	0.98	1.00	1.00	**0.01**	1.00	0.40	0.09	0.08
	Real estate	Quaternary	0.76	0.90	0.60	1.00	1.00	0.73	0.80	0.62
	Administration & support	Quaternary	**0.01**	0.29	**0.00**	1.00	0.80	0.78	0.68	**0.00**
	Public services	Quaternary	0.03	0.07	**0.01**	0.03	0.03	0.03	0.03	0.05
	Other services	Quaternary	0.46	0.77	0.82	0.47	0.69	0.05	0.16	0.98

5.3 Robustness checks

Our main research question is whether the sentiment of different industry segments has predictive power for macro-economic indicators. Our methodology is also applicable to other ways of segmenting firms, as *region* in which the are located or according to their *company size*. For our data, there are 10 regions and three company sizes. We re-estimate the 8 ARX models and perform the Granger Causality tests for the 20 regional blocks (i.e. 10 blocks for the Business Survey, 10 blocks for the Bank Survey). Likewise, we re-estimate the 8 ARX models and perform the Granger Causality tests for the 6 company size blocks (i.e. 3 blocks for the Business Survey, 3 blocks for the Bank Survey).

Similar as for the industry results discussed in Section 5.2, we find that the business sentiment has more incremental predictive power compared to the bank sentiment. Furthermore, Germany's largest geo-economical regions, Ruhr area and the Southern states, have most incremental predictive power for the macro-economic indicators to which their day-to-day business contributes most, i.e. IP-A1, IP-A2, IP-M, IP-E and IP-CaGo, IP-CoGo respectively. Finally, small- and medium-sized companies have more incremental predictive power than large companies. Germany is dominated by small- to medium-sized companies who are global market leaders in their segments, and, hence, those might be best at evaluating Germany's economy. Detailed results are available from the authors upon request.

6 Forecasting German macro-economic developments

We perform a rolling-window forecast exercise using a window of length $S = 30$. For each time window, we estimate the 8 ARX models. We use the same selection and estimation techniques as in Section 4.2, except for the standard Wald test and the ML estimator which are not available since the number of time series exceeds the time series length. Next, one-step-ahead forecasts are computed for $t = S + 1, \ldots, T$. We report the Mean Absolute Forecast Error, see equation (8), for each macro-economic indicator and each selection-estimation technique combination in Table 7.

Among the selection techniques, the proposed Granger Lasso test performs best. It attains the lowest value of the MAFE in 20 out of 24 cases (84% of the cases). The MAFEs when either all industries are used or when Granger Lasso Selection is used are close to each other. It turns out that the latter (overall) does not discard any of the industry blocks. In contrast, a much more parsimonious model is obtained using the Granger Lasso test. These parsimonious models lead to an improved forecast accuracy, in the majority of cases.

For the Adaptive Lasso, the Granger Lasso test leads to the lowest MAFE for 7 out of 8 macro-economic indicators. The MAFEs with the Granger Lasso test are, on average, 40% lower compared to the other selection techniques. After the first selection step where either an entire block of business or bank sentiment indicators is selected or not, the Adaptive

Table 7: Mean Absolute Forecast Error for the three selection techniques (rows), the three estimation techniques (columns), and the 8 macro-economic indicators (blocks).

Selection technique	Response	Estimation technique			Response	Estimation technique		
		Adaptive Lasso	Bayesian	Factor Model		Adaptive Lasso	Bayesian	Factor Model
All	IP-A1	1.460	0.921	1.275	IP-CaGo	2.734	1.892	3.147
Granger Lasso Selection		1.460	0.921	1.275		2.734	1.892	3.147
Granger Lasso test		1.138	0.962	0.937		3.707	1.834	2.926
All	IP-A2	1.462	0.817	1.207	IP-CoGo	1.142	0.609	0.918
Granger Lasso Selection		1.462	0.817	1.207		1.142	0.609	0.918
Granger Lasso test		0.567	0.640	1.006		0.777	0.617	0.915
All	IP-M	1.720	1.117	1.641	RT	2.025	1.109	1.723
Granger Lasso Selection		1.720	1.117	1.641		2.025	1.109	1.723
Granger Lasso test		1.688	1.090	1.342		1.140	1.035	1.510
All	IP-E	2.237	1.171	2.105	WS	1.524	0.530	0.800
Granger Lasso Selection		2.237	1.171	2.105		1.524	0.530	0.800
Granger Lasso test		1.249	0.959	1.601		0.566	0.685	0.677

Lasso allows some of the time series belonging to a one of the selected blocks to be discarded in this second stage. Further reducing the number of relevant predictor time series within the selected blocks improves forecast accuracy.

