



OECD Journal

Journal of Business Cycle Measurement and Analysis

Editor-in-Chief: Michael Graff



OECD Journal

Journal of Business Cycle Measurement and Analysis

Editor-in-Chief: Michael Graff

Volume 2012/2

A Joint Publication of OECD and CIRET

The OECD Journal collection comprises six periodicals: *OECD Journal: Economic Studies*, *OECD Journal: Financial Market Trends*, *OECD Journal: Competition Law and Policy*, *OECD Journal: General Papers*, *OECD Journal on Budgeting*, *OECD Journal: Journal of Business Cycle Measurement and Analysis*. Each periodical within OECD Journal retains its own ISSN, volume and issue numbering. The volume numbering of each journal corresponds to the calendar year, e.g. Volume 2011. Volume 2011 of *Journal of Business Cycle Measurement and Analysis* contains two issues.



CIRET



KOF

Centre for International Research
on Economic Tendency Surveys

The Ciret Office is located at
KOF Swiss Economic Institute, ETH Zurich,
Switzerland

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Periodical: *OECD Journal: Journal of Business Cycle Measurement and Analysis*
ISSN 1995-2880 (print)
ISSN 1995-2899 (online)

<http://dx.doi.org/10.1787/jbcma-v2012-2-en>

Corrigenda to OECD publications may be found on line at: www.oecd.org/publishing/corrigenda.

© OECD 2013

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

Table of contents

Measuring capacity utilisation in the Italian manufacturing sector: A comparison between time series and survey estimates	
by Marco Malgarini and Antonio Paradiso	5
Nowcasting Irish GDP	
by Antonello D'Agostino, Kieran McQuinn and Derry O'Brien.	21
Day-of-the-week effect in Consumer Confidence Index: The case of Turkey	
by Sadullah Çelik and Hüseyin Kaya	33
Constructing coincident and leading indices of economic activity for the Brazilian economy	
by João Victor Issler, Hilton Hostalacio Notini and Claudia Fontoura Rodrigues	43
Heuristic model selection for leading indicators in Russia and Germany	
by Ivan Savin, and Peter Winker.	67

Editors

Editor-in-Chief

Michael Graff, *ETH Zurich, Switzerland*

Associate Editor

Gyorgy Gyomai, *OECD, Paris, France*

Editorial Board

Chair

Martine Durand, *Chief Statistician, OECD, Paris, France*

Kai Carstensen, *Ifo Institute for Economic Research, Munich, Germany and University of Munich, Germany*

Marcelle Chauvet, *University of California Riverside, CA, USA*

Richard Curtin, *University of Michigan, Ann Arbor, MI, USA*

Jorgen Elmeskov, *OECD, Paris, France*

Horst Entorf, *University of Frankfurt, Germany*

Hélène Erkel-Rousse, *Ministry of Finance and Economy, Paris, France*

Roberto Golinelli, *University of Bologna, Italy*

Robert J. Gordon, *Northwestern University, Evanston, IL, USA*

Ullrich Heilemann, *University of Leipzig, Germany*

Kajal Lahiri, *State University of New York at Albany, NY, USA*

Marco Lippi, *University of Rome “La Sapienza”, Italy*

Marco Malgarini, *ISTAT, Rome, Italy*

Zbigniew Matkowski, *Warsaw School of Economics, Warszawa, Poland*

Gian Luigi Mazzi, *Eurostat, Luxembourg*

Ronny Nilsson, *former Associate Editor of the JBCMA, OECD, Paris, France*

Simon van Norden, *HEC Montréal, Quebec, Canada*

Karl Heinrich Oppenländer, *former President of IFO Institute, Munich, Germany*

Ataman Ozyildirim, *The Conference Board, New York, NY, USA*

Günter Poser, *former Editor-in-Chief of JBCMA, Technische Universität Darmstadt, Germany*

Gerd Ronning, *University of Tübingen, Germany*

Bernd Schips, *former Editor-in-Chief of JBCMA, ETH Zurich, Switzerland*

Jan-Egbert Sturm, *KOF, Swiss Institute for Business Cycle Research, ETH Zurich, Switzerland*

István P. Székely, *DG ECFIN, European Commission, Brussels, Belgium and Corvinus University, Budapest, Hungary*

András Vértes, *GKI Economic Research Co., Budapest, Hungary*

Martin Weale, *NIESR, National Institute of Economic and Social Research, London, UK*

Measuring capacity utilisation in the Italian manufacturing sector: A comparison between time series and survey estimates

by

Marco Malgarini and Antonio Paradiso*

The aim of this paper is to provide an interpretation of the measure of capacity utilisation provided by the European Union harmonised survey on the Italian manufacturing sector. In doing so, we evaluate its ability to correctly track cyclical turning points and its contribution in explaining consumer price index (CPI) inflation. The survey based measure results are a good co-incident indicator of business cycle, however it is generally outperformed by time series models in explaining inflation. We conclude that the standard “output gap” interpretation of the survey results is broadly confirmed by the data, however we cannot rule out at this stage that survey respondents may also consider the alternative “variable capacity utilisation” concept in answering the survey question.

Keywords: Capacity utilisation, co-integration, unobserved component models, VAR

JEL classification: E32, C22, E37

* Corresponding author: Antonio Paradiso – University of Rome la Sapienza; anto_paradiso@hotmail.com. Marco Malgarini – ISTAT, Piazza dell’Indipendenza, 4, 00185, Roma, Italy; mmalgarini@istat.it. The views expressed in the article are those of the author and do not involve the responsibility of ISTAT.

1. Introduction

Capacity utilisation plays an important role in the evaluation of economic activity, contributing to the explanation of the behaviour of investment, inflation, productivity, profits and output (see for example, Abel, 1981; Greenwood *et al.*, 1988; Corrado and Matthey, 1997; Jorgensen *et al.*, 2009). However, an official statistical measure of capacity utilisation does not exist. For the manufacturing sector, it is normally obtained from firm-level surveys where firms are asked to indicate their current operating rate. Alternatively, measures of capacity utilisation may be derived using time series methods. The latter option has the advantages of being free to generate, easily reproducible and consistently measured over time.

The aim of this paper is to discuss the empirical properties of the survey measure of capacity utilisation realised for Italy by ISTAT in the framework of the European Union (EU)-harmonised business surveys project, and to compare these properties with those obtainable from time series estimates in order to obtain an economic interpretation of the meaning of the survey measure. Various methods to estimate capacity are discussed in Section 2, while Section 3.1 checks for their ability to correctly gauge cyclical turning points and their explicative power with respect to consumer inflation.¹ Indeed, the most used method to evaluate the explicative power of capacity consists of estimating a single equation in which capital utilisation is considered as exogenous (see for example McElhattan, 1985; Garner, 1994). Treating capacity as exogenous may however, be incorrect. In fact, a change in price inflation is also likely to affect capacity utilisation, especially in the case of supply shocks. Therefore, a single equation does not provide an adequate framework for analysing how changes in capital utilisation affect price inflation and a vector autoregression (VAR) approach appears more appropriate. The VAR approach is implemented in Section 3.2, where different VARs are estimated up to the fourth quarter of 2007 and inflation forecasts are produced for the eight quarters ahead. Results show that the survey-based measure of capacity utilisation is a broadly co-incident indicator of the business cycle. However, time series measures generally outperform it in explaining inflation during the latest recession.

The paper is organised as follows: Section 2 presents various methods used for estimating the capacity utilisation (survey, unobserved component model, Shaikh-Moudud co-integration method); Section 3 compares the various measures of capacity utilisation through their capacity in tracking the business cycle phases and in explaining consumer price inflation; and, Section 4 concludes.

2. Methods for estimating capacity utilisation

2.1. The survey method

The EU-harmonised survey on the manufacturing sector includes, on a quarterly basis, a question on the level of capacity utilisation currently experienced by the firm. In Italy, the survey has been historically performed by Institute for Short-Term Economic

Studies (ISCO) and Institute for Studies and Economic Analyses (ISAE) and is now administered by National Institute for Statistics (ISTAT).² In the current version of the questionnaire,³ the exact wording of the question is as follows:

“What is your current rate of capacity utilisation (in % of full capacity)?”

Individual data is progressively aggregated to obtain sample averages using information on firm dimension and the relative economic importance of each industry branch as weights.

The standard interpretation of capacity utilisation is that it measures how close the manufacturing sector is to full capacity. In this sense, it is considered as an approximation of the “output gap” concept currently used in business cycle analysis and in standard new-Keynesian macro-models. In this sense, capacity utilisation is defined as the ratio among actual and potential output:

$$cu = \frac{IP}{cap} \cdot 100 \quad (1)$$

where cu = capacity utilisation; IP = industrial production index; cap = capacity index or potential output.

An alternative view is that of the so-called “variable capacity utilisation” approach, where the current level of capacity is derived as the solution of an optimisation problem, given the expected marginal benefit of raising utilisation and its associated expected marginal cost. In this case, capacity is not necessarily varying across the business cycle, but as a result of profit maximisation strategies.⁴

In (1), the numerator is easily observable, while potential output may be estimated on the basis of some hypothesis on the technology in use, or more generally making use of time series methods. Hence, estimates on cu depend upon the assumptions made about the denominator in (1). Indeed, potential output is a vague and hard-to-measure concept, which varies over time and according to general economic conditions. In a strict sense, the term refers to the sustainable maximum output that could be produced by an existing (installed) manufacturing plant and machinery. However, sometimes other factors such as labour are taken into account (Johansen, 1968). Measures of capacity utilisation consistent with the output gap interpretation may be derived using time series methods.

2.1.1. *Unobserved component method*

Time series methods may be used to estimate potential output and then derive a measure of capacity utilisation on the basis of equation (1). In fact, industrial production, as any other observable time series, can be decomposed in two or more components that are not directly observable, such as the trend, the seasonal component and the cycle. The basic idea is that the unobservable components can be identified by assuming that they affect the variable that can be observed. In addition, the underlying processes behind the unobservable variables have to be specified. After the model has been specified, both the unobservable and the observable variables may be estimated with maximum likelihood methods using the Kalman filter. In the following, we will estimate two different classes of unobserved component models for the industrial production index, alternatively assuming that the unobserved components are autonomously identified (univariate model) or that other variables exist which are potentially correlated with industrial production (bivariate and trivariate models).

2.2. Univariate unobserved component method

The univariate model is based on a simple trend-cycle decomposition of log of industrial production index:

$$IP_t = T_t + C_t + \varepsilon_t, \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \quad (2)$$

where T_t represents the trend, C_t the cycle, and ε_t the irregular component. The stochastic trend component is specified as:

$$\begin{aligned} T_t &= T_{t-1} + \beta_{t-1} + \eta_t, \eta_t \sim NID(0, \sigma_\eta^2) \\ \beta_t + \beta_{t-1} + \zeta_t, \zeta_t &\sim NID(0, \sigma_\zeta^2) \end{aligned} \quad (3)$$

where β_t is the slope of the trend T_t . The irregular ε_t , the level disturbance η_t and the slope disturbance ζ_t are mutually uncorrelated.

The cycle C_t has the following specification:

$$\begin{bmatrix} C_t \\ C_t^* \end{bmatrix} = \rho_C \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} C_{t-1} \\ C_{t-1}^* \end{bmatrix} + \begin{bmatrix} k_t \\ k_t^* \end{bmatrix} \quad (4)$$

where ρ_C is a damping factor; $0 < \lambda_j \leq \pi$ is the frequency of the cycle in radians; k_t and k_t^* are two mutually uncorrelated NID disturbances with zero mean and common variance σ_k^2 . The period of cycle is equal to $2\pi/\lambda_j$.⁵

Details on estimated output are reported in Table 1. The diagnostic checking rejects the presence of serial correlation, heteroskedasticity and non-normality.

Table 1. **Univariate unobserved component model for log IP (industrial production): Main results**

	IP	
Level	4.48	(0.00)
Slope	-0.01	(0.03)
Cycle component	0.076	
Residual tests		
Normality	4.49	
DW	1.96	
H(64)	0.82	
Q(17,12)	12.37	
R ²	0.49	

Notes: Sample period 1960q1-2009q4. Outliers relative to 1969q4, 1972q4, 1973q3, 1974q4, 2009q1 were added in the estimation. The P-values are in parenthesis. The normality test statistic is the Bowman-Shenton statistic having a Chi-square distribution with two degrees of freedom under the null hypothesis of normally distributed errors. We reject the null if the calculated probability exceeds the tabulated ones equal to 5.99 at 5% significance level and 9.21% at 1% significance level. H(h) is the heteroskedasticity test statistics distributed as a F(h,h) with (h,h) degrees of freedom. Under the null of no heteroskedasticity and for h = 64, the 5% critical value is 1.51. DW is the classical Durbin-Watson test. Q(P,d) is the Ljung Box statistics based on the sum of the first P autocorrelations and it is tested against a Chi-Square distribution with d degrees of freedom. The null hypothesis of no autocorrelation is tested against the alternative of autocorrelation. The critical value for 12 degrees of freedom is 21.03 at 5% significance level.

2.3. Multivariate unobserved component methods

The univariate model can be extended to include other variables that are assumed to contain information about the potential industrial production; in the following, information about unemployment, gross domestic product (GDP) and CPI inflation is included in addition to industrial production (IP) resulting in four different models:

1. bivariate model considering the relationship between IP and inflation;

2. a bivariate model considering the relationship between IP and the unemployment rate;
3. a bivariate model considering the relationship between IP and GDP; and,
4. a trivariate model considering the relationship among IP, GDP, and unemployment rate.

