MAGKS

Aachen Siegen Gießen Marburg Kassel Göttingen

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 01-2011

Ivan Savin and Peter Winker

Heuristic model selection for leading indicators in Russia and Germany

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Heuristic model selection for leading indicators in Russia and Germany

Ivan Savin, Peter Winker
Department of Economics, Justus Liebig University Giessen

Abstract

Business tendency survey indicators are widely recognized as a key instrument for business cycle forecasting. Their leading indicator property is assessed with regard to forecasting industrial production in Russia and Germany. For this purpose, vector autoregressive (VAR) models are specified and estimated to construct forecasts. As the potential number of lags included is large, we compare full–specified VAR models with subset models obtained using a Genetic Algorithm enabling 'holes' in multivariate lag structures. The problem is complicated by the fact that a structural break and seasonal variation of indicators have to be taken into account. The models allow for a comparison of the dynamic adjustment and the forecasting performance of the leading indicators for both countries revealing marked differences between Russia and Germany.

Keywords: Leading indicators, business cycle forecasts, VAR, model selection, genetic algorithms.

JEL Classification: C32, C52, C53, C61, E37.

1 Introduction

For many OECD countries, indicators based on business surveys became a standard tool for the analysis and forecasting of business cycles and for assessing models of expectation formation. Recent contributions and reviews on this literature include Chamberlin (2007), Claveria *et al.* (2007) and Goggin (2008). These indicators are also important as information on the current status of the economy given the publication delay of data on real production such as GDP or – with a smaller time lag – industrial production (Hüfner and Schröder 2002, p 317). Such information is of particular interest in periods of a rapid changing economic environment like the financial market crisis starting in 2008 or for economies in transition. For these reasons, business tendency surveys (BTSs) are conducted in an increasing number of countries mostly on a monthly basis. While in the late 1960s there were about 30 indicators from such surveys, which were available only for the main OECD countries, today, the number of these indicators exceeds 300, which cover more than fifty countries (Nardo 2003, p 645) including also economies in transition like Brazil, China, India and Russia (Nilsson and Brunet 2006).

For Germany – as for other industrial countries – there is a large body of empirical evidence that models forecasting the industrial production (IP) or real GDP, which include BTSs, outperform forecasts of univariate time series models (Hüfner and Schröder 2002, Benner and Meier 2004, Kholodilin and Siliverstovs 2006, Vogt 2007). Similar evidence can be found for other OECD countries (Bodo *et al.* 2000, Carnazza and Parigi 2003, Ozyldrim *et al.* 2010). For Russia, in contrast, there is less evidence on the usefulness of BTSs in forecasting IP or other aggregate measures of output (one of the exceptions is presented by Nilsson and Brunet (2006)). Possible reasons for that may be a relatively short history of these surveys in the country and the fact that the expectations of Russian managers are 'poorly calibrated' in a sense that Russian managers systematically overestimate their business situation, resulting in overoptimistic forecasts (Aukutsionek and Belianin 2001).

We use recent BTS–data for Germany and Russia to compare the relative information content. For Germany we consider the business expectation indicator taken from the ifo business cycle survey. For Russia, a similar indicator of expected future production is provided by both the Institute for the Economy in Transition (IET) and the Federal State Statistics Service (Rosstat). Our focus is on comparing the contribution of these indicators in modelling dynamic adjustment and forecasting IP. The modelling framework used are VAR models for constructing impulse response functions and unconditional forecasts, while the naive forecast and AR(p) processes serve as benchmarks.

Using the VAR modelling framework, we can avoid certain problems like simultaneity bias. As the potential number of lags included based on information criteria (IC) is large, we compare full-specified VAR models with subset models obtained using a Genetic Algorithm (GA) (Alberto *et al.* 2010, Savin and Winker forthcoming) allowing for 'holes' in multivariate lag structures. Employing all lags up to a predefined order one can maximize the model fit. However, given only a few non zero coefficients in the true data generating process, the standard approach results in a very unwelcome overfitting feature. The problem is complicated by the fact that we need to account for a structural break and seasonal variation of indicators. By incorporating subset models we aim to reduce variance in the model forecasts and compare the resulting dynamics for the two countries.

The focus of this work is threefold. First, it presents an application of a VAR modelling framework with BTSs on the example of Russia as an economy in transition and its comparison with Germany. Since none of the aforementioned authors analyzed BTS–indicators on their forecasting performance during a recession period,[1] the second contribution of our study is to determine whether BTSs in both countries provide useful information before and during the recent crisis. A third contribution of this paper is an illustration of the more flexible subset selection strategy based on GA (and the resulting advantages) in comparison to the standard approach.[2]

The rest of this paper is structured as follows. Section 2 presents the database for both countries (highlighting some peculiarities of the time series obtained from BTSs) and the VAR modelling framework used. Results of model specification and forecasting performance with the standard (*'take all up to the p-th lag'*) approach are reported in Section 3. Section 4 presents the heuristic algorithm and compares its results with the standard procedure. The paper concludes with a summary of the main findings and an outlook to further research in Section 5.

2 Data and modelling framework

2.1 Data description

The statistical indicators used in the following analysis are obtained through BTSs conducted in Germany and Russia. In both countries these indicators reflect businessmen's judgements on developments experienced in the recent

[1] The only exception for Germany is presented by Drechsel and Scheufele (2010).

[2] The latter focus is seen to be important since there are only few studies (Winker 1995, Winker 2000, Kapetanios 2007) demonstrating the application of the lag selection approach with 'holes' on VAR models.

past, their assessment of the current situation and their expectations for the next few months regarding their own business (in both data sets main responders are directors and deputy directors). The questions relate amongst others to future tendency of production, expected inflow of new orders, and business climate perception. The surveys are collected on a monthly basis and are usually available before the end of the respective month.[3]

The BTSs in Germany are conducted by the ifo institute (Munich) since 1949. The sample includes about 7000 enterprizes with half of them representing the manufacturing sector. The response rate is around 85%. The ifo BTSs, e.g., business climate and business expectation, have been extensively studied in the literature demonstrating their ability to improve forecasts of the German IP, real GDP and some monetary indicators (e.g., Langenmantel (1999), Flaig (2003), Mittnik and Zadrozny (2005)).

For Russia we take the information collected both by the Institute for the Economy in Transition (IET) and the Federal State Statistics Service (Rosstat), combining them in one data set. This is due to the data availability one faces with regard to the BTSs. For IET the data is publicly available for the period 02/1999–09/2009,[4] whereas the data of Rosstat is available since 07/1997, but has a structural break for the entire year 2005.[5] Hence, for the period 02/1999-12/2005 we use the IET data and for 01/2006–09/2009 Rosstat data. The coverage of the IET survey is relatively wide: with 227 administrative regions and 307 towns. Data are sampled for 1200 enterprizes from all major industries with a response rate of about 65% (Tsukhlo 2005). The coverage of the Rosstat survey is even larger (over 4000 companies) from mining, processing and service industries from all Russian regions.[6]

Because of the limited data availability for Russia we consider the period from 02/1999 to 09/2009 (128 observations in total), leaving one or two years for an out–of–sample forecast analysis between 11/2006 and 09/2009. Thus, we can consider the forecasting performance of our models both prior and during the crisis.[7] By restricting our data set in this way, we avoid to take into account the potential structural breaks due to reunification in 1990 and introduction of the Euro in 1999 for Germany and the period of financial crisis in Russia in 1998.

Considering Russia and Germany allows us to compare the relative per-

[3] In Russia the data is published a week after the end of the month.

[4] Officially, these BTSs are collected since 03/1992 to date.

[5] Due to the change in classification systems no observations are available.

[6] See the Federal State Statistics Service's web site (http://www.gks.ru/).

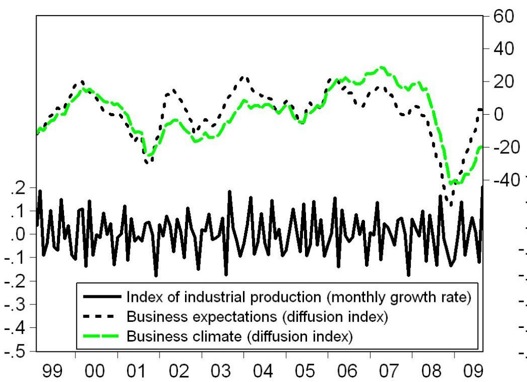
[7] Note that starting from the fourth quarter of 2008 GDP and IP growth rates for Russia and Germany become negative. For instance, in 2009 the actual IP growth rate in Germany reached the lowest level since the Second World War of approximately -16%.

formance of BTSs for a country in the process of transition, which collects this type of data only for around ten years, with an industrial country being one of the pioneers in using business survey data.