In line with the results of our simulation study, pre-selecting based on the Granger Lasso test is less favorable for the Bayesian shrinkage estimator compared to the other estimation techniques. Nevertheless, the Granger Lasso test in combination with the Bayesian shrinkage estimator still leads to the lowest MAFE for 5 out of 8 macro-economic indicators, with an average reduction in MAFE of 10%.

For the Factor Model, the Granger Lasso test consistently leads to the lowest MAFE. The MAFEs with the Granger Lasso test are, on average, 20% lower compared to the other selection techniques. Discarding the least predictive industry blocks in this high-dimensional data set and estimating the factors based on the most predictive industry blocks thus leads to important gains in forecast accuracy. This result is in line with Bai and Ng (2008) who find important gains in forecast accuracy from diffusion index models by not using all predictors but by using fewer, informative predictors.

Robustness checks. We investigate the robustness of the results to the choice of segmentation criterion. We repeat the same forecast exercise using the region segments and company size segments instead of the industry segments (cfr. Section 5.3). The conclusions obtained with either the industry, region or company size segments are very similar. For the regional segments, the Granger Lasso test is the best performing selection technique and attains the lowest value of the MAFE in 71% of the cases (17 out of 24). Similarly for the company size segments where the Granger Lasso test leads towards the lowest MAFE in 71% of the cases (17 out of 24). Detailed results are available from the authors upon request.

7 Discussion

This paper presents a high-dimensional Granger Causality test. It detects the most predictive industry segments for future macro-economic developments. For this purpose, we use both business and bank sentiment surveys answered by firms across Germany. Not all industry-specific sentiment indicators are equally predictive for all macro-economic indicators. Industries contain most predictive power for the macro-economic indicators most closely tied to their day-to-day business activities.

Our forecast exercise shows that important gains in forecast accuracy can be obtained by not using all industry segments, but by first selecting the most predictive ones using the Granger Lasso test. This selection of the most pertinent industry segments provides important information for institutes conducting these sentiment surveys. For instance, instead of equally spreading respondents among all segments, the number of respondents in predictive segments could be increased, whereas the number of respondents in non-predictive segments could be decreased. Alternatively, non-predictive segments could even be completely discarded, which provides an opportunity to obtain cost savings.

The identification of pertinent respondents also applies to consumer sentiment surveys. In the large literature on consumer sentiment, this topic has received little attention. We perform a similar exercise as described in this paper using a consumer sentiment survey data set from the National Bank of Belgium. Sentiment indicators are available for different classes of consumers' net disposable income, profession, employment status, education, age and gender. We study their predictive power for several retail trade indicators. The profession, education, and age sentiment indicators contain most predictive power. Again, important gains in forecast accuracy can be obtained by first selecting the most predictive sentiment indicators (for a specific target variable of interest) instead of using all indicators.

In our sentiment application, the Business Survey contains more predictive power than the Bank Survey. Future research could further deepen our understanding on the usefulness of bank sentiment. It would be interesting to investigate if this sentiment differs between, for instance, countries that are more or less severely hit by banking crises, and developed or developing countries. The study of sentiment with respect to the banking sector opens a rich area of new research on sentiment surveys.

Acknowledgments The authors gratefully acknowledge financial support from the FWO (Research Foundation Flanders, contract number 11N9913N). We would also like to thank EUWIFO for providing the data.

References

Abberger, K. (2007), "Qualitative business surveys and the assessment of employment A case study for Germany," *International Journal of Forecasting*, 23(2), 377–389.

Abegaz, F. and Wit, E. (2013), "Sparse time series chain graphical models for reconstructing genetic networks," *Biostatistics*, 14(3), 586–599.

Bahadori, M. and Liu, Y. (2013), "An examination of practical Granger causality inference," *Proceedings of the 2013 SIAM International Conference on Data Mining*.

Bai, J. and Ng, S. (2008), "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146(2), 304–317.

Ballings, M. and Van den Poel, D. (2015), "CRM in social media: Predicting increases in Facebook usage frequency," *European Journal of Operational Research*, 244(1), 248–260.

Blasco, N.; Corredor, P.; Del Rio, C. and Santamaria, R. (2005), "Bad news and Dow Jones make the Spanish stocks go round," *European Journal of Operational Research*, 163(1), 253–275.

Bruno, G. (2014), "Consumer confidence and consumption forecast: a non-parametric approach," *Empirica*, 41(1), 37–52.

Bühlmann, P. and Hothorn, T. (2010), "Twin Boosting: Improved feature selection and prediction," *Statistics and Computing*, 20(2), 119–138.

Chatterjee, A. and Lahiri, S. (2011), "Bootstrapping Lasso estimators," *Journal of the American Statistical Association*, 106(494), 608–625.