All the variables are considered in terms of logarithms, with the only exception of the CPI inflation. The estimation results are reported in Table 2. In all models, the variables are modelled jointly, where the stochastic trends are independent of each other and the stochastic cycles are modelled as similar cycles (for details, see Harvey, 1989). In Model 1, inflation is modelled as a local level model as suggested by Cogley *et al.* (2010). Inflation is seasonally unadjusted and for this reason, a seasonal component is included in the estimates.⁶ In models 2 and 4, the unemployment rate is expressed in log as suggested by Chen and Mills (2009). Five dummies, all concentrated in the 60s, are also added to the unemployment equation due to important changes which occurred in the Italian labour market (Modigliani *et al.*, 1986); however, starting the estimation in the first quarter of 1970 does not significantly alter the results. Various dummies are also added to the IP equations in order to account for possible outliers.⁷

Table 2. **Multivariate unobserved component models for log IP**

Model	Mult 1		Mult 2		Mult 3		Mult 4		
	IP	INFL	IP	Unem	IP	GDP	IP	UNEM	GDP
Level	4.52 (0.00)	1.34 (0.00)	4.53 (0.00)	2.12 (0.00)	4.47 (0.00)	14.02 (0.00)	4.52 (0.00)	14.05 (0.00)	1.99 (0.00)
Slope	-0.01 (0.07)	–	-0.01 (0.04)	0.02 (0.05)	-0.02 (0.01)	-0.01 (0.05)	-0.01 (0.02)	-0.001 (0.36)	0.02 (0.08)
Seasonal Chi ² test	–	5.38 (0.11)	–	–	–	–	–	–	–
Cycle component	0.099	0.96	0.11	0.049	0.068	0.024	0.094	0.034	0.065
Residual tests									
Normality ¹	3.61	15.96	6.11	5.28	5.24	3.31	8.35	2.69	42.66
DW ²	2.07	2.03	1.98	1.97	1.99	1.87	2.07	1.88	2.05
H ³	0.69	0.23	0.55	0.25	0.64	0.37	0.62	0.35	0.21
Q (P, 12) ⁴	10.05	13.85	13.89	20.25	10.31	15.34	8.42	11.82	13.79
R ²	0.54	0.49	0.56	0.61	0.52	0.52	0.56	0.57	0.51

Notes: Sample period 1960q1-2009q4.

The P-values are in parenthesis.

1. Bowman-Shenton statistic, Chi-square (2). We reject the null if the calculated probability exceeds the tabulated ones equal to 5.99 at 5% significance level and 9.21% at 1% significance level.
2. Durbin-Watson test.
3. H(h) is the heteroskedasticity test statistics distributed as a F(h,h) with (h,h) degrees of freedom. H = 62 (model 1, 3 and 4 and H = 61 (model 2). Under the null of no heteroskedasticity the 5% critical value is 1.52 and for h = 62 and 1.53 for h = 61.
4. Ljung Box statistics based on the sum of the first P autocorrelations and it is tested against a Chi-square (12). The critical value for 12 degrees of freedom is 21.03 at 5% significance level.

The diagnostic checking for Mult2 and Mult3 models are all satisfactory. Mult1 rejects the presence of serial correlation, heteroskedasticity and non-normality for the industrial production equation. Although dummy variables are used to pick up major outliers in the inflation series, normality is still rejected for the inflation residuals; given the erratic movements of inflation in the 1970s and the subsequent sharp fall in the early 1980s, this is not surprising. The diagnostic tests on residuals of Mult4 model are satisfactory for all equations except for the unemployment rate where residuals are non-normal even if several dummies are applied.⁸

2.4. Shaikh-Moudud (SM) co-integration method

An alternative way to estimate potential output is to use the co-integration approach introduced by Shaikh and Moudud (2004). The essential idea is that capacity co-varies with the capital stock⁹ over the long run, making co-integration the natural tool for measuring capacity. The Shaikh-Moudud framework consists of an identity and two behavioural equations. The identity can be stated as $Y_t = (Y/Y^*) \cdot (Y^*/K) \cdot K$, where Y = output of manufacturing firms, Y^* = potential output, K = capital stock of manufacturing firms. If we now define $v = (K/Y^*)$, and cu = capacity utilisation rate = (Y/Y^*) , then:

$$\log(Y_t) = \log(K_t) - \log(v_t) + \log(cu_t) \quad (5)$$

The IP index (Y) and capital stock (K) are observed variables, but to turn the previous identity into a model we need to specify capacity utilisation (cu) and capital to capacity ratio (v). As for the first point, Shaikh-Moudud assumes that Y fluctuates around capacity over the long run, so that the actual cu fluctuates around some desired or normal rate of capacity utilisation. This means that manufacturing firms are able to maintain some correspondence between manufacturing capacity and actual production. In log terms, this implies that:

$$\log(cu_t) = e_{cu,t} \quad (6)$$

where $e_{cu,t}$ is a random error term.¹⁰

Moreover, Shaikh-Moudud assume that the capital-capacity ratio (v) changes over time, in response to autonomous (co-efficient b_1) and embodied technical change; the latter in turn depends on the rate of capital accumulation (co-efficient b_2). Letting g_v = growth rate of the capital capacity ratio v and g_K = growth rate of the capital stock K , we have $g_v = b_1 + b_2 \cdot g_K$. In log terms, with an added random error term $e_{v,t}$, this leads to the following equation:

$$\log(v_t) = b_0 + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot \log(K_t) + e_{v,t} \quad (7)$$

where it is assumed that an autonomous technical change follows a quadratic trend. Equations (5)–(7) form a general model of the relationship among industrial production and the capital stock, given technical change and fluctuating capacity utilisation. We expect that $b_3 > 0$ and that $b_1 > 0$, $b_2 < 0$. These last two signs indicate that in the last 30 years Italy is growing at decreasing rates accordingly with Paradiso and Rao (2011). Combining (5), (6) and (7), we obtain the following estimable equation:

$$\log(Y_t) = a_0 + a_1 \cdot t + a_2 \cdot t^2 + a_3 \cdot \log(K_t) + e_t \quad (8)$$

where $a_0 = -b_0$, $a_1 = -b_1$, $a_2 = -b_2$, $a_3 = 1 - b_3$, and error term $e_t = e_{cu,t} - e_{v,t}$, where the assumption is that the error terms are independent from each other.

Equation (8) implies that $\log Y$ and $\log K$ are co-integrated, up to a linear and quadratic deterministic trend in the data. The estimation output obtained is reported in Table 3. The diagnostic augmented Dickey-Fuller test (ADF) on residuals is satisfactory.

Table 3. **Estimation result of the long run relationship**

Sample period	a_0	a_1	a_2	a_3
1970q1-2009q4	1.97 (1.45)	0.01 (5.72)	-4.40E-05 (6.67)	0.35 (2.89)
Residual ADF t-test	-4.51***			

Note: Y = Value added of manufacturing firms (annual data, source EUKLEMS updated with ISTAT dataset); K = Capital stock of manufacturing firms (annual data, source EUKLEMS updated with ISTAT dataset). The data is transformed into a quarterly frequency using the method proposed in Fernandez (1981).

The methodology used is the Engle and Granger approach. This approach consists of estimating a long-run relationship [i.e., equation (4)] and making an ADF t-test on residuals. For the residual ADF t-test, the lag length is chosen by SIC criteria. Critical values for test are suggested by MacKinnon (1991). Symbols *, **, *** represents, respectively, significance level of 10%, 5%, and 1% of residual ADF t-test. The reported t-statistics (in parenthesis) are corrected for bias stemming from serial correlation and heteroskedasticity (HAC) by the Newey-West procedure.

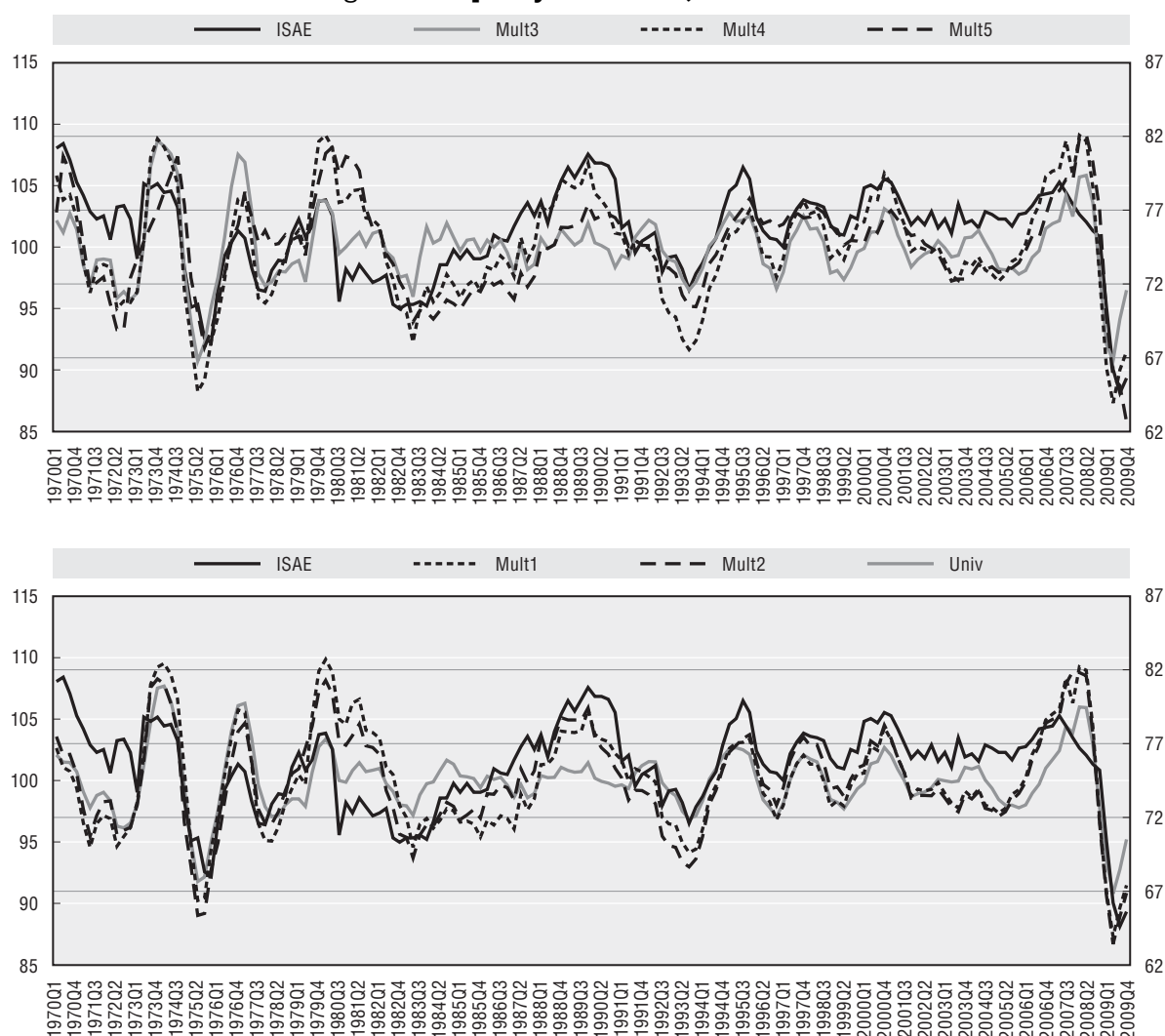
3. Interpreting the survey measure of capacity utilisation

Time series measures of capacity utilisation are derived under the “output gap” assumption. Therefore, if the survey-based measure is found to be consistent with its time-series counterparts and the business cycle, then answers provided by firms may be interpreted as an approximation of the output gap faced at the firm level. On the other hand, if time series and survey based measures are found to be significantly apart from each other, then it is well possible that the concept firms have in mind answering the survey question is different from the standard output gap assumption, possibly influenced by the optimisation process set in the “variable capacity utilisation” approach. In this respect, Figure 1 plots the various measures obtained in Section 2 in order to have a first visual appraisal of the difference existing between the various results. All the series show a similar pattern, even if in some cases downturn or upturn phases are under or over-emphasised; these characteristics reflect to a certain extent, the properties of the individual methods. For example, univariate and multivariate methods show that the trough reached in the current recession is roughly in line with that of 1975, whereas the other methods pose more emphasis on the recent recession.

3.1. Figure 2.3.1 Business cycle dating

In Table 4 we report the leads or lags turning points that capacity utilisation measures exhibit with respect to the benchmark business cycle dating for the Italian industry.¹¹ Peaks and troughs for the capacity utilisation series are identified by means of the Bry-Boschan-Harding-Pagan algorithm (Harding and Pagan, 2002). The output of the algorithm has not been taken at face value; in some cases it did not detect turning points that were instead evident from a visual inspection. In these cases the output is adjusted to consider these points.¹²

The univariate and bivariate IP-GDP methods correctly pick up all the turning points, although with some leads or lags in correspondence of some turning points. The others miss one turning point: three methods (A, D and G) miss the peak in the first quarter of 1992, whereas the ISAE survey method (G) misses the peak in the third quarter of 1974. All capacity utilisation measures identify extra cycles: univariate series identifies more additional extra cycles with respect to the others ($n = 8$). As for the values of capacity utilisation in correspondence of various peaks, it is interesting to note that only for bivariate IP-UR, trivariate, and Shaikh-Moudud methods are the highest values of capacity utilisation reached in the third quarter of 2007.

Figure 1. **Capacity utilisation, all methods**

Notes: Univ = Univariate model for industrial production; Mult1 = Bivariate model for industrial production and inflation; Mult2 = Bivariate model for industrial production and unemployment rate; Mult3 = Bivariate model for industrial production and GDP; Mult4 = Bivariate model for industrial production, unemployment rate and GDP; Mult5 = Shaikh-Moudud co-integration method; ISAE = ISAE measure of capacity utilisation. All series are represented on the left axis, except for ISAE, which is represented on the right axis.

Table 5 provides additional information on how well the competing capacity utilisation measures track economic fluctuations, focussing in particular on the timelessness of the detection. The empirical evidence suggests that, on average, the univariate series outperforms the others: the mean bias is zero. The trivariate and bivariate IP-UR estimates also show very good tracking properties. The Shaikh-Moudud estimate tends to be lagging at peaks, whereas the survey series tends to be on average a co-incident indicator of the reference cycle, albeit slightly lagging at troughs.