In both countries we deal with qualitative data. This means that enterprizes can characterize their situation as 'good', 'satisfactorily' or 'poor' and their business expectations as 'more favorable', 'unchanged' or 'more unfavorable'. For each sector, the balance of the percentage share of positive and negative answers is calculated. These balances are called 'index diffusions'. The composite indicators are then aggregated from the sector results.

The business expectation from the ifo survey refers to a planning horizon of six months according to the questionnaire. The indicator of future production for Russia (both IET and Rosstat) is calculated in a similar way, but with a time horizon of only three months. Obviously, this difference in the time horizon considered might result in differences of the dynamic adjustment in a VAR model setting (see Section 3).

We will also refer to the business climate indicator in our analysis. It has to be noted that this is calculated as an arithmetic mean of the results (balances) for the three questions related to production, total demand and finished goods stocks tendencies (with negative weight) for Russia, while a geometric mean of the present and future business situation is used for Germany (for further details see OECD (2003)). Thus, in contrast to the business expectations, these indicators can be referred to as composite leading indicators (see among others, Marcellino (2006) and Goldrian (2007)). Figure 1 shows the monthly IP growth rates (left axis) together with the business expectations and business climate perception (right axis) for Germany (left panel) and Russia (right panel), respectively.



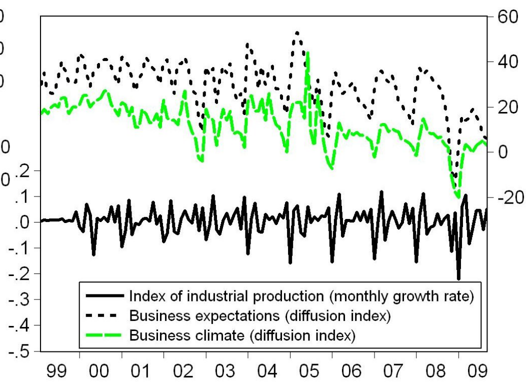


Figure 1: BTSs and IP monthly growth rates for Germany and Russia

An important difference between the Russian and German data is found with regard to the treatment of seasonal variation. While ifo asks its correspondents to provide seasonally adjusted estimates, both IET and Rosstat do not require this correction. As a result, the BTS indicators in Russia have a strong seasonal component (see Figure 1, right panel, dashed lines). Furthermore, the Figure exhibits that the seasonal pattern changes over time and becomes more pronounced from 2004 onward. Interestingly, if we consider data collected by the OECD available only up to 06/2007 (meanwhile the OECD statistical service terminated the data collection for Russia), a similar seasonal pattern is also observed for the period from 1995 to 2004. So far, we are not able to explain this surprising change in the seasonal pattern.[8]

Because the data on the Russian BTSs is not seasonally adjusted, we will also consider the IP indices for both countries as not seasonally adjusted. First, it will provide a more consistent basis for the comparison between the two countries. Second, instead of using some a priori methods of seasonal adjustment, which might have unwelcome side effects, e.g., the problem of spurious regression (for more details, see among others Meyer and Winker (2005)), introducing seasonal dummies in the model allows also to analyze this component and its effect on the dynamic adjustment process.

In addition, due to the necessity to merge two data sets for Russia, we account for a potential structural break in our VAR models for Russian data and for this reason add a shift dummy consisting of zeros until 12/2005 and ones starting from 01/2006 onwards. Furthermore, we include its first difference and lagged values of the first difference (the same number of lags as for the corresponding VAR model) to capture the dynamics given in the models. Also, interaction terms between the seasonal dummies and the shift dummy are included in the models for Russia. Thus, one provides a more flexible model accounting for the change in the seasonal structure for Russian BTSs.[9]

As for IP indices, for Germany, we use the data provided by the Federal Statistical Office with the base year 2005. The index accounts for under 80% of total production in the manufacturing sector and covers enterprizes also in former East Germany.[10] For Russia we consider the Rosstat data with the base year 2002 and also a broad coverage of firms and major industries.

Using monthly IP index (instead of quarterly GDP) has certain advantages. First, available at monthly frequency IP requires no aggregation of the

[8]Our exchange with IET regarding this finding did not result in a satisfying explanation. Nevertheless, it will have to be taken into account for the empirical model.

[9]Significance of the interaction terms for the corresponding VAR models is confirmed via the F-test at the 1% significance level.

[10]See the Federal Statistical Office's web site (http://www.destatis.de).

BTSs and, hence, no loss of information. Second, the number of observations for IP is by all means higher increasing the degrees of freedom.

However, it should be noted that for Germany IP covers less than 50% of total economic activity (Kholodilin and Siliverstovs 2006, p 240). For Russia this figure is even closer to 30% (Domsch and Lidokhover 2007). Moreover, in both countries the service sector becomes increasingly relevant, which might exhibit a different dynamic in comparison to IP. In contrast, as it was mentioned, the BTS indices cover all main industries of the two economies, including IP, services and trade (Tsukhlo 2005). Thus, we acknowledge some differences in the structure of the IP index we forecast and the BTSs we use.

2.2 Model specification

Given that no a priori information is available on the mutual dynamic adjustment of IP and leading indicators, a vector autoregressive (VAR) model is chosen as a reduced form modelling framework:

$$\begin{aligned} Y_t &= \sum_{i=1}^{p} \beta_{1,i} Y_{t-i} + \sum_{j=1}^{q} \gamma_{1,j} E_{t-j} + \sum_{d=1}^{12} \delta_{1,d} S_{d,t} + \varepsilon_{1,t} \\ E_t &= \sum_{i=1}^{p} \beta_{2,i} Y_{t-i} + \sum_{j=1}^{q} \gamma_{2,j} E_{t-j} + \sum_{d=1}^{12} \delta_{2,d} S_{d,t} + \varepsilon_{2,t}, \end{aligned} \tag{1}$$

where Y stands for the IP growth rate, E for the BTS indicator and S_d for seasonal dummies. Our focus is on identifying the dynamic dependence between the variables of interest and whether information from BTSs can improve forecast accuracy for IP. When evaluating the forecasting performance, we will consider a naive forecast (NF) as one benchmark. The NF forecasts are modeled as mean average values of the IP growth rates over the respective estimation period, i.e., the h–step–ahead–forecast $\widehat{Y}_{t+h}$ made at time t is a mean average of the variable Y in the period between 1 and t:

$$\widehat{Y}_{t+h} = \frac{1}{t} \sum_{\eta=1}^{t} Y_\eta. \tag{2}$$

Thus, this benchmark might not be hard to beat, but allows one to compare goodness of our models prior and during the recession. This is due to similar performance of NF before and during the recession, which is not the case for, e.g., a forecast based on the last observation Y_t. Alternatively, we consider

AR(p)-processes (against which our models have to compete) to analyze the contribution of the BTSs in the VAR models.

In a first step of the model specification, one has to consider possible nonstationarity of the time series. Thereby, it has to be taken into account that leading indicators refer to future changes in output. The variables of interest are the growth rate of output, approximated by the monthly IP growth rate, and the BTSs. In Appendix 6.1, we present results of unit root tests. Although the results are not perfectly conclusive, they allow the assumption that variables behave as stationary time series.

The lag length selection for the VAR models is based on the Schwarz (SIC) and the Akaike information criteria (AIC). The latter IC is known to be asymptotically inconsistent (Shibata 1976), but penalizes free parameters less strongly than SIC.[11]

2.3 Goodness of forecasts

To assess the quality of forecasts we will consider two measures. First, the root mean squared forecast error (*RMSFE*)

$$\mathit{RMSFE} = \sqrt{\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} e_t^2}, \tag{3}$$

where e_t is the difference between the actual IP growth rate (y_t) and its forecast ($\widehat{y}_t$); T_1 and T_2 indicate the first and the last period in the out–of–sample forecast, respectively. The model with the smallest *RMSFE* is the best one.

The second accuracy measure is 'Theil's U' (TU) (Theil 1966):

$$TU = \sqrt{\sum_{t=T_1}^{T_2} e_t^2} \Bigg/ \sqrt{\sum_{t=T_1}^{T_2} u_t^2}. \tag{4}$$

This measure can be interpreted as the ratio of the *RMSFE* of the analyzed model as compared to that of a reference model with forecast errors u_t.