Christiansen, C.; Eriksen, J. and Moller, S. (2014), "Forecasting US recessions: The role of sentiment," *Journal of Banking & Finance*, 49, 459–468.

Claveria, O.; Pons, E. and Ramos, R. (2007), "Business and consumer expectations and macroeconomic forecasts," *International Journal of Forecasting*, 23(1), 47–69.

Cotsomitis, J. and Kwan, A. (2006), "Can consumer confidence forecast household spending? Evidence from the European commission business and consumer surveys," *Southern Economic Association*, 72(3), 597–610.

Davidson, R. and MacKinnon, J. (1998), "Graphical methods for investigating the size and power of hypothesis tests," *The Manchester School*, 66(No. 1), 1–26.

Dell'Ariccia, G.; Detragiache, E. and Rajan, R. (2008), "The real effects of banking crises," *Journal of Financial Intermediation*, 17, 89–112.

Dreger, C. and Kholodilin, D. (2013), "Forecasting Private Consumption by Consumer Surveys," *Journal of Forecasting*, 32(1), 10–18.

Fernandez, A.; Gonzalez, F. and Suarez, N. (2013), "The real effect of banking crises: Finance or asset allocation effects? Some international evidence," *Journal of Banking & Finance*, 37(7), 2419–2433.

Fujita, A.; Sato, J.; Garay-Malpartida, H.; Yamaguchi, R.; Miyano, S.; Sogayar, M. and

Ferreira, C. (2007), "Modeling gene expression regulatory networks with the sparse vector autoregressive model," *BMC Systems Biology*, 1, No. 39.

Gelper, S.; Wilms, I. and Croux, C. (2015), "Identifying demand effects in a large network of product categories," *Journal of Retailing. Available online 10 June 2015, doi:10.1016/j.jretai.2015.05.005.*

Ghosh, S.; Troutt, M.; Thornton, J. and Offodile, O. (2010), "An empirical method for assessing the research relevance gap," *European Journal of Operational Research*, 201(3), 942–948.

Hansson, J.; Jansson, P. and Lof, M. (2005), "Business survey data: Do they help in forecasting GDP growth," *International Journal of Forecasting*, 21(2), 377–389.

Hastie, T.; R., T. and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer.

Huang, T.; Fildes, R. and Soopramanien, D. (2014), "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem," *European Journal of Operational Research*, 237(2), 738–748.

Inoue, A. and Kilian, L. (2008), "How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation," *Journal of the American Statistical Association*, 103(482), 511–522.

Klein, L. and Oezmucur, S. (2010), "The use of consumer and business surveys in forecasting," *Economic Modelling*, 27(6), 1453–1462.

Kreiss, J. and Lahiri, S. (2012), *Bootstrap methods for time series. In: Rao, T., Rao, S. and Rao, C. (Eds.) Handbook of Statistics 30. Time Series Analysis: Methods and Applications*, North Holland.

Kroszner, R.; Laeven, L. and Klingebiel, D. (2007), "Banking crises, financial dependence and growth," *Journal of Financial Economics*, 84(1), 187–228.

Kumar, V.; Leone, R. and Gaskins, J. (1995), "Aggregate and disaggregate sector forecasting using consumer confidence measures," *International Journal of Forecasting*, 11(3), 361–377.

Lancaster, H. (1949), “The combination of probabilities arising from data in discrete distributions,” *Biometrika*, 36, 370–382.

Lemmens, A.; Croux, C. and Dekimpe, M. (2005), “On the predictive content of production surveys: A pan-European study,” *International Journal of Forecasting*, 21(2), 363–375.

Litterman, R. (1986), “Forecasting with Bayesian vector autoregressions: Five years of experience,” *Journal of Business & Economic Statistics*, 4, 25–38.

Martinsen, K.; Ravazzolo, F. and Wulfsberg, F. (2014), “Forecasting macroeconomic variables using disaggregate survey data,” *International Journal of Forecasting*, 30(1), 65–77.

Meinshausen, N.; Meier, L. and Bühlmann, P. (2009), “p-values for high-dimensional regression,” *Journal of American Statistical Association*, 104(488), 1671–1681.

Sahoo, B. and Acharya, D. (2010), “An alternative approach to monetary aggregation in DEA,” *European Journal of Operational Research*, 204(3), 672–682.

Stock, J. and Watson, M. (2002), “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20(2), 147–162.

Vuchelen, J. (2004), “Consumer sentiment and macroeconomic forecasts,” *Journal of Economic Psychology*, 25(4), 493–506.

Wasserman, L. and Roeder, K. (2009), “High dimensional variable selection,” *The Annals of Statistics*, 37(5A), 2178–2201.

Zou, H. (2006), “The adaptive lasso and its oracle properties,” *Journal of the American Statistical Association*, 101, 1418–1429.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be