In what follows, we focus on the main economic recessions of the last 40 years: 1980-83 (caused by the 1979 energy crisis); 2001-03 (caused by the dot-com bubble); and, 2007-ongoing recession (caused by many factors, including the subprime mortgage crisis and the sharp rise in oil and food prices). Figure 2 makes a comparison between the recessions (in length and amplitude) of various capacity utilisation paths during these phases. In other words, we compare the various recession phases in terms of duration and amplitude. As always,

Table 4. **Turning point analysis: Leads (–) and lags (+) with respect to the growth-cycle reference dating**

Method	Trough	Peak	Trough	Peak	Trough	Peak	Trough	Peak
Growth cycle reference								
	1971q2	1974q3	1975q3	1976q4	1977q4	1981q2	1983q3	1985q1
	1985q3	1989q4	1991q3	1992q1	1993q3	1995q3	1996q4	1997q4
	1998q4	2000q4	2001q4	2003q1	2005q1	2007q3	2009q2	
A Univariate UC model	–2 (97.78)	0 (107.78)**	0 (91.77)**	0 (106.28)	1 (97.05)	0 (103.35)	1 (97.16)	0 (101.53)
B Bivariate IP-INFL UC model	–2 (94.55)	0 (109.52)	0 (90.24)	–1 (105.75)	1 (95.07)	0 (109.85)**	1 (94.55)	–
C Bivariate IP-UR UC model	–2 (95.47)	–1 (108.26)	0 (89)	0 (104.67)	1 (96.46)	0 (108.15)	1 (93.69)	–
D Bivariate IP-GDP UC model	–2 (96.88)	–1 (108.68)**	0 (90.72)	–1 (107.52)	0 (96.86)	–1 (103.79)	1 (95.95)	0 (102.17)
E Trivariate IP-GDP-UR UC model	–2* (96.27)	–1 (108.79)	0 (88.22)	0 (104.1)	0 (95.43)	0 (109.08)	1 (92.26)	–
F Shaikh-Moudud (SM) method	–2 (96.69)	2 (107.37)	1 (91.84)	0 (104.47)	1 (100.15)	1 (108.14)	1 (93.9)	1 (100.18)
G ISAE survey	1 (75)	–	2 (68)	–1 (75.6)	0 (71.5)	0 (77.7)	–1 (70.3)	1 (75.5)

Method	Trough	Peak	Trough	Peak	Trough	Peak	Trough	Number of extra cycles
Growth cycle reference dating								
	1993q3	1995q4	1996q4	2000q4	2003q2	2007q3	2009q2	
A Univariate UC model	0 (96.87)	–2 (102.66)	0 (97.02)	0 (102.68)	0 (99.87)	2 (105.98)	0 (90.79)	8
B Bivariate IP-INFL UC model	0 (94.14)	0 (103.36)	0 (96.8)	0 (104.31)	1 (97.76)	2 (109.24)	0 (87.26)	6
C Bivariate IP-UR UC model	0 (92.96)	0 (103.73)	0 (97.99)	0 (104.49)	1 (97.44)	1 (108.87)**	0 (86.67)	6
D Bivariate IP-GDP UC model	0 (96.49)	–3 (102.77)	0 (96.6)	0 (103.1)	0 (99.18)	3 (105.81)	–1 (89.57)	6
E Trivariate IP-GDP-UR UC model	0 (91.65)	0 (102.99)	0 (97.53)	0 (105.98)	1 (97.28)	2 (109.08)**	0 (87.29)	6
F Shaikh-Moudud (SM) method	1 (95.12)	0 (103.93)	0 (101.63)	1 (102.97)	0 (97.23)	3 (109.14)**	–	6
G ISAE survey	0 (71.6)	–1 (79.1)	1 (74.5)	0 (79.1)**	0 (75.4)	–1 (78.9)	1* (64.6)	4

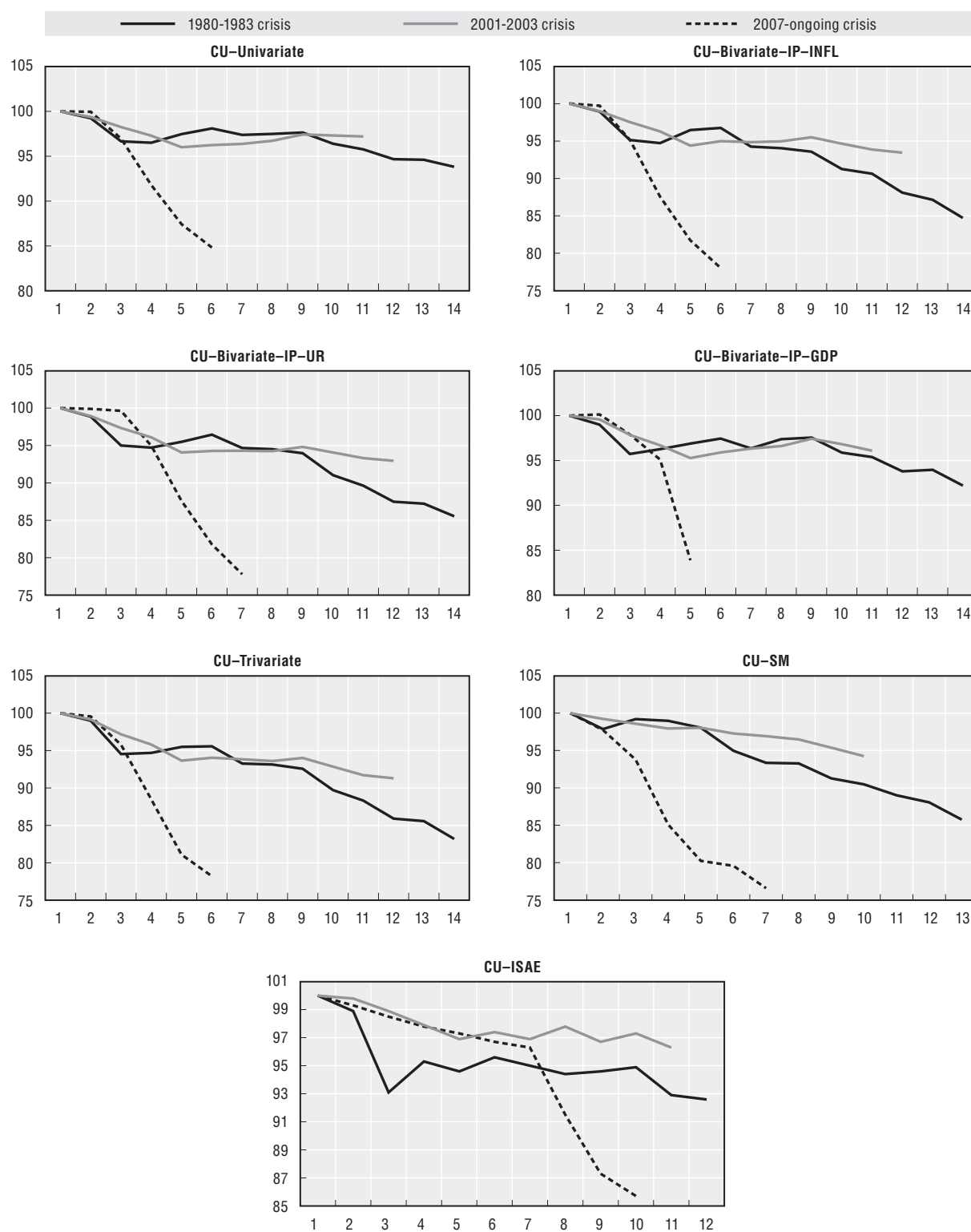
Notes: * Indicates the lead or lag (with respect to the reference dating) of capacity utilisation that is not classified as such by the Bry-Boschan algorithm but are evident on visual inspection; In parenthesis are the values of capacity utilisation in correspondence of peaks and troughs; ** indicates the maximum value of capacity utilisation observed in correspondence of peaks.

Table 5. **Average lags of the various capacity utilisation measures**

Method	Average Lag at		
	Peaks	Troughs	All
Univariate UC model	0	0	0
Bivariate IP-INFL UC model	+0.17	+0.12	+0.14
Bivariate IP-UR UC model	0	+0.12	+0.07
Bivariate IP-GDP UC model	–0.43	–0.12	–0.33
Trivariate IP-GDP-UR UC model	+0.17	0	+0.07
Shaikh-Moudud (SM) method	+1.14	+0.29	+0.71
ISAE survey	–0.33	+0.5	+0.14

Note: A positive (negative) sign indicate a lag (lead).

Figure 2. Recessions comparisons



Note : Time on horizontal axis is in quarters.

all the results are compared with the survey based capacity utilisation. In all measures used, the last cycle is characterised by the most severe recession. The survey capacity utilisation shows a recession duration of basically the same size of previous recessions. Referring to this last measure, it is interesting to note that up until seven quarters, the recession is of the same amplitude of the others and only after this point is there a severe slump in capacity utilisation.

3.2. Inflation and predictive power of capacity utilisation

So far, we have compared the properties of the different capacity utilisation measures with respect to the growth cycle dating. An alternative approach is to test to what extent capacity utilisation contributes to explain consumer price inflation. The most used method consists of estimating a single equation in which capacity utilisation is assumed to be an exogenous variable that explains inflation rate. We guess that bidirectional causality between capital utilisation and inflation rate is more consistent with the economic theory.

The linkage between these two variables can be understood using the familiar aggregate demand-aggregate supply (AD-AS) framework. A positive shift in aggregate demand will initially cause output and capacity utilisation to rise. As time passes, prices will begin to adjust and inflation will rise. As a consequence of rising prices, output and capacity utilisation rates will fall back to their initial levels and inflation will shrink.¹³ A shift in aggregate supply affecting prices positively will lead to a reduction in output and in capacity utilisation. However, reductions in output and capacity utilisation will constrain price dynamics, and consequently lead to output reduction. In the long run, no additional capacity will be built up, and we expect that capacity and actual output will move in a way that the capacity utilisation rate will return to its initial value.

This discussion suggests that a positive demand shock will cause a temporary positive effect on capacity utilisation and inflation; whereas a negative supply shock (for example a rise in oil price) will determine a temporary positive effect on inflation and a temporary negative effect on capacity utilisation.

VAR and associated impulse response functions are used to study the interaction of different measures of capacity utilisation and inflation. The standard form of the VAR model is:

$$\begin{bmatrix} \pi_t^c \\ cu_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \sum_{k=1}^K A_k \begin{bmatrix} \pi_{t-k}^c \\ cu_{t-k} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \quad (9)$$

where π^c is the cyclical component of the inflation¹⁴ (since that π is $I(1)$ ¹⁵) and cu is the capacity utilisation. $e_{1,t}$ and $e_{2,t}$ are errors terms which are composed of the structural innovations (that is, demand and supply shocks).

The theory does not specify a time frame for the relationship between cu and π , so we let the data determine the number k of quarters lags to include. Specifically, we set k basing on the Schwartz Information Criteria (SIC).¹⁶ On the basis of the VAR model, we employ the impulse response functions to illustrate how the two endogenous variables relate to each other over time. The identification of demand and supply shocks is achieved by assuming delayed impacts of demand shock on inflation, i.e. a Cholesky economy.¹⁷ All the capacity utilisation measures exhibit an impulse response function in accordance with the theoretical considerations except for univariate and bivariate IP-GDP measures: the response of the capacity utilisation to a supply shock is positive for the most part of

period.¹⁸ For conserving space we do not report the results, but these are available upon request from the authors.

The last important test is the out-of-sample forecasting performance of inflation. We have estimated each model for the period from the first quarter of 1970 to the fourth quarter of 2007. These estimates are used to predict the observations for the period from the first quarter of 2008 to the fourth quarter of 2009 (the period of current recession). These forecasts are compared with the actual values. The results are reported in Table 6.

Table 6. Out-of-sample performance of various capacity utilisation measures

	MAPE	RMSPE	% CDCP
Univariate UC			
Capacity utilis.	4.39	0.47	86
Inflation (cycle component)	48.37	6.03	100
Bivariate IP-INFL			
Capacity utilis.	6.93	0.77	100
Inflation (cycle component)	74.51	12.61	86
Bivariate IP-GDP			
Capacity utilis.	5.63	0.61	71
Inflation (cycle component)	30.70	4.17	86
Bivariate IP-UR			
Capacity utilis.	6.21	0.78	71
Inflation (cycle component)	195.89	32.13	86
Trivariate IP-GDP-UR			
Capacity utilis.	7.4	0.82	100
Inflation (cycle component)	41.37	6.09	100
Shaikh-Moudud (SM)			
Capacity utilis.	7.74	0.94	86
Inflation (cycle component)	76.12	10.91	86
ISAE survey			
Capacity utilis.	7.71	1.03	100
Inflation (cycle component)	80.27	9.2	100

Note: MAPE (Mean Absolute Percentage Error) = $\left[\frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \right] \cdot 100$, where A_t is the actual value and F_t is the forecast value. The difference between A_t and F_t is divided by the actual value A_t again. The absolute value of this calculation is summed for every fitted or forecast point in time and divided again by the number of fitted points n . This makes it a percentage error so one can compare the error of fitted time series that differ in level.

$$\text{RMSPE (Root Mean Square Percentage Error)} = \left[\frac{1}{n} \sum_{t=1}^n \left(\frac{A_t - F_t}{A_t} \right)^2 \right] \cdot 100$$

$$\% \text{ CDCP (Correct Direction Change Predictions)} = \frac{1}{n} \sum_{t=1}^n H_t \cdot 100, \text{ where } H_t = 1 \text{ if } (A_{t+1} - A_t) \cdot (F_{t+1} - F_t) > 0, \text{ otherwise}$$

$H_t = 0$. This indicator measures the ability of the model to predict direction changes irrespective of the forecast error magnitude.

The forecasts accuracy indexes show that, in general (with the exception of the bivariate IP-UR capacity utilisation estimation), survey based capacity utilisation exhibits a root mean square percentage error in line with that obtained using time series capacity utilisation measures. However, in some cases time series methods outperform the survey. Overall, the trivariate capacity utilisation estimation has stronger predictive power for the inflation rate.