As it was mentioned in Section 1 we build a few series of out–of–sample forecasts (for one- and two–year periods) for the IP growth rates for both countries using the VAR models specified (see Sections 3-4). To investigate dynamics of the prediction errors, the so called 'rolling windows' method is used (for more information see, e.g., Swansan (1998)). Doing this we either fix the estimation period, so that its length is the same for different overlapping

[11] The latter fact might be considered as an advantage for small data samples.

forecasting samples - fixed estimation windows (FEW), or let it increase always starting with 02/1999 - increasing estimation windows (IEW). We specify and estimate models for the forecasting exercise based on evolving estimation samples, while the analysis of stationarity, causality and impulse response functions are based on the full sample. Due to the limited number of observations after the structural break in Russia (01/2006), we include interaction terms only for a part of the simulations (forecasting samples) performing other forecasts without these flexible seasonal components.[12]

3 Results of the standard VAR model selection

According to the standard approach to the model selection problem (see, e.g., Lütkepohl (2007)), one compares different lag lengths for the VAR model choosing the one with the best balance between the model fit ($\widehat{\Sigma}$) and the total number of parameters to be estimated ($k = m^2p + md$).[13] Thereby, in this Section we impose the standard restriction that all lags up to a predefined order will be included (*'take all up to the p-th lag'* approach) and that this lag length is the same for all variables in all equations ($p = q$ in the notation of equation (1)). Due to the monthly data frequency, the maximum lag length (p_{max}) is set to 13. The results are provided in Table 8 in Appendix 6.2.

In order to test for causality between the series we apply the Granger test on the VAR models with the IP growth rate and business survey indicators as endogenous variables. Furthermore, we analyze the impulse response functions of the VAR models to obtain additional information on the dynamic adjustment path (these results are available on request). In short, according to the qualitative analysis we can expect the BTSs to provide additional information for forecasting IP in Germany, but not (or only to a minor extent) in Russia. Possible reasons for this could be the 'poor calibration' of Russian BTSs, their shorter expectation horizon and/or a higher estimation variance caused by the limited data quality (the change in the seasonal pattern mentioned in Section 2). In addition, one can find that the business expectations have a stronger explanatory power as compared to the business climate for the future IP in Germany,[14] whereas no evident difference between the BTSs is found for Russia.

[12]Due to the same reason we modify the size of the fixed estimation samples for simulations with and without the interaction terms (101 and 93 observations, respectively).

[13]$m = 2$ is the number of endogenous variables, p is the number of lags and $d = 12$ and 27 is the number of exogenous variables for Germany and Russia, respectively.

[14]This finding is intuitively clear: in the period of six months firms can adjust their production to respective changes, while their reaction to current shocks is less flexible.

In the following, resulting VAR models are compared in regard to their out–of–sample performance using static and dynamic forecast approaches. While a static forecast is a 1–step–ahead approach that employs only actual observations in (1), the dynamic approach is an h–step–ahead procedure that also employs predicted values for IP and BTSs.

3.1 Static forecasts

First, consider results with IEW for an evaluation period of one and two years (Table 1). One can see that the German IP is forecasted best by the VAR model with business expectations specified according to SIC, while the VAR with the business climate based on AIC is the second-best choice. Both of these models have an *RMSFE* smaller than the NF forecasts (benchmark). Similarly, the forecasts built for the Russian IP based on the VAR models specified by SIC are more accurate than the ones based on AIC. One can also see that the improvement is of a similar scale as in Germany, although in absolute values the forecast errors for Russia are even smaller.[15]

Table 1: Forecasting performance of the built VAR models

Model specification			Germany	Russia	Germany	Russia
			10/2007–09/2009		10/2008–09/2009	
RMSFE	NF forecast		10.0544	7.7724	10.2511	8.9200
	AR forecast	AIC	6.3691	4.6519	8.0974	6.1679
		SIC	6.3898	4.7360	8.2208	6.3296
	Business expectations	AIC	5.5188	5.3519	6.9972	6.2860
		SIC	4.9840	3.8418	6.0813	5.0850
	Business climate	AIC	5.3051	4.0459	6.6216	5.4377
		SIC	5.4239	3.8520	6.7384	5.1057
TU in regard to NF	Business expectations	AIC	0.5489	0.6886	0.6826	0.7047
		SIC	0.4957	0.4943	0.5932	0.5701
	Business climate	AIC	0.5276	0.5205	0.6459	0.6096
		SIC	0.5395	0.4956	0.6573	0.5724

To make sure that the produced forecasts (including the ones of respective

[15] Since in Table 1 forecasting results for two different indices are provided, the latter finding does not imply that the forecast for Russia is better.

AR(p)-processes) are not biased we test an equation of the following form:[16]

$$Y_t = \alpha + \beta\widehat{Y}_t + \nu_t. \tag{5}$$

In particular, we test whether α and β do not differ significantly from 0 and 1, respectively. For this purpose we consider t-tests for separate coefficients and p-values of the Wald test as a joint test. Results for the period 10/2008–09/2009 with IEW are presented in Table 9 in Appendix 6.3 and provide some evidence that the respective hypotheses can not be rejected.[17]

Next, to analyze goodness of the resulting forecasts for other periods, consider TU in regard to the NF forecast for overlapping windows within 11/2006–09/2009. Results for SIC[18] with IEW are given in Figure 2 for one-year period (for two–year period and FEW see Appendix 6.4). On the X-axis the forecasted period is given, while the Y-axis indicates TU values.[19] While prior to the end of 2008 (when the crisis has become pronounced also in IP indices) the TU values vary approximately between 0.3–0.5, the relative performance of the VAR models decreases dramatically afterwards.

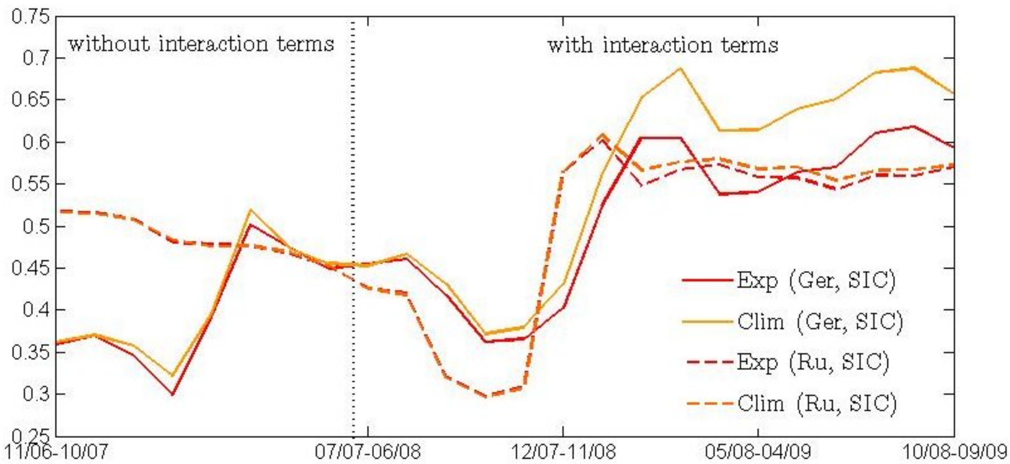


Figure 2: One–year forecast accuracy in relation to NF with IEW

The question, however, is whether the BTS indicators are useful in producing those forecasts (both prior and during the recession). To this end,

[16]This can be referred to as Mincer–Zarnowitz regression (Mincer and Zarnowitz 1969).
[17]Similar results are also obtained for other out–of–sample experiments.
[18]Results for AIC are available on request.
[19]In the following we denote in our VAR models Germany with 'Ger' and Russia with 'Ru', while 'Exp' and 'Clim' stand for business expectations and business climate.

the VAR models need to be compared with some better benchmark than NF. The AR(p)-processes are most suitable for this type of comparison, because the contribution of the BTSs as additional indicators is expected to become evident. Nonetheless, as one can see on Figure 3, the information content of BTSs is not clear, in particular for Russia. While for Germany the VAR models outperform the AR ones (with some minor exceptions), for Russia a high variance (that is even higher for AIC[20]) in the resulting performance is obtained. This is due to the variance in the lag length selected by IC resulting in a higher forecast error. Thus, the difference in the forecast accuracy (in relation to AR) for the period 12/2007–11/2008 is about 10% lower than for the same length period shifted one month before or two months later.

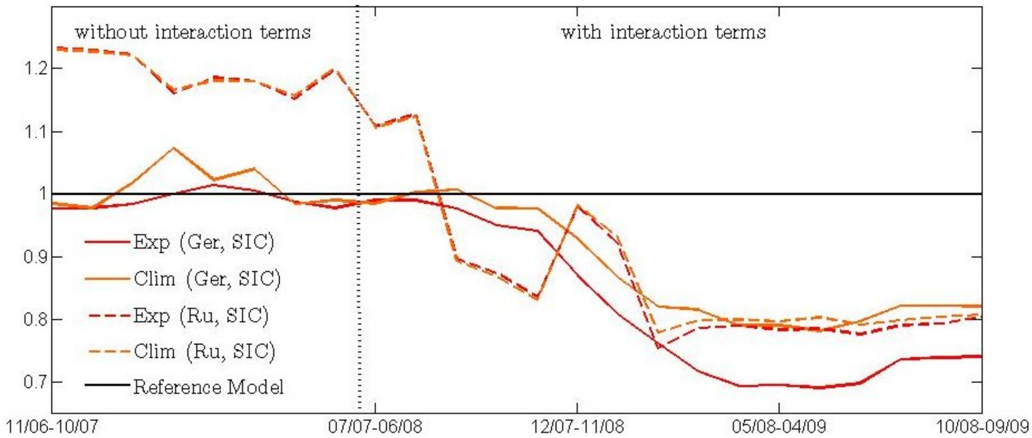


Figure 3: One–year forecast accuracy in relation to AR with IEW

Due to the high variance in the forecasting performance identified, the contribution of the BTSs is ambiguous. Nor their role prior and during the crisis is clear. Before one applies any significance tests on differences in out–of–sample performance between the VAR and AR models, we want to reduce the variance via an alternative model selection strategy allowing for 'holes' in the multivariate lag structures (see Section 4).

[20]These results are also available on request and for two–year period and FEW can be found in Appendix 6.5. For the two–year forecasts it is also evident that the interaction terms improve the forecasting accuracy of Russian VAR models.

3.2 Dynamic forecast

To compare the forecasts of our models with real data, we construct dynamic forecasts for the entire year 2009 (01/2009–12/2009) using the rest of the sample as estimation sample. The resulting forecasts for all three alternatives (AR models, VAR models with business expectations and VAR models with business climate) via SIC are presented in Figure 4.

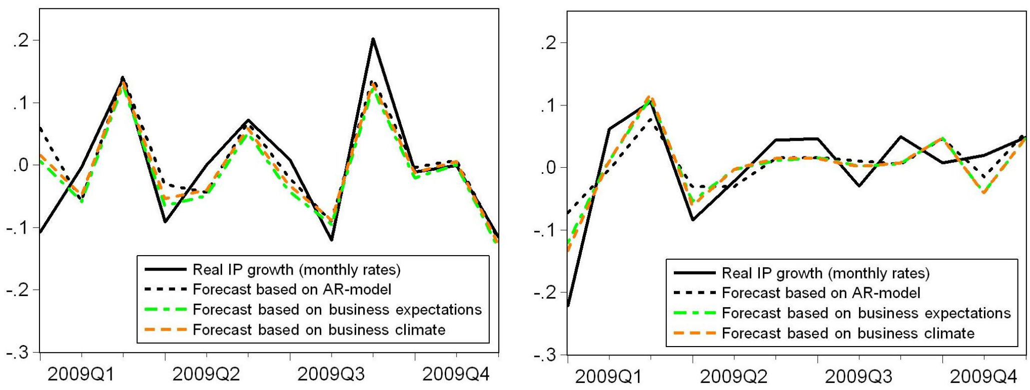


Figure 4: Prediction dynamics for the IP growth rate in 2009

The resulting VAR forecasts have minor differences predicting fairly similar IP growth dynamics. Estimating yearly IP growth rates (calculated as average of monthly year-to-year growth rates)[21] using the best unbiased model specification alternatives according to Table 1 (in particular, VAR with business expectations via SIC for both countries), one obtains the results given in Table 2.

Table 2: Prediction accuracy for the IP growth rate in 2009

	NF forecasts	AR forecasts	VAR forecasts	Actual rates
Germany	−14.45%	-6.13%	−17.51%	-15.99%
Russia	+0.11%	-4.16%	−5.68%	-10.84%

Both of the VAR–forecasts outperform the predictions of the corresponding benchmarks (AR and NF) in the two countries. It is also clear that the VAR–forecasted IP growth rate is much closer to the actual rate for Germany

[21]This methodology is also used by Thompson Reuters.

than for Russia. The large difference between the forecasted and the actual rates in the latter case might be due to the shorter expectation horizon of Russian BTSs. Thus, while the survey indicators in Russia can be accurate enough in the one–step–ahead forecast (see Figure 2), they are obviously inferior for the multi–step–ahead approach.

4 Heuristic algorithm and resulting models

In the standard bivariate VAR model with 13 lags and seasonal dummies considered in our study, the rich lag structure requires $k = 76$ parameters (52 lags and 24 dummies) to be estimated via the seemingly unrelated regressions (SUR) method. For Russia this number is even higher ($k = 106$) due to the shift dummies and the interaction terms included. Having only 128 observations in total (including a period for an out–of–sample forecast), this would leave only a very moderate number of degrees of freedom left. If only a small portion of these parameters has non–zero coefficients in the real DGP, this results in a high estimation variance and forecast errors. This motivates us to use an alternative subset selection strategy allowing for 'holes' in multivariate lag structures, which is seen to be a more flexible form to represent the real DGP.

Due to the discrete search space and $2^{m^2 p}$ potential submodels we can neither apply standard optimization methods nor enumerate all possible solutions for high or even medium m and p (as, e.g., two and 13) comparing them based on IC. This cannot be done even using efficient branch–and–bound algorithms as described in Gatu *et al.* (2008).

In the last decade a large number of studies has been devoted to subset VAR selection methods. Among them are a sequential bottom–up (top–down) inclusion (deletion) of individual regressors (Lütkepohl 2007) and its analog based on respective coefficient's t-statistics (Brüggemann and Lütkepohl 2001). Another option is to set certain model prior probabilities shrinking the parameter search space. The resulting model averaging strategies are well discussed in Kapetanios *et al.* (2008). However, these methods investigate only a certain portion of all potential submodels. An alternative is presented by the least absolute shrinkage and selection operator (Lasso) that was initially suggested by Tibshirani (1996) as a constrained version of the ordinary least squares estimator, but later is also applied to VAR models (Hsu *et al.* 2007). Nevertheless, due to the shrinkage parameter the method exhibits a substantial estimation bias (Hsu *et al.* 2007, Savin 2010).

To overcome a priori restrictions on the search space that might result in local optima, one can take advantage of optimization heuristics that mimic natural evolution processes (for an overview of the methods as well as for guidelines of their implementation see Gilli and Winker (2009)).

In our study we apply Genetic Algorithms (GA), which are a population–based heuristic that operate on a set of solutions (population).[22] The members in the GA population (chromosomes) are represented as bit strings, in which each position (gene) has two possible values (1 and 0), corresponding to included and not included regressors, respectively. In each generation GA replace parts of a population with new chromosomes (children) aimed to represent better solutions for a given problem.

Applying GA on our model selection problem we consider different scenarios. At first, GA are used with IC as an objective function with the search space consisting either of all lags to be potentially included (2^{m^2p}) or all lags and exogenous variables (constant, seasonal dummies plus shift dummies and interaction terms for Russia). Based on IC, GA forecasts exhibit a strong overfitting feature. Explaining the training data set better than the standard model selection approach, the heuristic demonstrates a lower forecasting performance. To overcome this overfitting, different constants (greater than one) as a multiplicator for the IC's penalty term were tested.[23] However, no stable improvement is identified. This can be due to the short sample size of our data limiting both the training data set and the evaluation period. Moreover, testing the forecasting performance during the recession presents an additional challenge for the subset selection strategy.

Alternatively, certain 'weights' increasing the penalties for lags of a higher order were examined. In particular, some improvement for AIC is identified by penalizing lags in a geometric progression ($\omega_p = 2^p/2$).[24] Thus, similarly to the standard approach, the p-th lag (instead of $p-1$) is penalized by more than $f(1,T)$. In addition, since the BTSs have a planning horizon of three to six months, one would expect these lags to be more informative than the ones of a higher order. Nevertheless, the choice of weights is rather ad hoc; and for SIC, no stable improvement is identified for the models with BTSs for Germany and none for the ones for Russia.

Realizing the limitations of IC in our study, in the following we consider *RMSFE* as an objective criterion for GA . In particular, selecting a subset of lags, one estimates its out–of–sample performance on a 'pre-testing' period (which is twice as long as the actual out-of–sample period and precedes it) and uses the respective forecast error as a measure of goodness for a can-

[22] In Savin and Winker (forthcoming) this method is compared with a different heuristic (Threshold Accepting) and is found as more efficient for discrete subset selection problems.