4. Conclusions

This paper provides an interpretation for the capacity utilisation measure obtained from the EU-harmonised survey on the Italian manufacturing sector. From a turning point analysis of the last 40 years, it emerges that the measure obtained from the survey is a broadly co-incident indicator of the Italian business cycle, providing results similar to those obtained with time-series methods used to calculate potential output. In this sense, results are consistent with the standard “output gap” interpretation of the survey question. However if the output gap interpretation is correct, capacity utilisation should also show good explanatory properties with respect to consumer price inflation. In this sense, the survey measure has performed rather poorly, possibly hinting to different interpretation of the question on behalf of survey respondents. A possibility is that they derive their level of capacity utilisation as a result of an optimisation problem as in Christiano, Eichenbaum and Evans (2005). Further research is needed to more properly assess the relationship among the survey measures of capacity utilisation and the variable utilisation concept used in the literature. For the time being, our conclusion is that the survey results may be easily interpreted as a measure of the output gap, even if we cannot rule out the possibility that firms are also considering profits and costs behaviour in setting their level of capacity utilisation over time.

Notes

1. One of the basic propositions in macroeconomics is that inflation accelerates as capacity utilisation moves higher. For empirical studies which examine this evidence see McElhattan (1985), Garner (1994), and Emery and Chih-Ping Chang (1997).
2. Starting from 1 January 2011, ISTAT took all the functions previously exerted by ISAE including the realisation of business and consumers surveys.
3. More information on the harmonised project may be found here: http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm.
4. See Smets and Wouters (2007); Christiano, Eichenbaum and Evans (2005).
5. The decomposition presented above (local linear trend, stationary trigonometric cycle and an irregular component) is the same applied by Harvey (1989) to the GDP analysis.
6. The seasonal component is estimated as a trigonometric form (see Harvey (1989) for further

details) and is given by $\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}$, where γ_t is the seasonal component and each $\gamma_{j,t}$ is generated by

$$\begin{bmatrix} \gamma_{j,t} \\ * \\ \gamma_{j,t} \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ * \\ \gamma_{j,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ * \\ \omega_{j,t} \end{bmatrix}, \quad j = 1 \dots, [s/2] \\ t = 1 \dots T$$

where $\gamma_j = 2\pi j/s$ is the frequency, in radians, and the seasonal disturbances ω_t and ω_t^* are two mutually uncorrelated NID disturbances with zero mean and common variance σ_ω^2 . For s , the component at $j = s/2$ collapses to $\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \omega_{j,t}$.

7. The following impulse dummies are included in the IP equation: 1969q4, 1972q4, 1979q4, 2009q1.
8. It is interesting to note that the same problem also emerges in the unemployment residuals in a study by Chen and Mills (2009) conducted on the output gap estimation referring to the entire euro area.
9. For capital stock, we use the annual measure provided by EU-KLEMS. Quarterly data are derived using the Fernandez (1981) method implemented in the official Eurostat software ECOTRIM.
10. For the hypothesis that capacity utilisation fluctuates around the normal rate over the long-run, see for example Hein et al. (2010). In general, there is no guarantee that the rate of capacity

utilisation will converge to one (or any other specific number) in the long run. For example, Lavoie *et al.* (2004) have argued that the actual utilisation rate may deviate from its desired level in the long run, fluctuating around its desired level. See also Dutt (1997) and Cassetti (2006).

11. See ISTAT (2011).
12. Readers should bear in mind that survey based capacity utilisation is available in real time and it is not corrected with the availability of new information. On the other hand, both the dating exercise and the time series methods are “*ex post*”, in the sense that they consider the information that is available at the moment of writing. This information may be different both from what was available when the survey measure has been published and from what will be available, for example, five years from now. In this sense, performance comparisons may be considered as “unfair”. Nevertheless, we believe that the analysis do provide significant evidence on the relationship among survey based information and the business cycle. We however thank an anonymous referee for having raised this point.
13. In this context, we are assuming that AD curve does not lead to a shift in the AS curve. This implies that we are thinking in terms of an exogenous growth model framework. We thank an anonymous referee for having raised this point.
14. The cyclical component of inflation is obtained through the following univariate unobserved component model:

$$\pi_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t, t = 1, \dots, T$$

where μ_t represents the trend estimated as a driftless random walk, ψ_t a stochastic cycle, and γ_t a stochastic seasonal component. This univariate specification follows Koopman *et al.* (2009).
15. ADF and KPSS tests on inflation suggest that this variable is I(1); tests results are available with the authors upon request. Similar results for US and Australia are also obtained by Russell and Banerjee (2010).
16. The SIC model selection criterion is based on the log likelihood function adjusted by a penalty for the number of parameters, and performs well for model selection with large sample sizes (see Lutkepohl, 1991).
17. See Chapter 5 of Enders (2003) for an explanation.
18. For these models the effect of demand shock to inflation and capacity utilisation fade out after 40 periods. Shock responses are however statistically significant for six or seven periods.

References

- Abel, A.B. (1981), “A Dynamic Model of Investment and Capacity Utilization”, *Quarterly Journal of Economics*, 69(3), pp. 379-403.
- Bergheim, S. (2008), *Long-run Growth Forecasting*, Springer-Verlag, Berlin.
- Bruno, G. and E. Otranto (2008), “Models to Date the Business Cycle: The Italian Case”, *Economic Modelling*, 25(5), pp. 899-911.
- Cassetti, M. (2006), “A Note on the Long-Run Behaviour of Kaleckian Models”, *Review of Political Economy*, 18, No. 4, pp. 497-508.
- Chen, X. and T.C. Mills (2009), “Measuring the Euro Area Output Gap Using Multivariate Unobserved Components Models Containing Phase Shifts”, University of Glasgow, *Working Paper*, 2009-35.
- Cogley, T., G.E. Primiceri and T. Sargent (2010), “Inflation-Gap Persistence in the US”, *American Economic Journal: Macroeconomics*, 2(1), pp. 43-69.
- Corrado, C. and J. Matthey (1997), “Capacity Utilization”, *Journal of Economic Perspectives*, 11(1), pp. 151-167.
- Christiano, L.J., M. Eichenbaum, and C.L. Evans (2005), “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy”, *Journal of Political Economy*, 113, pp. 1-45.
- Dutt, A.K. (2006), “Aggregate Demand, Aggregate Supply and Economic Growth”, *International Review of Applied Economics*, 20, No. 3, pp. 319-336.
- Emery, K.M. and C. Chih-Ping (1997), “Is There a Stable Relationship Between Capacity Utilization and Inflation”, *Economic Review*, Federal Reserve Bank of Dallas, pp. 14-20.
- Enders, W. (2003), *Applied Econometric Time Series*, New York: Wiley.

- European Union, KLEMS website, www.euklems.net.
- Fernandez, R.B. (1981), "A Methodological Note on the Estimation of Time Series"; *Review of Economic Statistics*, 78, pp. 67-77.
- Garner, A.C. (1994), "Capacity Utilization and U.S. Inflation", *Economic Review, Federal Reserve Bank of Kansas City*, pp. 1-21.
- Greenwood, J., Z. Hercowitz and G.W. Huffman (1988), "Investment, Capacity Utilization, and the Real Business Cycle", *American Economic Review*, 78(3), pp. 402-417.
- Harding, D. and A. Pagan (2002), "Dissecting the Cycle: A Methodological Investigation", *Journal of Monetary Economics*, 49(2), pp. 265-381.
- Harvey, A.C. (1989), *Forecasting, Structural Time Series, and the Kalman Filter*, Cambridge: Cambridge University Press.
- Hein, E., M. Lavoie and T. Van Treeck (2010), "Harroldian Instability and the "Normal Rate" of Capacity Utilization in Kaleckian Models of Distribution and Growth – A Survey", *Metroeconomica*, DOI: 10.1111/j.1467-999X.2010.04106.x.
- Istituto nazionale di statistica (ISTAT) (2011), "Rapporto annuale: La situazione del Paese nel 2010".
- Johansen, L. (1968), "Production Functions and the Concept of Capacity", *Collection Economie e Mathematique et Econometrie*, 2, pp. 46-72.
- Jorgensen, B.N., J. Li and G. Sadka (2009), "Capacity Utilization, Profit Margins, and Stock Returns", Columbia University, *Working Paper Series*.
- Koopman, S. J. et al. (2009), *Structural Time Series Analyser, Modeller and Predictor – Stamp 8.2*, London: Timberlake Consultants.
- Lavoie, M., G. Rodriguez, and M. Seccareccia (2004), "Similitudes and Discrepancies in Post-Keynesian and Marxist Theories of Investment: A Theoretical and Empirical Investigation", *International Review of Applied Economics*, 18 (2), pp. 127-149.
- Lutkepohl, H. (1991), *Introduction to Multiply Time Series*. New York: Springer.
- MacKinnon, J.G. (1991), "Critical Values for Cointegration Tests", in R.F. Engle and C.W.J. Granger (eds), *Long Run Economic Relationship*, Oxford: Oxford University Press, pp. 267-276.
- McElhattan, R. (1985), "Inflation, Supply Shocks and Stable-Inflation Rate of Capacity Utilization", *Economic Review, Federal Reserve Bank of San Francisco*, pp. 45-63.
- Modigliani F., F. Padoa Schioppa and N. Rossi (1986), "Aggregate Unemployment in Italy, 1960-1983", *Economica, London School of Economics and Political Science*, 53, 210, pp. 245-273.
- Paradiso, A. and B.B. Rao (2011), "How to Offset the Negative Trend Growth Rate the Italian Economy?", *Applied Economics Letters*, No. 15, pp. 1479-1483.
- Russell, B. and A. Banerjee (2008), "The Long-Run Phillips Curve and Non-Stationary Inflation", *Journal of Macroeconomics*, Vol. 30, pp. 1792-1815.
- Shaikh, A.M. and J.K. Moudud (2004), "Measuring Capacity Utilization in OECD Countries: A Cointegration Method", The Jerome Levy Economics Institute of Bard College, *Working Paper* 415.
- Shapiro, M. (1989), "Assessing the Federal Reserve's Measures of Capacity Utilization", *Brooking Papers on Economic Activity*, Vol. 1, pp. 181-241.
- Smets, F. and R. Wouters (2007), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *The American Economic review*, Vol. 97, pp. 586-606.

Nowcasting Irish GDP*

by

Antonello D'Agostino,** Kieran McQuinn and Derry O'Brien

In this paper we present a dynamic factor model that produces nowcasts and backcasts of Irish quarterly GDP using timely data from a panel dataset of 35 indicators. We apply a recently developed methodology, whereby numerous potentially useful indicator series for Irish GDP can be availed of in a parsimonious manner and the unsynchronised nature of the release calendar for a wide range of higher frequency indicators can be handled. The nowcasts in this paper are generated by using dynamic factor analysis to extract common factors from the panel dataset. Bridge equations are then used to relate these factors to quarterly GDP estimates. We conduct an out-of-sample forecasting simulation exercise, where the results of the nowcasting exercise are compared with those of a standard benchmark model.

JEL classification: C33, C53, E52.

Keywords: GDP, forecasting, factors, Ireland.

* The views expressed in this paper are our own, and do not necessarily reflect the views of the European Central Bank or the Central Bank of Ireland.

** Corresponding author: European Central Bank, Postfach 16 03 19, 600 66, Frankfurt am Main, Germany; antonello.dagostino@ecb.int.

1. Introduction

In this paper early estimates or nowcasts and backcasts of quarterly Irish GDP are presented. A nowcast (backcast) is a prediction of the present (near past) of a series with a publication lag. We draw upon the factor analysis based literature in seeking to distill significant information from relatively large amounts of variables. The approach follows that of Giannone, Reichlin and Small (2008) who produce nowcasts of output for the US using a dynamic factor model. The nowcasting methodology has been applied to many countries including Euro Area (Angelini *et al.*, 2010), Norway (Aastveit and Trovik, 2011), Switzerland (Siliverstovs and Khodidilov, 2010), New Zealand (Matheson, 2011). For a survey of the literature see Banbura, Giannone and Reichlin (2011).

A nowcast estimate of GDP is obtained in two steps. In the first step, monthly indicators are used to estimate factors. These factors are then used as regressors in an associated bridge equation. In an Irish case, we compile a monthly panel dataset of 35 variables. In terms of the timeliness of Irish GDP releases, for the first two months in any given quarter, the most recent available release of GDP is for the second last quarter. By the end of the third month in each quarter, releases of GDP are available for the previous quarter. In this paper we generate estimates for the current quarter (nowcast) and for the previous quarter (backcast). In the case of the latter, this is only done when no release is available *i.e.* for the first two months of the quarter.

We generate the nowcasts using a pseudo-real time approach. By this we mean that when a nowcast is derived from the data in every quarter, the data availability which existed at that quarter, is replicated exactly. In essence we are seeking to replicate the timeliness which would have pertained for an analyst at the time the GDP estimate was formulated. By pseudo we mean that it is the final vintage of this data that is used to produce the forecasts and not the “real-time” data available at the time of the forecast, however there is evidence that the qualitative results of the model are unaffected by data revisions (see Liebermann, 2011).

This approach does give rise to what has been referred to as the “jagged edge” issue; that is, some data series do not have observations for the most recent month or two, so the panel from which the factor is derived is unbalanced in nature. In addressing this problem we follow the same two-step approach as Giannone, Reichlin and Small (2008), in which current and recent GDP is estimated conditional on the number of indicators for which current information is already available.

This modeling approach represents a significant addition to the policy-analysis tool kit of the Irish Central Bank. The coherent nature of policy making within the Eurosystem necessitates the provision of timely and accurate estimates of output growth by euro member states. Evaluating the present state of the economy and generating “credible” short-term forecasts has often been a complex task of combining information from both qualitative and quantitative based sources usually available at different time delays. Qualitative, survey-type information concerning present conditions within the economy

tends to be available on a timely basis, whereas data more typically used in model based forecasts is often only available at a significant time lag. Additionally, many timely and useful variables are released at monthly intervals whereas the variable of interest, GDP, is normally on a quarterly basis. Therefore, from a policy perspective the flexibility afforded by the nowcasting model is a particularly attractive feature of this approach.