[23] These constants are intended to improve the convergence speed of IC (Winker 2000).

[24] Thus, IC penalties depend also on these weights ($f(k, \omega_p, T)$) and, e.g., for AIC are as follows: $2\sum_{p=1}^{13} k\omega_p/T$.

didate solution. The intuition for this is that one can assume a model to demonstrate good out–of–sample performance, if the model well predicts the foregoing period. To this end, we concentrate on the selection within lags, but not within exogenous variables. First, one has not enough observations to estimate *RMSFE* on the 'pre-testing' period for interaction terms. Second, one avoids the risk to remove the controls due to their lower information content in comparison to lags.

The GA algorithm used in this study is very similar to the one described in Savin and Winker (forthcoming) in detail. Some difference is that due to the smaller discrete search space we use a population of $p = 100$ solutions and preserve only one elitist chromosome (K^*). Besides, the number of generations $G = 100$ (computational resources) is sufficient for one restart.[25]

Comparing the full-specified models with subset models obtained using GA we consider both their in–sample and out–of–sample fit. While the in–sample analysis is made for the whole sample 02/1999–09/2009, the out-of-sample exercise is conducted (as described before) via rolling windows either of one or two years covering IP before and during the recession. In addition, dynamic forecasts based on the resulting GA subsets are constructed.

4.1 In–sample fit

Two alternative model selection strategies, *'take all up to p-th lag'* and GA, are compared by means of the number of free parameters to be estimated (k)[26] and adjusted R^2 (for the IP growth rate's equation) (Table 3).

As one can see, for Germany GA include an intermediate number of parameters between AIC and SIC demonstrating a slightly lower in–sample fit than the standard lag selection strategy. A similar finding on the number of parameters selected is made for Russia, whereas the models' goodness–of–fit is very close or even slightly outperforms the standard strategy. Thus, omitting certain lags in the resulting GA subsets does not produce any substantial loss in the in–sample fit.

Now we are interested whether GA are able to reduce the variance in our out–of–sample forecasts identified in Section 3.

4.2 Static forecasts and significance testing

The resulting out–of–sample fit is our main focus (in comparison to the in-sample one) due to two reasons. First, a good out–of–sample performance

[25]The corresponding CPU time for 100 generations using Matlab 7.10 on a Pentium IV 2.67 GHz requires approximately 940s.

[26]This includes both endogenous and exogenous variables.

Table 3: In–sample fit of different models

Model specification			'all up to p-th lag'		GA	
			k	R^2_{adj}	k	R^2_{adj}
Germany	Business expectations	AIC	44	0.8692	34	0.8401
		SIC	32	0.8661		
	Business climate	AIC	40	0.8694	34	0.8462
		SIC	32	0.8539		
Russia	Business expectations	AIC	64	0.7963	59	0.7925
		SIC	56	0.7853		
	Business climate	AIC	64	0.7926	58	0.7935
		SIC	56	0.7843		

gives a vital evidence in support of a particular model specification. Thus, a well specified model captures the given data set, but does not overfit it with too many predictors. Otherwise, forecasts based on the parameter estimates can be poor due to the estimates' variance. Second, for those interested in forecasting real production output it is more important that a specified model can produce good forecasts than fit to a certain data set.

To compare the forecasting performance of our VAR models specified based on the two alternative strategies, consider the resulting TUs (with NF as a benchmark) for one– and two–year out–of–sample periods (Table 4).

Table 4: Forecasting performance of the built VAR models in comparison

Model specification			Standard	GA	Standard	GA
			10/2007–09/2009		10/2008–09/2009	
Germany	Business expectations	AIC	0.5489	0.5084	0.6826	0.5973
		SIC	0.4957		0.5932	
	Business climate	AIC	0.5276	0.5065	0.6459	0.6182
		SIC	0.5395		0.6573	
Russia	Business expectations	AIC	0.6886	0.4721	0.7047	0.5544
		SIC	0.4943		0.5701	
	Business climate	AIC	0.5205	0.4832	0.6096	0.5366
		SIC	0.4956		0.5724	

One can see that the GA forecasts based on RMSFE both for one- and two–year periods always outperform the forecasts based on the standard model selection approach via AIC. Moreover, they either outperform the standard strategy based on SIC (e.g., forecasts for Russia and business climate for Germany) or do not significantly vary in their performance (e.g., forecasts based on business expectations for Germany). This also holds true for other overlapping windows within the period 11/2006–09/2009 (see Figure 5).[27] Important to note is that, in contrast to the standard model selection, GA allow to reduce variation in the forecast accuracy. Thus, one avoids certain spikes in the TU dynamics as, e.g., for the period 04/2007–03/2008 for Germany or 01/2008–12/2008 for Russia.

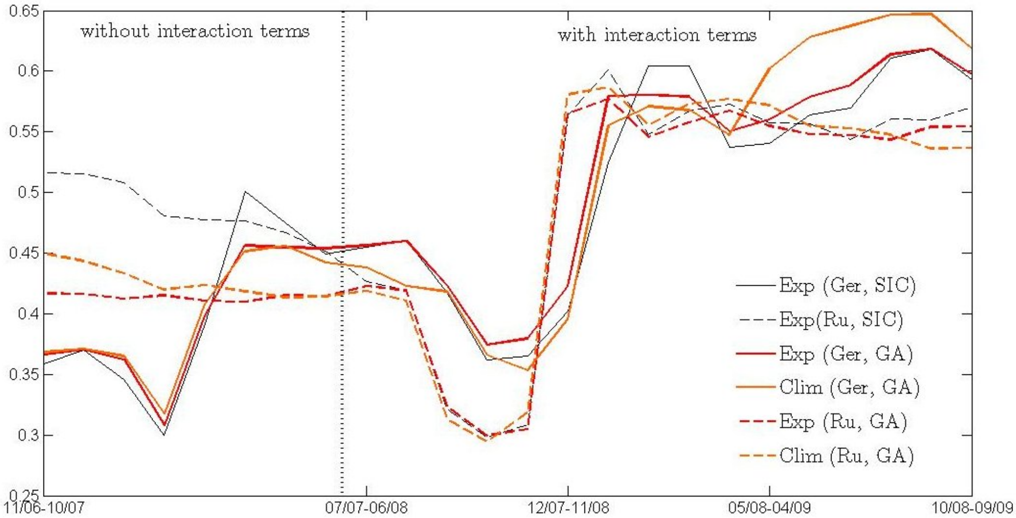


Figure 5: One–year forecast accuracy in relation to NF with IEW (GA)

Similarly, one can consider the resulting forecasting accuracy of our models in relation to the AR–benchmark (Figure 6).[28] Obviously, here we also obtain less variance in forecasting performance of our models. An important difference to the results of the standard strategy is that there is a much smaller difference between the AR and VAR forecasts for Russia.

Since the obtained TUs alone do not provide any evidence on statistical significance of the differences in forecasting performance between the AR and VAR models, we apply the Giacomini and White (2006) test (GW test

[27] Results of the standard model selection strategy for business expectations via SIC (the best alternative identified in Section 3) are plotted for comparative reasons.

[28] In this case both AR and VAR model structures are optimized via GA.

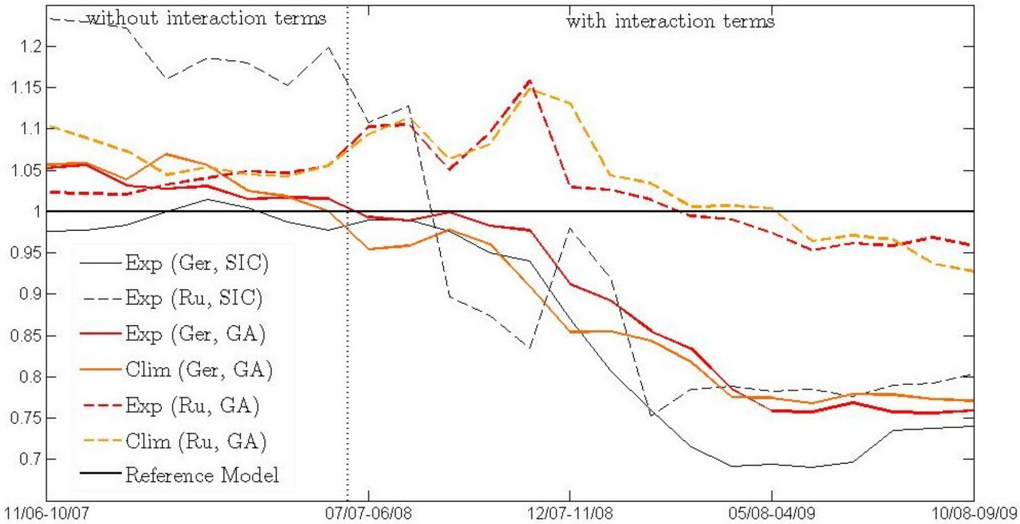


Figure 6: One–year forecast accuracy in relation to AR with IEW (GA)

henceforth) of conditional predictive ability.[29] The GW test is most suitable for our study for several reasons. First, it takes into account the effect of estimation uncertainty (West 1996) on the relative forecasting performance. Second, it can evaluate forecasts both for nested and nonnested models, which is particularly relevant for our case with an AR model being potentially nested within a VAR model for certain rolling windows. Third, presenting a generalized approach to existing studies in this field (as, e.g., Diebold and Mariano (1995)), the test allows for structural shifts at unknown dates (note the shift in the BTSs for Russia) and for both FEW and IEW approaches.

Defining $\Delta L_{m,t+1}$ as the loss difference of our VAR model forecast (e^2_{t+1}) and the benchmark (u^2_{t+1}) for a 1–step–ahead forecast, the null-hypothesis of the GW test is $E\left[h_t \Delta L_{m,t+1} | F_t\right] = 0$, where F_t is an information set available at time t. Then, the Wald-type test statistic is as follows:

$$GW^1_{m,n} = n \left(n^{-1} \sum_{t=T_1}^{T_2-1} h_t \Delta L_{m,t+1} \right)' \hat{\Omega}_n^{-1} \left(n^{-1} \sum_{t=T_1}^{T_2-1} h_t \Delta L_{m,t+1} \right) \overset{\alpha}{\sim} \chi^2_q, \quad (6)$$

where $n = T_2 - T_1 - 1 + 1$ is the number of out–of–sample forecasts, m

[29]So far, available literature for Germany (Kholodilin and Siliverstovs 2006, Drechsel and Scheufele 2010) indicates that the difference is significant only to some extent.

is the estimation sample, $\hat{\Omega}_n = n^{-1}\sum_{t=T_1}^{T_2-1}(h_t\Delta L_{m,t+1}) \times (h_t\Delta L_{m,t+1})'$ is a consistent variance estimator and the test function h_t is set to $[1\ \Delta L_{m,t}]$.[30]

To test the significance of our survey indicators before and during the recession, we split the sample 11/2006–09/2009 into a pre–crisis and a crisis period, where the latter starts from 11/2008. This choice is based on the dramatic change in the forecasting accuracy of the VAR models in regard to NF observed for both standard and heuristic model selection strategies. To this end, we employ the IEW approach constructing 1–step–ahead forecasts for the entire pre–crisis ('prior') and crisis ('during') subsamples. Results of the pairwise test of equal conditional predictive ability are given in Table 5.

Table 5: Conditional predictive ability test results[a]

Model specification			'all up to p-th lag'		GA	
			prior	during	prior	during
Germany	Business expectations	AIC	0.8810	0.8490	1.0010	0.7609*
		SIC	0.9530	0.7415		
	Business climate	AIC	0.8677	0.8034*	1.0240	0.7727**
		SIC	0.9788	0.8204		
Russia	Business expectations	AIC	1.1599	1.0241**	1.1078**	0.9270
		SIC	1.1755*	0.7989		
	Business climate	AIC	1.2335*	0.8791	1.1202*	0.9288
		SIC	1.1714*	0.8021*		

[a]The entries are the relative *RMSFE*s of corresponding VAR models with respect to AR benchmark with **,* denoting significance level of the Giacomini and White (2006) test at the 5 and 10%, respectively.

While for Germany constructed VAR models demonstrate significantly better forecasting performance only in the latter subsample[31] (during the recession), resulting VAR models for Russia significantly (but negatively) differ from the benchmark in the former one (prior to the crisis). This finding demonstrates a marked difference between the results obtained via the heuristic strategy (right panel) and the '*all up to p-th lag*' method (left panel). Thus, employing the standard approach one could obtain evidence confirming either significantly superior or inferior performance of the VAR models during the recession depending on the BTS and IC chosen. However, since the

[30]Setting the conditioning set in this way, we expect past loss differences to predict their future dynamics or, in other words, to indicate the more accurate model (Giacomini and White 2006)[p. 1551].

[31]Note that this result is obtained for the short data sample taken. Employing a larger sample would provide different results (see, e.g., Drechsel and Scheufele (2010)).

heuristic approach consistently demonstrates better forecasting performance (Figures 5 and 6), its results can be considered as more justified alternative.

4.3 Dynamic forecast

Employing the subset models obtained via GA for AR and VAR (with business expectations) models in Russia and Germany, we can construct dynamic forecasts (as in Section 3.2) for the year 2009 (Table 6).

Table 6: Prediction accuracy for the IP growth rate in 2009 (with GA)

	NF forecasts	AR forecasts (via GA)	VAR forecasts (via GA)	Actual rates
Germany	−14.45%	−6.13%	−16.59%	-15.99%
Russia	+0.11%	−7.52%	−7.42%	-10.84%

As one might expect based on Section 4.2, forecasts constructed for IP in Russia via GA subsets demonstrate some better approximation of the actual growth rates. This holds for both AR and VAR models, whereas their difference is very small. The latter finding is consistent with the *GW* test results. For Germany, in contrast, GA do not make any difference for the AR model yielding the same prediction. The VAR forecast, however, is slightly improved in comparison to the standard strategy.

5 Conclusion and outlook

While there are many studies providing evidence that leading indicators improve univariate time series models forecasting real output in OECD countries, this is by far not the case for developing countries like Russia. Moreover, there is a lack of research examining the indicators' performance during the crisis. Therefore, our study contributes to the existing literature in two ways: by analyzing BTSs with regard to their performance in predicting IP in Russia and Germany both prior and during the recent economic recession.

For this purpose, we specify VAR models accounting for a structural break and seasonal variation in the BTSs. Due to the limited data sample available and the large number of parameters to be estimated, high variance in estimation errors and, hence, resulting forecasts is obtained. To reduce this problem, an alternative subset selection strategy (GA) allowing for 'holes' in the lag structures is implemented. Since the data set is too small for IC to identify the 'correct' model, RMSFE for a pre-testing sample is taken as an objective criterion. The heuristic strategy is assessed in its in–sample and out–of–sample fit demonstrating superior forecasting performance.

Constructing 1–step–ahead forecasts for rolling windows of one and two years length with both IEW and FEW, we demonstrate the VAR models to outperform the AR benchmark during the crisis in Germany. In contrast, there is no significant evidence on this in Russia. These results are confirmed by the conditional predictive ability test and are consistent with both Granger causality test and impulse response function analysis.

Also, we show that these results can be obtained employing GA, but not the standard '*all up to p-th lag*' strategy. This seem to be a result of the very limited data sample, where a comprehensive model selection strategy is required to optimize multivariate lag structures sparing some degrees of freedom and, hence, reducing variance in resulting forecasts.

In conclusion, we argue that the heuristic methods are an effective tool of model selection also for short time series models. For future study remains their comparison with other alternatives as, e.g., bayesian model averaging or shrinkage estimators. Furthermore, in a few years it is worth to repeat the experiment on a larger data set (for Russia) to test whether the BTSs would provide any better results.

Acknowledgements Thanks are due to participants of the 5th ifo workshop on macroeconomics and business cycles in Dresden and of the 3rd international conference of the ERCIM in London for their constructive comments that helped to improve the paper. All shortcomings are our responsibility.

Financial support from the German Academic Exchange Service (DAAD) and the EU Commission through MRTN-CT-2006-034270 COMISEF is gratefully acknowledged.

References

Alberto, I., A. Beamonte, P. Gargallo, P.M. Mateo and M. Salvador (2010). Variable selection in STAR models with neighbourhood effects using genetic algorithms. *Journal of Forecasting* **29**(8), 728–750.

Aukutsionek, S.P. and A.V. Belianin (2001). Quality of forecasts and business performance: A survey study of Russian managers. *Journal of Economic Psychology* **22**(5), 661–692.

Benner, J. and C.-P. Meier (2004). Prognosegüte alternativer Frühindikatoren für die Konjunktur in Deutschland. *Journal of Economics and Statistics* **224**(6), 639–652.