To place the nowcasting exercise in context, in Section 2 we discuss both the type and timeliness of Irish macroeconomic data releases. In Section 3, the nowcasting model is presented along with the results of an out-of-sample forecast simulation. Section 4 offers some concluding comments.

2. Real-time data flow and the indicator dataset

The Quarterly National Accounts (QNA) releases of the Irish Central Statistics Office (CSO) provide the most comprehensive available information on recent developments in the Irish economy. The QNA provide estimates of GDP and its main output and expenditure components with the income accounts only available on an annual basis. The current release delay is no later than 90 days, meaning that GDP growth for a reference quarter is available towards the end of the subsequent quarter. In view of this significant release delay, conjunctural assessments of the Irish economy would benefit greatly from an early indicator of quarterly GDP of sufficient accuracy and timeliness. Such an early indicator could avail of the many releases of monthly Irish data.

In order to illustrate more fully the benefits of producing backcasts and nowcasts for Irish GDP, it is helpful to examine the timeline of releases of Irish macroeconomic data more carefully. Let's suppose we are currently half way through a particular quarter and label the quarter q_t . The release of the quarterly national accounts for the previous quarter q_{t-1} , will still not appear for another six weeks or so i.e. close to the end of q_t . However there is considerable monthly information now available for the previous quarter including monthly releases of retail sales and industrial production for each month of the previous quarter. This paper proposes to make use of the available monthly information to produce a timely backcast of GDP i.e. an estimate of GDP for q_{t-1} . At the current juncture in the middle of q_t , there is already partial information available on the current quarter including the unemployment rate and exchange rate data for the first month of the quarter. The methodology adopted in this paper also makes use of these monthly data to produce a timely nowcast of GDP for q_t . Such a nowcast yields significant benefits in terms of providing important timely information on the state of the economy. The official quarterly national accounts for q_t are released at the end of q_{t+1} . The nowcasts and backcasts for GDP can be computed at any point during q_t but clearly the nowcasts are more reliable as we proceed through the quarter when more and more information on monthly indicators can be incorporated.

The GDP nowcasting model incorporates information from a large set of more timely, higher frequency indicators that try to capture conjunctural developments in the Irish and international economies. There are 35 indicator series in the conditioning set. The full list of indicators along with their respective sources and release delays are presented in Table 1. These series are part of a larger set of series used by the Central Bank of Ireland in projection exercises but the series in the conditioning set satisfied a number of additional criteria including having a sufficiently timely release delay. The series are generally of monthly frequency and are significantly more timely than the GDP releases, with the

longest release delay for the monthly series at about 40 days. Each of the series must also be sufficiently long for modeling purposes. The dataset begins in January 1985 and is unbalanced at the end of the sample reflecting the different release delays of the indicators. The structure of the dataset should be largely the same, at least for the set of monthly series, at each monthly update of the quarterly GDP nowcast. The model attempts to nowcast year-on-year GDP growth for a given quarter and the indicator series undergo transformations before entering the model. Typically, the series are converted to year-on-year growth rates helping to avoid the excessive volatility of quarter-on-quarter growth rates.

Table 1. **List of variables used in the factor analysis**

Name	Frequency	Timeliness (approx. days)	Source
Live register	M	10	www.cso.ie/prlabfor.htm
Retail sales	M	40	www.cso.ie/prservices.htm
Car sales	M	10	www.cso.ie/prtransport.htm
Unemployment rate	M	10	www.cso.ie/prlabfor.htm
Industrial production	M	40	www.cso.ie/prind.htm
Real M1	M	30	www.centralbank.ie
Real M3	M	30	www.centralbank.ie
Real private sector credit	M	30	www.centralbank.ie
CPI sub-indices (7 series)	M	30	www.cso.ie/prprices.htm
House completions	M	20	www.environ.ie/
House registrations	M	20	www.environ.ie/
Consumer sentiment index	M	8	www.esri.ie/
Index of consumer expectations	M	8	www.esri.ie/
Euro sterling exchange rate	M	0	www.centralbank.ie
Euro dollar exchange rate	M	0	www.centralbank.ie
Euro area consumer and business surveys (11 series)	M	3	http://ec.europa.eu/
Extra euro area import demand of Irish trading partners	Q	Variable	ECB
Intra euro area competitors' prices on export side	Q	Variable	ECB
Gross domestic product	Q	90	www.cso.ie/prprices.htm

The dataset contains direct measures of economic activity and price dynamics along with indirect measures such as business and consumer sentiment surveys. The monthly indicators presented in Table 1 can be broadly grouped in terms of type and release delay (in brackets) as follows:

- hard data 1 (release delays of less than 10 days) – includes live register of unemployment benefit claimants, the unemployment rate and car sales,
- hard data 2 (between 20 and 40 days) – includes housing data, retail sales and industrial production,
- price data (within 12 days) – consumer price sub-indices,
- survey data (within 8 days) – Irish consumer sentiment surveys and euro area consumer and business sentiment surveys,
- financial data 1 (at the end of reference month) – euro sterling and euro dollar exchanges rates,
- financial data 2 (within 30 days) – monetary aggregates and private sector credit.

The importance of developments in trading partners for the Irish economy is captured by the inclusion of quarterly weighted import demand and competitiveness variables along

with higher frequency indicators such as consumer and business sentiment indicators for the euro area and euro exchange rate movements *vis-à-vis* pound sterling and the US dollar. Sentiment indicators offer very timely information on conditions in the domestic and international economies. Exchange rate data are daily but they enter the model as monthly averages. Almost each sector of the economy is represented but efforts are made to adequately cover particularly those sectors with both higher weighting and more volatile outturns. Industrial output, which accounts for about a quarter of GDP at factor cost, is an important source of volatility, as illustrated in Table 2. The volatility is particularly pronounced in certain manufacturing sub-sectors, such as the manufacture of basic chemicals, and this can present significant challenges in a forecasting context. The overall monthly industrial output index is included as an indicator.

Table 2. **Mean absolute deviations of year-on-year growth rates by sector**

	Mean absolute deviation	Share of GDP at factor cost
Agriculture	28.0	2.3
Industry (excl. construction)	31.4	25.1
Building and construction	16.7	8.5
Distribution, transport and communication	8.0	15.6
Public administration and defence	2.4	3.4
Other services	8.1	46.2

Note: Shares are approximate, due to non-additivity of the chained-linked data, and do not add to 100.

It is worth noting that the explanatory power of industrial production indices may be limited by the fact that the monthly industrial production series are not adjusted for royalties and licence services imports whereas GDP is adjusted as these inputs are not regarded as value added. In this respect, it is worth observing that the increasing use of service inputs over time may not be adequately taken into account (data on services inputs are only available quarterly with the Balance of International Payments, which is released at the same time as the QNA). Monetary aggregates have been shown to be useful in short-term forecasting of activity and are also included.

The contribution of the construction sector to GDP growth has undergone significant changes during this decade and indicators such as housing completions, housing registrations and private sector credit are included to capture activity in the sector. Activity in the market services sector is accounted for primarily by the monthly retail sales and car sales indices. The inclusion of a range of public finance series was also considered but unfortunately the data were not available at a high frequency over a sufficiently long period of time.¹ Altogether, seven subindices of the CPI remained in our final set of indicators after the various steps of the selection process. Finally, there are two labour market indicators i.e. the monthly unemployment rate and the numbers on the live register of unemployment benefit claimants.

3. The model

In this section we outline the dynamic factor model (Giannone, Reichlin and Small, 2008) used to generate the monthly estimates of GDP. The estimation strategy with this approach is twofold: in the first a set of factors are extracted from a panel of monthly indicators; in the second step, the GDP series is projected onto the factors via a bridge equation.

The Giannone, Reichlin and Small (2008) model can be summarised as follows. A vector of n monthly, stationary (standardised), year-on-year variables $x_t = (x_{1,t}, x_{2,t}, \dots, x_{n,t})'$, $t = 1, 2, \dots, T$ is assumed to have the following dynamic factor model characterisation:²

$$x_t = \chi_t + \xi_t = \Lambda f_t + \xi_t \quad (1)$$

$$f_t = \sum_{i=1}^p A_i f_{t-i} + \zeta_t \quad (2)$$

$$\zeta_t = B\eta_t \quad (3)$$

where x_t in eq.(1) is the sum of two orthogonal components, the common component χ_t , and the idiosyncratic component ξ_t . The common component is the product of an $n \times r$ matrix of loadings Λ and a $r \times 1$ vector of latent factors f_t . The idiosyncratic component is a multivariate white noise with diagonal covariance matrix Σ_ξ . Factor dynamics are described in eq.(2), which is a VAR(p). $A_1, A_2, \dots, A_p$ are matrices of parameters and $\zeta_t: N(0, BB')$, where B is a $(r \times q)$ matrix³ with $q \leq r$; $\eta_t: N(0, I_q)$.

In the Appendix we outline how consistent estimates of the parameters of the model can be obtained. Using these estimates, the factors can be estimated in the following manner:

$$\hat{F}_t = \text{proj}[F_t | x_1, \dots, x_T; \hat{\Lambda}, \hat{A}, \hat{B}, \hat{\Sigma}_\xi]$$

that is, by applying the Kalman filter to the state-space representation obtained by replacing estimated parameters in the factor representation:

$$x_t = \hat{\Lambda} f_t + \xi_t \quad (4)$$

$$f_t = \sum_{i=1}^p \hat{A}_i f_{t-i} + \zeta_t \quad (5)$$

The Kalman filter can be also used to evaluate the degree of precision of the factor estimates

$$V_k = E[(F_t - \hat{F}_t)(F_t - \hat{F}_t)' | x_1, \dots, x_T; \hat{\Lambda}, \hat{A}, \hat{B}, \hat{\Sigma}_\xi].$$

while, the estimates of the signal and their degree of precision are given, respectively by

$$\chi_t = \text{Proj}[\chi_t | x_1, \dots, x_T; \hat{\Lambda}, \hat{A}, \hat{B}, \hat{\Sigma}_\xi] = \hat{\Lambda} \hat{F}_t$$

$$E(\chi_t - \hat{\chi}_t)^2 = \hat{\Lambda}' V_0 \hat{\Lambda}$$

This framework is adapted to estimate the factors on the basis of an incomplete dataset, i.e. a dataset which contains some missing values corresponding to data which has not yet been released. We impose that the variance of the idiosyncratic part of the series with the missing observation is infinite at time t . This implies that no weight is put on the missing variable in the computation of the factors at time t .

We define the level of the GDP for a quarter q as the average of latent, monthly observations of GDP within that quarter, $GDP_t^q = \frac{1}{3}(GDP_t^m + GDP_{t-1}^m + GDP_{t-2}^m)$; where GDP_t^m denotes the unobservable latent realisation of GDP at the monthly frequency. Year-on-year GDP growth is computed as $gdp_t^q = GDP_t^q - GDP_{t-12}^q$ within the same quarter q . Quarterly factors are computed by averaging over the year-on-year monthly factors $f_t^q = \frac{1}{3}(f_t^m + f_{t-1}^m + f_{t-2}^m)$.

The estimates of year-on-year changes of GDP, on quarterly variables, are computed with the following bridge equation:

$$\widehat{gdp}_t^q = \hat{\beta}' \hat{f}_t^q \quad (6)$$

where $\widehat{gdp}_t^q$ denotes the yearly estimated growth rate of GDP and $\hat{\beta}$ is a $r \times 1$ vector of estimated parameters; this is computed for all the available t 's belonging to a certain

quarter q . Backcasts, nowcasts and forecasts of the GDP series can be computed every month as soon as new information becomes available.

The forecast error is defined as the difference between the (ex post) realised and the estimated value of gdp : $\varepsilon_t^q = gdp_t^q - \widehat{gdp}_t^q$. We assume that $\varepsilon_t^q \sim N(0, \sigma_\varepsilon^2)$ and that ξ_t , ζ_t and ε_t are mutually independent at all leads and lags.

3.1. Model evaluation

To evaluate the forecast performance of the modelling approach, we perform a pseudo real-time out of sample simulation. In using the pseudo real-time approach, we are seeking to replicate the actual data availability situation which pertained at the time the nowcast/forecast is generated. Therefore, the parameters of the model are generated recursively based on the data availability at a particular quarter.

The out of sample simulation procedure is as follows. The exercise begins by estimating the model on a sub-sample called the estimation window 1980:Q1 to 1996:Q4. The estimated parameters are then used to backcast and nowcast GDP. The estimation window is updated sequentially with one observation and the parameters are re-estimated based on the new sample available. The estimates of GDP are again generated using the new sample. This procedure is then iterated until the end of the sample.

We evaluate the performance of the model by generating the Mean Squared Forecast Error (MSFE), which is defined as

$$MSFE = \frac{1}{(t_1 - t_0 + 1)} \sum_{t=t_0}^{t_1} (gdp_t^q - \widehat{gdp}_t^q)^2, \quad (7)$$

The MSFE refers to Mean Squared Nowcast Error (MSNE) when the evaluation is computed for the months within the current quarter or refers to Mean Squared Backcast Error (MSBE) when the evaluation is computed in the months falling in the next quarter.

We also compare the accuracy of the models estimates with that of a benchmark model.⁴ In our case we take, as the benchmark model, the average of the last four most recently available year-on-year GDP changes. The results in the model simulation are generated with a specification with one dynamic factor, one static factor and the VAR for the factors of order 4. This specification results in the lowest mean square forecast error in a training sample from 1990 Quarter 1 to 1996 Quarter 4 as well as in the sample used for the simulation.⁵

3.2. Results

We now compare the forecast performance of the model in terms of both nowcasts and backcasts vis-à-vis that of the benchmark model. Table 3 presents the mean squared errors (MSE) for the different applications. These are presented for the case where the nowcast or the backcast is generated for each of the three different months in each quarter.