Bodo, G., R. Golinelli and G. Parigi (2000). Forecasting industrial production in the Euro area. *Empirical Economics* **25**(4), 541–561.

Box, G.E.P. and D.A. Pierce (1970). Distribution of residual autocorrelation in autoregressive moving average time series models. *Journal of the American Statistical Association* **65**(332), 1509–1526.

Brüggemann, R. and H. Lütkepohl (2001). Lag selection in subset VAR models with an application to a U.S. monetary system. In: *Econometric Studies – A Festschrift in Honour of Joachim Frohn* (R. Friedmann, L. Knüppel and H. Lütkepohl, Eds.). pp. 107–128. LIT. Münster.

Carnazza, P. and G. Parigi (2003). Tentative business confidence indicators for the Italian economy. *Journal of Forecasting* **22**(8), 587–602.

Chamberlin, G. (2007). Forecasting GDP using external data sources. *Economic & Labour Market Review* **1**(8), 18–23.

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

Dickey, D.A. and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**(366), 427–431.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**(3), 253–263.

Domsch, M.E. and T. Lidokhover (2007). The broader historical, social and economic context of the current situation in Russia. In: *Human Resource Management in Russia* (M.E. Domsch and T. Lidokhover, Eds.). pp. 1–22. Aldershot. Hampshire, UK.

Drechsel, K. and R. Scheufele (2010). Should we trust in leading indicators? Evidence from the recent recession. Technical Report 10-10. IWH Discussion Papers.

Flaig, G. (2003). Seasonal and cyclical properties of ifo business test variables. *Journal of Economics and Statistics* **223**(5), 556–570.

Gatu, C., E. J. Kontoghiorghes, M. Gilli and P. Winker (2008). An efficient branch-and-bound strategy for subset vector autoregressive model selection. *Journal of Economic Dynamics & Control* **32**, 1949–1963.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* **74**(6), 1545–1578.

Gilli, M. and P. Winker (2009). Heuristic optimization methods in econometrics. In: *Handbook of Computational Econometrics* (D. A. Belsley and E. Kontoghiorghes, Eds.). pp. 81–119. Wiley. Chichester.

Goggin, J. (2008). An analysis of the potential of the European Commission business and consumer surveys for macroeconomic forecasting. *ESRI Quarterly Economic Commentary* **4**, 46–67.

Goldrian, G. (2007). A leading indicator composed of survey data: appropriate construction and prognostic significance. In: *Handbook of Survey-Based Business Cycle Analysis* (G. Goldrian, Ed.). pp. 151–158. Edward Elgar. Cheltenham.

Hsu, N.-J., H.-L. Hung and Y.-M. Chang (2007). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis* **52**(7), 3645–3657.

Hüfner, F. and M. Schröder (2002). Forecasting German industrial production: An econometric comparison of ifo- and ZEW-business expectations. *Journal of Economics and Statistics* **222**(3), 316–336.

Kapetanios, G. (2007). Variable selection in regression models using non-standard optimization of information criteria. *Computational Statistics & Data Analysis* **52**(1), 4–15.

Kapetanios, G., V. Labhard and S. Price (2008). Forecasting using Bayesian and information-theoretic model averaging: An application to U.K. inflation. *Journal of Business & Economic Statistics* **26**(1), 33–41.

Kholodilin, K. A. and B. Siliverstovs (2006). On the forecasting properties of the alternative leading indicators for the German GDP: Recent evidence. *Journal of Economics and Statistics* **226**(3), 234–259.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**(1-3), 159–178.

Langenmantel, E. (1999). Das ifo-Geschäftsklima als Indikator für die Prognose des Bruttoinlandproduktes. *IFO Schnelldienst* **16/17**, 16–21.

Lanne, M., H. Lütkepohl and P. Saikkonen (2002). Comparison of unit root tests for time series with level shifts. *Journal of Time Series Analysis* **23**(3), 667–685.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis. Corr. 2nd printing.* Springer. Berlin, Heidelberg, New York.

Marcellino, M. (2006). Leading indicators. In: *Handbook of Economic Forecasting* (G. Elliott, C. Granger and A. Timmermann, Eds.). pp. 892–924. Elsevier. Amsterdam.

Meyer, M. and P. Winker (2005). Using HP filtered data for econometric analysis: some evidence from Monte Carlo simulations. *Allgemeines Statistisches Archiv* **89**(3), 303–320.

Mincer, J. A. and V. Zarnowitz (1969). The evaluation of economic forecasts. In: *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* (J. A. Mincer, Ed.). pp. 1–46. NBER. New York.

Mittnik, S. and P. Zadrozny (2005). Forecasting quarterly German GDP at monthly intervals using monthly ifo business conditions data. In: *Ifo Survey Data in Business Cycle Analysis and Monetary Policy Analysis* (J.-E. Sturm and T. Wollmershäuser, Eds.). pp. 19–49. Physica-Verlag. Heidelberg.

Nardo, M. (2003). The quantification of qualitative business surveys: a critical assessment. *Journal of Economic Surveys* **47**(2), 211–223.

Nilsson, R. and O. Brunet (2006). Composite leading indicators for major OECD non-member economies: Brazil, China, India, Indonesia, Russian Federation, South Africa. Technical Report 2006/1. OECD Statistics Working Papers.

OECD (2003). *Business Tendency Surveys: A Handbook.* OECD Organisation for Economic Co–operation and Development. http://www.oecd.org/dataoecd/29/61/31837055.pdf.

Ozyldrim, A., B. Schaitkin and V. Zarnowitz (2010). Business cycles in the Euro area defined with coincident economic indicators and predicted with leading economic indicators. *Journal of Forecasting* **29**(1-2), 6–28.

Savin, I. (2010). A comparative study of the lasso-type and heuristic model selection methods. Technical Report 42. COMISEF Working Papers.

Savin, I. and P. Winker (forthcoming). Heuristic optimization methods for dynamic panel data model selection. Application on the Russian innovative performance. *Computational Economics.*

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**(1), 117–126.

Swansan, N. R. (1998). Money and output viewed through a rolling window. *Journal of Monetary Economics* **41**(3), 455–474.

Theil, H. (1966). *Applied Economic Forecasting.* Rand McNally. Chicago.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* **58**(1), 267–288.

Tsukhlo, S. V. (2005). *The IET surveys, organisational problems and analysis opportunities (in Russian).* Institute for the Economy in Transition. http://www.iet.ru/files/persona/tsukhlo/council_24.03.05.pdf.

Vogt, G. (2007). The forecasting performance of ifo-indicators under real-time conditions. *Journal of Economics and Statistics* **227**(1), 87–101.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* **64**(5), 1067–1084.

Winker, P. (1995). Identification of multivariate AR-models by threshold accepting. *Computational Statistics & Data Analysis* **20**, 295–307.

Winker, P. (2000). Optimized multivariate lag structure selection. *Computational Economics* **16**, 87–103.

6 Appendix

6.1 Tests of nonstationarity

To test for possible nonstationarity of the variables we apply the Augmented Dickey-Fuller-test (ADF–test) (Dickey and Fuller 1979) to variables, where no structural shift is assumed, and an UR–test (see Lanne *et al.* (2002)) to the BTSs of Russia with structural breaks (UR–test).[32] The null–hypothesis (H_0) of both, the ADF– and the UR–tests, is that the variable under consideration has a unit root. The alternative hypothesis is that it is stationary. For all series the test is run with a constant and the number of lags included to be selected based on SIC with a maximum lag order of 12 months. The test statistics are summarized in the lower left entries in Table 7.

Table 7: Unit root test results

	Germany			Russia		
	lags	ADF–test	KPSS–test	lags	ADF/UR–test	KPSS–test
IP growth rate	12	1.76	0.24	11	−2.25	0.11
Business expectations	1	3.08**	0.18	1	−2.18	
Business climate	2	−2.47	0.13	0	−5.39***	

***/**/* denote significance at the 1, 5 and 10% level, respectively.

From a theoretical point of view it is expected that the IP growth rates and the BTSs are stationary. However, this assumption is not confirmed for IP growth rate[33] and two of the four survey indicators. These results might have been expected due to the low power of the test against H_0.

Starting with H_0 of stationary series and running the KPSS–test (Kwiatkowski *et al.* 1992) to the variables with no structural shifts, including again a constant and determining the bandwidth with the Newey–West procedure, results in no rejections of the hypothesis (see upper right entries in Table 7). As a result, based on the tests employed one still has at least one nonstationary variable (business expectations for Russia).

[32] The latter test is implemented using the JMulTi free statistical software package available at http://www.jmulti.de/.

[33] We believe that the main reason for this is the extraordinary drop of the IP growth rates at the end of 2008 due to the global economic crisis. In fact, considering the shorter sample until 09/2008 supports the stationarity assumption (results available on request).

6.2 Lag length selection

After a specific model specification is selected, we test for possible residual autocorrelation. To this end, we employ the multivariate extension of the test suggested by Box and Pierce (1970), namely the Ljung–Box portmanteau test with the maximum lag length of 13 (Table 8) with its p–values in parentheses.

Table 8: Lag length selection results

	SIC				AIC			
	Germany		Russia		Germany		Russia	
	lags	Test	lags	Test	lags	Test	lags	Test
Business expectations	2	58.2 (0.07)	1	62.8 (0.08)	5	45.0 (0.06)	3	53.5 (0.08)
Business climate	2	53.9 (0.15)	1	74.5 (0.01)	4	30.2 (0.74)	3	40.0 (0.52)

It becomes obvious that the models for the German data do not exhibit residual autocorrelation at the 5% significance level when choosing the lag length suggested by SIC. This also holds true for the Russian VAR model with the business expectations, but not for the business climate. This result might be due to changes in the seasonal pattern of the time series described in Section 2. A more detailed analysis of this feature is left for future research. For the current analysis we stick to the lag orders presented in Table 8. In contrast, selecting the lag length according to AIC one has to deal with larger models (as expected), but identifies no autocorrelation at the 5% level.

6.3 Forecasts' bias analysis

Table 9: Forecasts' bias analysis

Model specification		$\alpha = 0$		$\beta = 1$		$\alpha = 0$ and $\beta = 1$	
		t-statistics				p-values	
		Germany	Russia	Germany	Russia	Germany	Russia
SIC	AR model	-1.63	-1.52	-0.49	1.06	0.19	0.30
	Business expectations	-0.60	-0.66	-0.27	0.75	0.81	0.63
	Business climate	-0.97	-1.13	-0.45	0.88	0.55	0.43
AIC	AR model	-0.84	-1.52	-0.94	0.75	0.44	0.34
	Business expectations	-1.31	-1.50	-0.45	0.20	0.37	0.35
	Business climate	-1.66	-1.22	0.03	0.45	0.28	0.49

6.4 Forecast accuracy in relation to NF

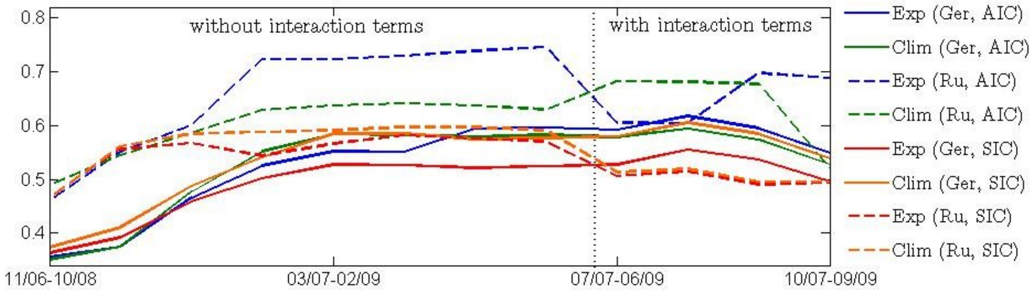


Figure 7: Two–year forecast accuracy in relation to NF with IEW

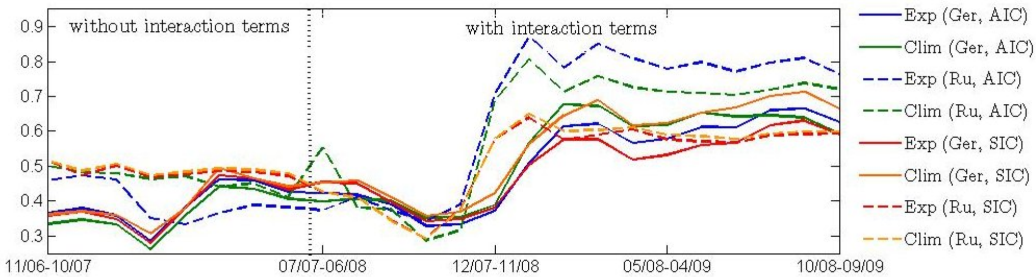


Figure 8: One–year forecast accuracy in relation to NF with FEW

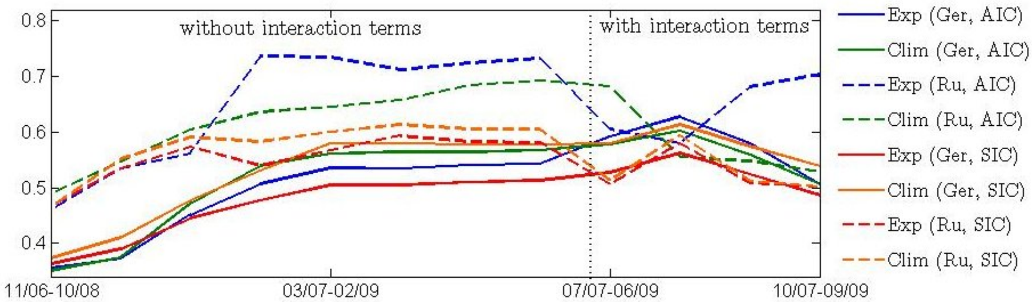


Figure 9: Two–year forecast accuracy in relation to NF with FEW

6.5 Forecast accuracy in relation to AR

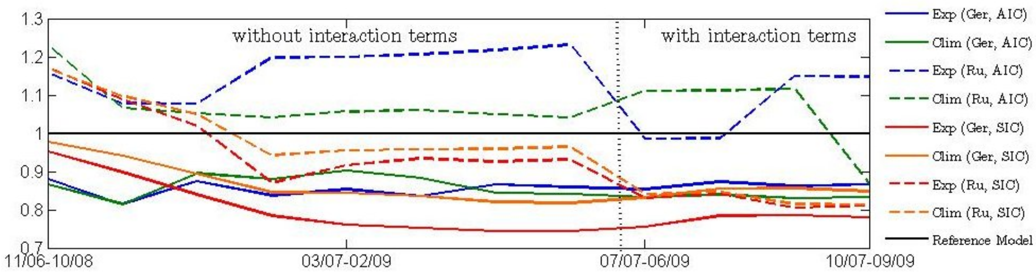


Figure 10: Two–year forecast accuracy in relation to AR with IEW

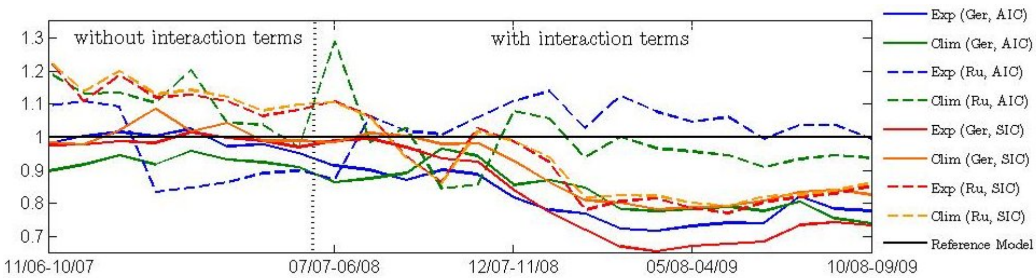


Figure 11: One–year forecast accuracy in relation to AR with FEW

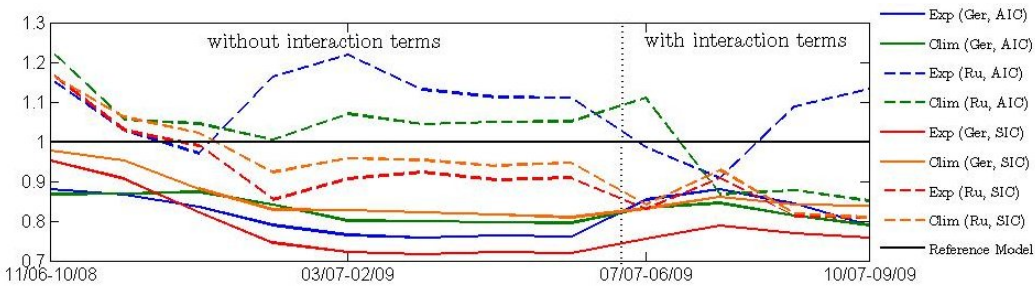


Figure 12: Two–year forecast accuracy in relation to AR with FEW