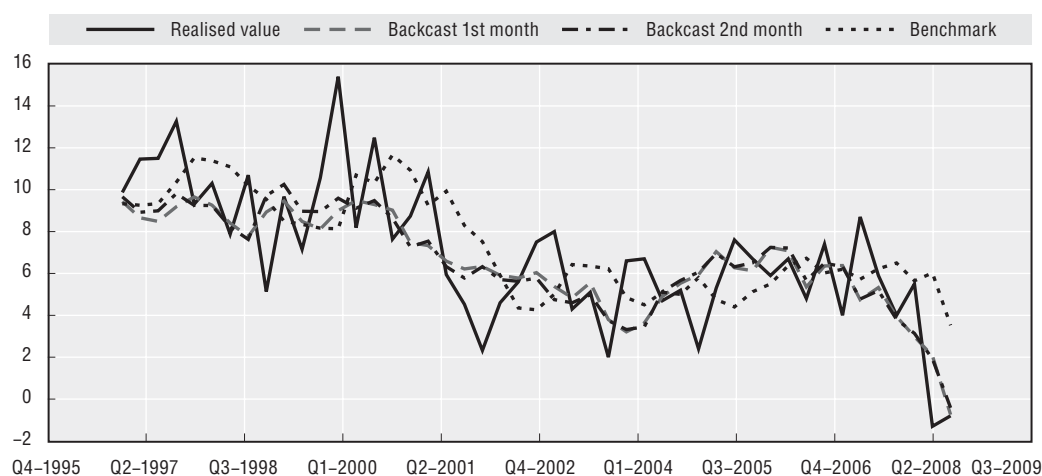
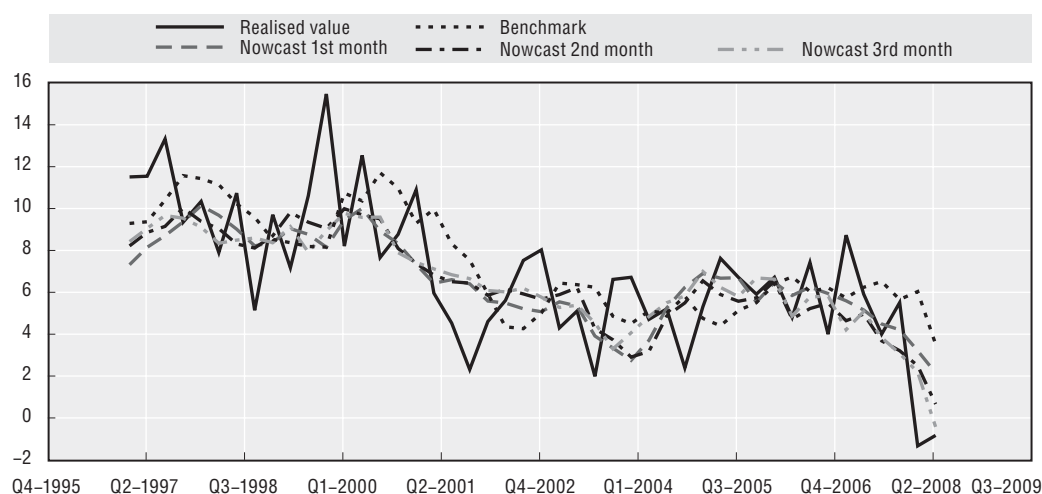
It can be seen from Table 3 that in both the case of the backcasts and the nowcasts, the mean squared backcast error (MSBE) and the mean squared nowcast error (MSNE) of the benchmark model is considerably greater than the model proposed here. In terms of the month in the quarter the now/backcast is generated, it is evident, as one would expect, that as one moves from the first month to the second and onto the third month, there are notable declines in the MSBE and the MSNE.

In Figures 1 and 2, we plot the backcast and the nowcast respectively along with the observed series and the results from the benchmark model. From Figure 1, it may be

Table 3. **Mean squared errors (MSE) for backcasts and nowcasts**

Model	MSBE	MSNE
1st month	5.317	6.145
2nd month	5.034	5.570
3rd month		5.475
Benchmark	8.189	8.361

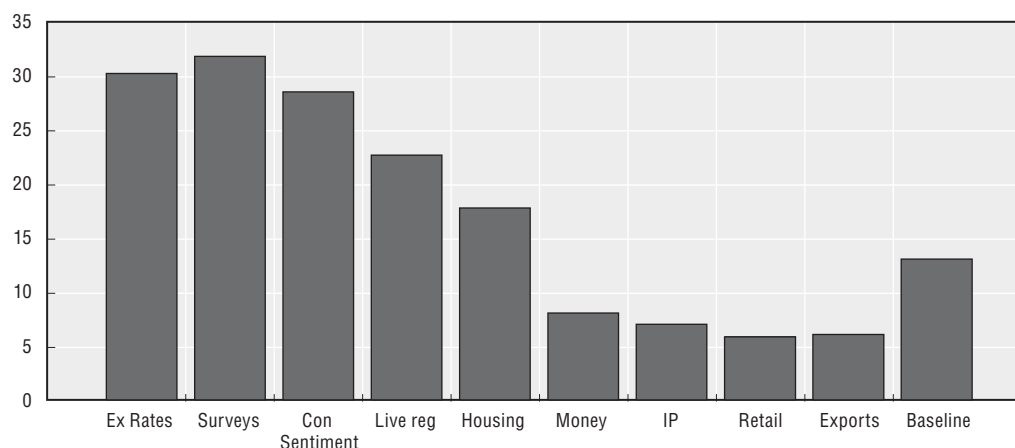
observed that the backcast generated for the second month tracks the observed series quite well, particularly when compared with the estimate of the benchmark. In the case of the nowcast estimates in Figure 2, the estimate generated for the third month can also be seen to improve on that estimated in the first and second months of the quarter.

Figure 1. **Comparison of backcasting performance**Figure 2. **Comparison of nowcasting performance**

As a final exercise, we examine the effect of individual data releases on the forecast accuracy of the approach. In Figure 3, we plot the MSFE for the nowcast associated with the

addition of each data release. This is compared with the score for the benchmark. As can be seen from Table 1, the first set of available releases contains exchange rates while the last is for extra euro area demand for Irish exports and competitors' prices on the export side. Apart from surveys, a marginal improvement can be observed relative to the forecasting performance of the previous release. The improvement is particularly notable for the releases of the live register of unemployment claimants, housing data and monetary aggregates.

Figure 3. **Impact of data releases on forecast accuracy**



4. Conclusions

The use of the dynamic factor model framework for producing nowcasts and backcasts represents a significant addition to the forecasting tool kit of the Central Bank of Ireland. In providing timely estimates of quarterly GDP, the approach has some appealing features. A large panel dataset of potential indicators for GDP may be parsimoniously employed through the dynamic factor model methodology, and at any time during a quarter the estimation approach can handle and exploit the unsynchronised flow of generally higher frequency data.

In evaluating the dynamic factor model, a pseudo-real time approach is followed in that the data availability situation which existed for each quarter, is replicated when computing the model estimates. An out-of-sample simulation is performed where the estimates of the model are compared with that of a benchmark approach. We find that the mean squared forecast errors for both the nowcasts and the backcasts are considerably smaller than those of the benchmark model. Unsurprisingly, the later in the quarter the nowcast or the backcast is generated the more accurate the estimate proves to be relative to the observed series. This suggests that macroeconomic data releases have an important informational content; it is hence worth updating the forecast many times within the same quarter to incorporate the most recent information.

Notes

1. There is the option of interpolating some of the series and we may revisit this at some future date as some of these series could prove useful.
2. T is the last date available in the monthly year-on-year dataset.

3. We assume $B'B = \Sigma$.
4. The standard benchmark model in this literature is the constant growth model. However, owing to the particularly volatile nature of Irish quarter-on-quarter GDP changes, we elect to use, as the standard GDP transformation, year-on-year changes. Therefore, for such a transformation, the average growth rate is a more appropriate model rather than the constant growth rate. Nonetheless, we also compare our results with those of the standard benchmark model and the results do not change. The results are available upon request, from the authors.
5. Given that the number of dynamic factors is equal to the number of static factors, the dataset is characterised by low dynamics which are captured by the lags in the factors VAR.

References

- Aastveit, K. and T.G. Tørvik (2007), "Nowcasting Norwegian GDP: The Role of Asset Prices in a Small Open Economy", *Norwegian Central Bank Working Paper* 2007/9.
- Angelini, E., G. Camba-Méndez, D. Giannone, G. Rünstler and L. Reichlin (2008), "Short-term Forecasts of Euro Area GDP Growth", *Working Paper Series* 949, European Central Bank.
- Banbura, M., D. Giannone and L. Reichlin (2010), "Nowcasting" in M.P. Clements and D.F. Hendry (eds.), *Oxford Handbook on Economic Forecasting*, Oxford University Press, USA.
- Giannone, D., L. Reichlin and D. Small (2008), "Nowcasting: The Real-Time Informational Content of Macroeconomic Data", *Journal of Monetary Economics*, Vol. 55(4), Elsevier, pp. 665-676.
- Liebermann, J. (2011), "Real-Time Nowcasting of GDP: Factor Model Versus Professional Forecasters", *Central Bank of Ireland Research Technical Paper* 03/RT/11.
- Matheson, T.D. (2010), "An Analysis of the Informational Content of New Zealand Data Releases: The Importance of Business Opinion Surveys", *Economic Modelling*, Elsevier, Vol. 27(1), pp. 304-314.
- Siliverstovs, B. and K.A. Kholodilin (2010), "Assessing the Real-Time Informational Content of Macroeconomic Data Releases for Now-/Forecasting GDP: Evidence for Switzerland", *Discussion Papers of DIW Berlin* 970, DIW Berlin, German Institute for Economic Research.
- Stock, J.H. and M. Watson (2002a), "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association*, 97(460), pp. 147-162.
- Stock, J.H. and M. Watson (2002b), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20(2), pp. 147-162.

APPENDIX A

Parameters estimation

In this Appendix, we outline how consistent estimates of the parameters of the dynamic factor model are obtained.

Suppose that $z_{it} = y_{it} - \hat{\mu}_i$ and that $x_{it} = \frac{1}{\hat{\sigma}_i}(y_{it} - \hat{\mu}_i)$, where $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$ and $\hat{\sigma}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu}_i)^2}$. Consider the following estimator of the common factors: $(\tilde{F}_t, \hat{\Lambda}) = \underset{\tilde{F}_t, \hat{\Lambda}}{\operatorname{argmin}} \sum_{t=1}^T \sum_{i=1}^n (z_{it} - \Lambda_i \tilde{F}_t)^2$.

The correlation matrix of the observables (y_t) can be defined as:

$$S = \frac{1}{T} \sum_{t=1}^T x_t x_t'$$

Let's define D the $r \times r$ diagonal matrix with diagonal elements given by the r largest eigenvalues of S and V the $n \times r$ matrix of the corresponding eigenvectors subject to the normalisation $V'V = I_r$. Factors are estimated as:

$$\tilde{F}_t = \Lambda x_t$$

and the factor loadings $\hat{\Lambda}$ are estimated by regressing the variables on the estimated factors:

$$\hat{\Lambda} = \sum_{t=1}^T x_t \tilde{F}_t' (\sum_{t=1}^T \tilde{F}_t \tilde{F}_t')^{-1}$$

and the covariance matrix of the idiosyncratic component is estimated as:

$$\hat{\Sigma}_\xi = \operatorname{diags}(S - VDV)$$

The other parameters $\hat{A}$ and Σ are estimated by running a VAR on the estimated factors:

$$\hat{A} = \sum_{t=2}^T \tilde{F}_t \tilde{F}_{t-1}' (\sum_{t=2}^T \tilde{F}_{t-1} \tilde{F}_{t-1}')^{-1}$$

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=2}^T \tilde{F}_t \tilde{F}_t' - \hat{A} (\frac{1}{T-1} \sum_{t=2}^T \tilde{F}_{t-1} \tilde{F}_{t-1}') \hat{A}'$$

Finally, P is defined as the $q \times q$ diagonal matrix with the entries given by the largest q eigenvalues of $\hat{\Sigma}$ and by M the $r \times q$ matrix of the corresponding eigenvectors, then:

$$\hat{B} = MP^{\frac{1}{2}}$$

Day-of-the-week effect in Consumer Confidence Index: The case of Turkey

by

Sadullah Çelik and Hüseyin Kaya*

The aim of this study is to examine the validity of the day-of-the-week effect on both mean and volatility for changes in Consumer Confidence Index in Turkey. To the best of our knowledge, there is no previous study on this topic for an emerging market. Employing the E-GARCH method, we are able to validate day-of-the-week effect both in mean and volatility of the daily changes in the Consumer Confidence Index. In our findings, the mean equation exhibits only a Friday effect and the lowest volatility is also observed for Friday. Additionally, we use nonparametric stochastic dominance (SD) approach by employing several SD tests and verify the existence of Friday effects.

Keywords: Day-of-the-week effect, Consumer Confidence Index, volatility, E-GARCH, stochastic dominance

JEL classification: C14, C32, D11, D12

* Corresponding author: Çelik: Marmara University, Faculty of Economics, Department of Economics (Eng.), GoztepeYerleskesi, Kadikoy, 34722, Istanbul, Turkey. scelik@marmara.edu.tr.

Kaya: Bahcesehir University, Faculty of Economics and Administrative Sciences, Department of Economics, Besiktas, 34353, Istanbul, Turkey.

The authors would like to acknowledge the research grant provided by Marmara University Scientific Research Projects Unit under SOS-D-120309-0054.

1. Introduction

One of the well-known market anomalies, the day-of-the-week effect has been the subject matter of many studies in the literature. Most of these focussed on the analysis of stock returns while some were concerned with return on foreign exchange rates.

In one of the earliest studies, Cross (1973) showed that the mean return on Friday is higher than the mean return on Monday in the S&P 500 index during the period of 1953 to 1970. His finding established a foundation for the day-of-the-week effect in literature, in which many studies have questioned whether markets are inclined to rise towards the end of the week. Following Cross, many empirical studies found that the return on Monday is lower than the return on any other day of the week. In a recent attempt, Basher and Sadorsky (2006) examined the behaviour of 21 emerging stock markets by including unconditional and conditional risk factors. Their main finding was that there is a strong day-of-the-week effect in a few emerging stock markets even under the inclusion of conditional market risk.

On the other hand, there is also a vast literature which considers the return in foreign exchange rates rather than the return on stocks within the day-of-the-week effect theory. Among others, McFarland *et al.* (1982), Hilliard and Tucker (1992), Cornett *et al.* (1995), Aydogan and Booth (2003), Berument *et al.* (2007) concluded in favour of the day-of-the-week effect when they evaluated the depreciation of exchange rates. Lenkkeri *et al.* (2006) extended the research to European securitised real estate index returns. They found that eight out of eleven European countries exhibit the highest rates of return on Friday.

While the day-of-the-week effect in financial markets is regarded as an anomaly (Baker *et al.* 2008) or as a puzzling empirical finding (Wang *et al.*, 1997) it has hardly been rationalised theoretically. One argument is that firms tend to release bad news toward or after the close on Friday because of the fear of panic selling which could cause negative returns on Monday. Damodaran (1989) shows that the delayed bad news hypothesis can explain only a small portion of the negative returns on Monday. None of the other arguments such as delay between trading and settlement periods (Lakonishok and Levi, 1982), measurement errors in stock prices (Keim and Stambaugh, 1984) or institutional aspects of stock markets (Flannery and Protopapadakis, 1988) have been capable of fully explaining the day-of-the-week effect in financial markets in a conceptualised nature. Nonetheless, this anomaly exists and the literature attributes its presence to the dynamics of the financial markets.

Previous research has shown the close relationship between financial markets and the day-of-the-week effect, and also the close relationship between financial markets and Consumer Confidence Indexes (Mishkin, 1978; Gunes and Çelik, 2010). However, the link between the day-of-the-week effect and Consumer Confidence Index has not been examined.

The present study tries to contribute to the literature by analysing the day-of-the-week effect on both the mean and volatility for daily changes in the CBNC-e Consumer

Confidence Index (CCI) in Turkey. One should expect that changes in daily consumer confidence should be sensitive to daily changes in financial variables through the wealth effect. Indeed, as previous studies already demonstrated that exchange rate markets and the stock markets exhibit day-of-the-week effect in Turkey (Basher and Sadorsky, 2006, Berument *et al.*, 2007) it is intriguing to claim that anomalies in financial variables may pollute consumer confidence. It is possible to detect such a spill over effect by an analysis of the day-of-the-week effect in the CCI. It will also stand as evidence for the dynamic nature of expectation formation and decision making processes in emerging markets, where participants act relatively faster than developed and deep financial markets.

Our empirical findings demonstrate that daily changes in the CNBC-e Consumer Confidence Index exhibits the day-of-the-week effect. Additionally, we find that there is a day-of-the-week effect on the variability of the index using both parametric and nonparametric methodologies. The rest of this study is organised as follows: Section 2 includes a brief discussion on the Consumer Confidence Index in the Turkey; Section 3 contains the data and methodology; Section 4 includes the empirical findings; and, Section 5 concludes.

2. Consumer Confidence Index for Turkey

In Turkey, there are two consumer confidence indices: the CNBC-e Consumer Confidence Index and CBRT-TURKSTAT Consumer Confidence Index. In this study, we use the CNBC-e Consumer Confidence Index for the empirical analysis as we are able to obtain its daily data set from the survey provider. The index is composed of the following questions:

1. We would like to learn about your current economic situation. How would you compare your (and your family's) current financial situation with last year?

Answer choices: Better Worse Same No idea

2. What do you think your (and your family's) future financial situation will be in a year?

Answer choices: Better Worse Same No idea

3. How would you compare your current expectations of the Turkish economy with the previous month?

Answer choices: Better Worse Same No idea

4. How do you think the Turkish economy's situation will be in a year?

Answer choices: Better Worse Same No idea

5. Do you think that the current period is a good time to buy durable consumer goods such as a TV set, a refrigerator and furniture, or vehicles or a residence?

Answer choices: Good time Bad time No idea

The index is calculated according to following formula:

$$\text{Index value} = (\text{Current period value} / \text{Base period value}) * 100$$

The current period's value for each question is being calculated as:

$$([(\text{Number of optimistic answers for the question} - \text{Number of pessimistic answers for the question} / \text{Number of surveys}) * 100) + 100$$

The current period values of each question are summed up to obtain the current period's value for the overall index. The methodology used to compile and to calculate the index has been adopted from the Michigan University Index of Consumer Sentiment. The base period of the index is set as January 2002 and the value of the index at this period is 100. The index has a point of scale ranging from 0 to 200. The sample used to collect survey data is chosen from a database maintained by the survey provider. The database contains records of approximately 15 million individuals. The index is a compilation of 720 completed surveys. The survey data is obtained from the respondents between the 27th day of the previous month and the 26th day of the current month. The distribution of the completed surveys meets seven criteria:

1. 70% are selected from Istanbul, Ankara and Izmir; 30% are selected from other cities and large districts in Turkey.
2. 60% are selected from the 36-55 age group; 40% are selected from the 18-35 age group.
3. 50:50 distribution of male and female respondents.
4. 50% of the total surveys are composed of new records.
5. A minimum of 30% of new records belongs to individuals who had been successfully surveyed in the previous month.
6. A maximum 20% of the 720 completed surveys may be composed of additional respondents and these respondents are not called again in the following month.
7. One of the most important criteria is that respondents are not surveyed more than twice. This helps to minimise the biases in the answers of the respondents.

3. Data and methodology

We use data from the CNBC-e Consumer Confidence Index for the period of 1 January 2003 to 25 July 2008. The CCI is calculated using the data obtained from the CNBC-e Consumer Confidence Index survey provider. Change in the daily index is calculated as follows:

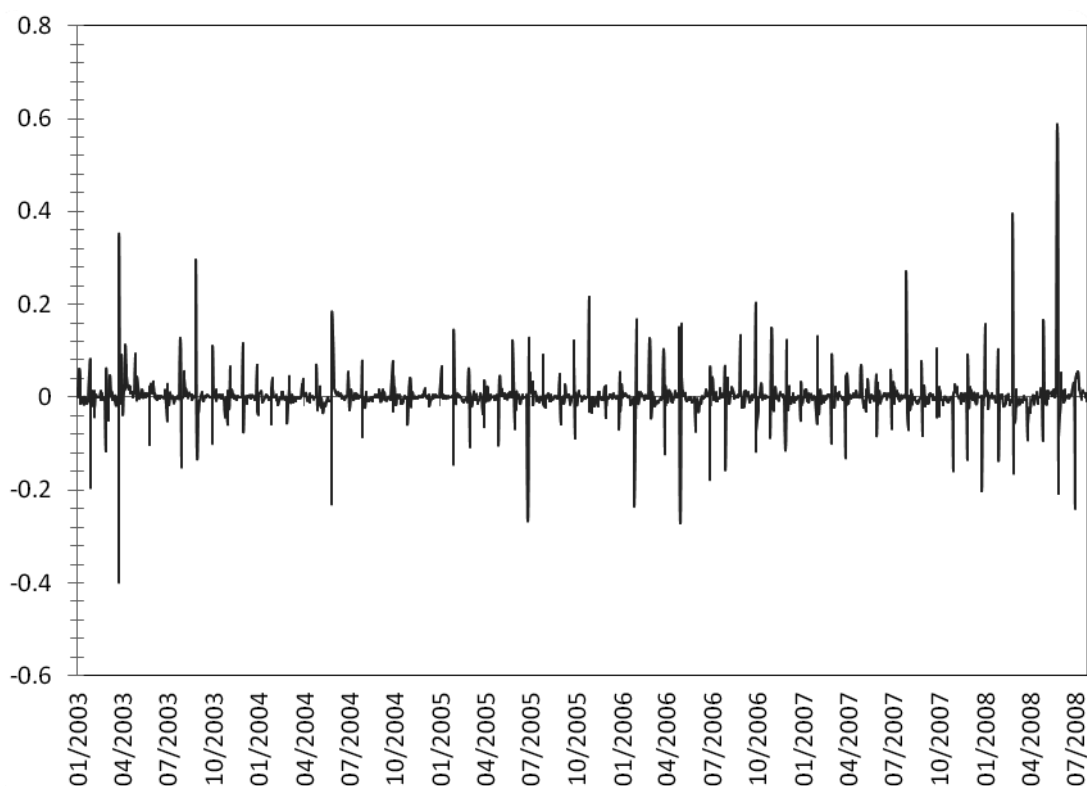
$$z_t = \ln\left(\frac{I_t}{I_{t-1}}\right)$$

where z is the daily changes in CCI and I stands for the CCI index value.

Table 1 summarises the main descriptive statistics of the CCI for the days of week. Kurtosis for all days of the week is very high and therefore the null hypothesis of normality is rejected. Friday and Tuesday have the lowest mean value, whereas Monday has the highest. The daily change on Monday and Tuesday has the lowest and highest standard deviations, respectively.

Table 1. **Descriptive statistics**

	Monday	Tuesday	Wednesday	Thursday	Friday
Mean	0.0025	-0.0009	0.0004	-0.0023	-0.0009
Median	0.0000	0.0000	0.0000	0.0000	0.0000
Maximum	0.1491	0.4604	0.3323	0.2961	0.2395
Minimum	-0.3105	-0.2739	-0.4983	-0.2579	-0.3157
Std. Dev.	0.0344	0.0506	0.0461	0.0440	0.0376
Skewness	-1.4587	1.4092	-3.2903	0.1153	-2.3254
Kurtosis	30.285	31.706	61.290	24.343	35.840
Jarque-Bera	9.004	9.984	41.436	5.505	13.293
Probability	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	287	288	289	290	290

Figure 1. **Daily change of the Consumer Confidence Index**
(1 January 2003-25 July 2008)

3.1. Ordinary least squares analysis

To investigate the day-of-the-week effect on daily changes in the CCI, first we employ the following regression model by employing the ordinary least square (OLS) technique:

$$z_t = \beta_1 D_{Mt} + \beta_2 D_{Tt} + \beta_3 D_{Wt} + \beta_4 D_{THt} + \beta_5 D_{Ft} + \sum_{i=1}^4 \beta_{5+i} z_{t-i} + \varepsilon_t \quad (1)$$

where z_t is the daily change in the Consumer Confidence Index, D_{Mt} , D_{Tt} , D_{Wt} , D_{THt} and D_{Ft} are dummy variables for Monday, Tuesday, Wednesday, Thursday and Friday respectively,

and ε_t is the error term. We include a one week delay lag for daily changes in the CCI to overcome the problem of autocorrelation.

3.2. E-GARCH method

Another serious problem the data set could have is inconstant variance of residuals and most probably time varying. ARCH models, introduced by Engle (1982), are proposed to correct the variability of error variance, by modelling error variances as a function of past errors. In this study, we employ the Exponential GARCH (E-GARCH) to analyse the day-of-the-week effect. The E-GARCH model was introduced by Nelson (1991) and is widely employed to analyse financial variables. One of the main advantages of the EGARCH is that unlike the GARCH model, it does not impose non-negativity constraints on parameters. On the other hand, the EGARCH allows for asymmetry in responsiveness to shocks.

The generalised structure of the E-GARCH model, including dummy variables, uses the following:

Mean equation:

$$z_t = \beta_1 D_{Mt} + \beta_2 D_{Tt} + \beta_3 D_{Wt} + \beta_4 D_{THt} + \beta_5 D_{Ft} + \sum_{i=1}^4 \beta_{5+i} z_{t-i} + v_t \quad (2a)$$

where v_t denotes the heteroskedastic error term, and

Variance equation:

$$\begin{aligned} \log(\sigma_t^2) = & \alpha_1 D_{Mt} + \alpha_2 D_{Tt} + \alpha_3 D_{Wt} + \alpha_4 D_{THt} + \alpha_5 D_{Ft} \\ & + \sum_{i=1}^q \beta_i \log(\sigma_{t-i}^2) + \sum_{j=1}^p \delta_j \left| \frac{v_{t-i}}{\sigma_{t-i}} - E\left(\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right) \right| + \sum_{k=1}^r \lambda_k \frac{v_{t-k}}{\sigma_{t-k}} \end{aligned} \quad (2b)$$

In Equation (2b), the left hand side is the logarithm of conditional variance of error term. This model assumes that the leverage effect is exponential. Since the model is specified in logarithms, the conditional variances are constrained to be non-negative.

3.3. Stochastic dominance approach

In order to strengthen the finding of an existence of the day-of-the-week effect in the CCI, we also perform a non-parametric stochastic dominance approach which uses the realised distribution of change in the CCI. There are three main types of stochastic dominance (SD): first order SD (FSD); second order SD (SSD); and, third order SD (TSD).

SD results imply a hierarchy in which the FSD implies SSD and in turn the SSD implies TSD. However, the reverse order is not true. To analyse the day-of-the-week effect in daily changes in the CCI by using the SD approach we employed three different SD tests. The first proposed by Anderson (1996), the second by Davidson and Duclos (2000) and the third by Barrett and Donald (2003). The Anderson (1996) SD tests are based on t-statistics comparing the object calculated in two independent samples. On the other hand, the Davidson and Duclos (2000) SD test is based on tests of inequality constraints which in turn lead to small number comparisons that are made by choosing the fixed number of arbitrary points. The Barrett and Donald (2003) SD test is a Kolmogorov-Smirnov type test that makes comparisons at all points.

4. Empirical results

Table 2 depicts the OLS results of Equation (1). The findings of the OLS estimation indicate that none of the daily dummy variables are significant. We performed a Wald coefficient restriction test to see whether the daily dummy variables are jointly significant but we cannot reject the null hypothesis of the dummy variable's joint insignificance.¹ As a result, the OLS estimation shows that there is no-day-of-the-week effect on daily changes in the Consumer Confidence Index. The Breusch-Godfrey Serial Correlation Lagrange Multiplier (LM) test with four lags and the Ljung-Box Q show that there is no serial correlation. This result assumes the existence of a constant variance. To check whether conditional variance has time varying properties, we performed the Lagrange Multiplier Autoregressive Conditional Heteroskedastic test² as suggested by Engle (1982) using four lags³. The calculated test statistic, which is observed as R^2 and comes from regressing the squared residuals on its first four lags, is 12.99; higher than the critical value of 9.48 at 5% significance level. Hence the null hypothesis of the constant conditional variance is rejected and we need to take varying conditional variance into account.

Table 2. Ordinary least square results

Variable	Coefficient	Std. error	t-statistic	Prob.
D_M	0.0016	0.0024	0.6442	0.5196
D_T	-0.0008	0.0024	-0.3154	0.7525
D_W	0.0004	0.0024	0.1483	0.8821
D_{TH}	-0.0021	0.0024	-0.8963	0.3703
D_F	-0.0015	0.0024	-0.6172	0.5372
Z_{t-1}	-0.2849	0.0263	-10.8102	0.0000
Z_{t-2}	-0.1686	0.0272	-6.1898	0.0000
Z_{t-3}	-0.1057	0.0272	-3.8835	0.0001
Z_{t-4}	-0.0706	0.0263	-2.6815	0.0074

Notes: Breusch-Godfrey LM test statistic: 4.08 (0.39).

Ljung-Box Q(36): 14.71 (0.99).

ARCH LM Test: 10.99 (0.02).

By modelling the conditional variance of equation (1) as an E-GARCH (1,1) process, we estimate the mean equation (2a) and variance equation (2b) jointly. In the estimation of variance equation, we exclude the Wednesday dummy variable to fix the dummy trap problem. Table 3 shows the results of the E-GARCH model.

Estimates of the mean equation indicate that there is a day-of-the-week effect. All dummy variables except Friday are insignificant. We also tested the coefficients of D_{Mt} , D_{Tt} , D_{Wt} and D_{THt} jointly and found that joint effect is also insignificant.⁴ Only Friday is significant and has a negative effect on the daily change of the CCI. This result shows that consumers become pessimistic as the weekend approaches.⁵ One possible reason for this rather interesting empirical finding is that since Friday is the last working day of the week, respondents bear the whole week's cumulative fatigue (mental and physical) which may depress the survey participants. Further empirical research is needed in order to reach rather well-built conclusions.

Estimation results of the variance equation show that the coefficients of three of the four dummy variables: Monday, Thursday and Friday, are significant. This finding shows the existence of the day-of-the-week effect for the volatility equation. Among others, Friday has the lowest volatility. To check whether the Wednesday dummy variable is

Table 3. **Results of the E-GARCH method (1,1)**

Variable	Coefficient	Std. error	z-statistic	Prob.
D_M	0.0024	0.0016	1.4989	0.1339
D_T	0.0039	0.0028	1.3682	0.1713
D_W	0.0051	0.0036	1.4376	0.1505
D_{TH}	0.0019	0.0021	0.9342	0.3502
D_F	-0.0042	0.0014	-3.0079	0.0026
Z_{t-1}	-0.0120	0.0519	-0.2306	0.8176
Z_{t-2}	-0.0832	0.0334	-2.4905	0.0128
Z_{t-3}	-0.0825	0.0335	-2.4632	0.0138
Z_{t-4}	-0.0491	0.0306	-1.6055	0.1084
Variance equation				
δ	0.5076	0.0418	12.1298	0.0000
λ	0.1160	0.0330	3.5178	0.0004
β	0.0293	0.0308	0.9513	0.3415
α_1	-0.6082	0.0504	-12.0638	0.0000
α_2	0.0505	0.0465	1.0864	0.2773
α_3	-0.2448	0.0487	-5.0242	0.0000
α_4	-0.7508	0.0371	-20.2491	0.0000

Note: Ljung-Box Q (36): 21.211 (0.976).

Table 4. **Stochastic dominance test results**

Test		M against F	T against F	W against F	Th against F
Anderson (1996)	FSD	0.438	0.435	0.582	0.131
	SSD	0.402	0.196	0.285	0.040
	TSD	0.361	0.179	0.237	0.044
Davidson and Duclos (2000)	FSD	0.459	0.428	0.561	0.110
	SSD	0.639	0.086	0.600	0.438
	TSD	0.393	0.127	0.513	0.521
Barrett and Donald (2003)	FSD	0.450	0.403	0.549	0.450
	SSD	0.735	0.013	0.396	0.186
	TSD	0.689	0.004	0.281	0.260

Notes: Monday (M), Tuesday (T), Wednesday (W), Thursday (Th), Friday (F).

For Anderson (1996), we report *p-values* for simulated and 10 evolution points. For Davidson and Duclos (2000), we report *p-values* of the Wald test and 10 evolution points. For Barrett and Donald (2003), we report *p-values* of the KS1 model.

significant, we performed the test by replacing Tuesday with Wednesday and conclude that Wednesday is also insignificant.

The final empirical part to strengthen the findings for Friday employed the nonparametric stochastic dominance approach. If Friday had the significantly lowest daily change, we would be able to find that all days dominate Friday in some order. The columns show the *p-values* of all tests considered in this study for the daily change comparisons under the null hypothesis that Monday (M), Tuesday (T), Wednesday (W), Thursday (Th) dominate Friday (F).

The main finding of our test results is that all days dominate Friday in first order. Both the Anderson (1996) SD test and Davidson and Duclos (2000) SD test suggest that Thursday and Tuesday cannot dominate Friday at the second and third order, respectively. However, as the SD results imply hierarchy, we conclude that all other days of the week dominate Friday, strengthening our parametric results with the non-parametric stochastic dominance approach results.

5. Conclusion

This paper examines the day-of-the-week effect on daily changes in Consumer Confidence Index and their volatility. We started by employing OLS and found no day-of-the-week effect. Considering the time varying variance, we assessed the day-of-the-week effect with the E-GARCH specification. Our estimates of the mean equation show that daily changes of the Consumer Confidence Index in Turkey exhibit the Friday effect. Moreover, estimates of the variance equation indicate that there is a day-of-the-week effect on the variability of the confidence index. Finally, we employed several nonparametric SD tests and verified the Friday effect. These results validate the intrinsic nature of expectation formation, which is characterised by economic, cultural, psychological and sociological factors.

Notes

1. F-statistics and the corresponding probability of the Wald test statistic are 0.34 and 0.886 respectively.
2. We employ the White, Harvey and Breusch-Pagan-Godfrey heteroscedasticity tests and found that there is no statistical evidence of significance from these types of heteroscedasticity test. We did however find that the conditional variance is varying because it is affected by its past values.
3. Up to five lag order results are the same. For a higher order up to eight, the null hypothesis is rejected at 10% significance level.
4. F-statistics and the corresponding probability of Wald test statistic are 1.42 and 0.223 respectively.
5. It is understood that the day-of-the-week effect may bias the Consumer Confidence Index due to seasonality, however this possibility is not within the scope of our research.

References

- Anderson, G. (1996), "Nonparametric tests for stochastic dominance", *Econometrica*, Vol. 64, pp. 1183-93.
- Aydogan, K. and G. Booth (2003), "Calendar anomalies in the Turkish foreign exchange markets", *Applied Financial Economics*, Vol. 13, pp. 353-60.
- Baker, H.K., A. Rahman and S. Saadi (2008), "The day-of-the-week effect and conditional volatility: Sensitivity of error distributional assumptions", *Review of Financial Economics*, Vol. 17, pp. 280-95.
- Barrett, G.F. and S.G. Donald (2003), "Consistent tests for stochastic dominance", *Econometrica*, 71, 71-104.
- Basher, S.A. and P. Sadorsky (2006), "Day-of-the-week effects in emerging stock markets", *Applied Economic Letters*, Vol. 13, pp. 621-28.
- Berument, H., M.N. Coskun and A. Sahin (2007), "Day-of-the-week effect on foreign exchange market volatility: Evidence from Turkey", *Research in International Business and Finance*, Vol. 21, pp. 87-97.
- Cornett M.M., T.V. Schwarz and A.C. Szakmary (1995), "Seasonalities and intraday return patterns in the foreign currency futures market", *Journal of Banking and Finance*, Vol. 19, pp. 843-69.
- Cross, F. (1973), "The behavior of stock prices on Fridays and Mondays", *Financial Analyst Journal*, November-December, Vol. 29, pp. 67-9.
- Damodaran, A. (1989), "The weekend effect in information releases: A study of earnings and dividend announcements", *Review of Financial Studies*, Vol. 2, pp. 607-23.
- Davidson, R. and Y.J. Duclos (2000), "Statistical inference for stochastic dominance and for the measurement of poverty and inequality", *Econometrica*, Vol. 68, pp. 1435-64.
- Engle, R. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation", *Econometrica*, Vol. 50, pp. 987-1007.
- Flannery, M.J. and A.A. Protopapadakis (1988), "From T-bills to common stocks: Investigating the generality of intra-week return seasonality", *Journal of Finance*, Vol. 33, pp. 431-450.

- Gunes, H. and S. Çelik (2010), "Consumer confidence and financial market variables in an emerging market: The case of Turkey", *International Research Journal of Finance and Economics*, Vol. 48, pp. 169-86.
- Hilliard, J.E. and A.L. Tucker (1992), "A note on weekday, intraday, and overnight patterns in the interbank foreign exchange and listed currency options markets", *Journal of Banking and Finance*, Vol. 16, pp. 1159-71.
- Keim, D.B. and R.F. Stambaugh (1984), "A further investigation of the weekend effect in stock returns", *Journal of Finance*, Vol. 39, pp. 819-40.
- Lakonishok, J. and M. Levi (1982), "Weekend effects on stock returns: A note", *Journal of Finance*, Vol. 37, pp. 883-89.
- Lenkkeri, V., W. Marquering and B. Strunkmann-Meister (2006), "The Friday effect in European securitized real estate index returns", *Journal of Real Estate Finance and Economics*, Vol. 33, pp. 31-50.
- Mishkin, F.S. (1978), "Consumer sentiment and spending on durable goods", *Brookings Papers on Economic Activity*, Vol. 1, pp. 217-32.
- Nelson, D.B. (1991), "Conditional heteroskedasticity in asset returns: A new approach", *Econometrica*, Vol. 59, pp. 347-70.
- Nelson, D.B. and C.Q. Cao (1992), "Inequality constraints in the univariate GARCH model", *Journal of Business and Economic Statistics*, Vol. 10, pp. 229-35.
- Wang, K., Y. Li and J. Erickson (1997), "A new look at the Monday effect", *Journal of Finance*, Vol. 52, pp. 2171-186.

Constructing coincident and leading indices of economic activity for the Brazilian economy

by

João Victor Issler, Hilton Hostalacio Notini and Claudia Fontoura Rodrigues*

This paper has three original contributions. The first is the reconstruction effort of the series of employment and income to allow the creation of a new coincident index for the Brazilian economic activity. The second is the construction of a coincident index of the economic activity for Brazil, and from it, (re) establish a chronology of recessions in the recent past of the Brazilian economy. The coincident index follows the methodology proposed by The Conference Board (TCB) and it covers the period 1980:1 to 2007:11. The third is the construction and evaluation of many leading indicators of economic activity for Brazil which fills an important gap in the Brazilian Business-Cycle literature.

Keywords: Coincident and Leading Indicators, Business Cycles, Common Features, Latent Factor Analysis

JEL codes: C32, E32

* Corresponding author: João Victor Issler, Graduate School of Economics – EPGE, Getulio Vargas Foundation, Praia de Botafogo 190, s. 1100, Rio de Janeiro, RJ 22250-900, Brazil.

We gratefully acknowledge the comments and suggestions of Marcelle Chauvet, Kajal Lahiri, Paulo Picchetti, the Editor in Charge (Michael Graff), two anonymous referees, and participants of CIRET 2008 in Santiago, Chile. All errors are ours. We thank Marcia Waleria Machado and Rafael Burjack for excellent research assistance and thank CNPq-Brazil, FAPERJ and INCT for financial support.

1. Introduction

An important concern of any modern society is: what is the current “state” of the economy and what should be the state of the economy in the near future? Entrepreneurs and individuals are interested in this question because their profits and welfare are a function of it. Governments also have an interest in the subject for budgetary and welfare issues. Unfortunately, no one possesses a series that represents the “state of the economy” because it is a latent variable, or rather, it is unobservable.

Since Burns and Mitchell (1946), there has been a great deal of interest in making inferences about the “state of the economy” from sets of monthly variables that are believed to be either concurrent or that lead the economy’s business cycles (the so called “coincident” and “leading” indicators, respectively). The most educated estimate of US turning points is embodied in the binary variable announced by the NBER Business Cycle Dating Committee. These announcements are based on the consensus of a panel of experts and they are made some time (usually six months to one year) after the time of a turning point in the business cycle. The NBER summarises its deliberations as follows:

“The NBER does not define a recession in terms of two consecutive quarters of decline in real GNP. Rather, a recession is a recurring period of decline in total output, income, employment, and trade, usually lasting from six months to a year, and marked by widespread contractions in many sectors of the economy.”

www.nber.org/cycles

The time it takes for the NBER committee to deliberate and decide that a turning point has occurred is often too long to make these announcements practically useful. This gives importance to two constructed indices, namely the coincident index and the leading indicator index. The traditional coincident index constructed by the Department of Commerce is a combination of four representative monthly variables on total industrial production, income, employment and sales. The Conference Board Index (TCB) uses a simple average of the standardised differenced (logged) series, which is a way of treating equally the fluctuations of all four series in computing the index. The TCB approach is somewhat heuristic, since it requires no estimation of a formal econometric model. Despite that, it works surprisingly well in practice for the US; see the comparison in Issler and Vahid (2006) using the TCB index and alternative econometric-based indices in trying to replicate the NBER dating decisions.¹

As an alternative to heuristic methods such as TCB, several authors have proposed methods of building indices supported by sophisticated econometric and statistic techniques. Stock and Watson (1998a, 1998b, 1989, 1993a, 1993b) were the first to apply the tools of modern time-series econometrics to build an approach able to construct leading and coincident indices, detect turning points of economic activity, and to predict the probability of a recession.

An important empirical drawback of Stock and Watson’s approach was its failure to detect the US recession in 1990-1991. Many papers tried to improve on Stock and Watson’s method, while keeping the formal building block of a structural econometric model (see

e.g., Chauvet, 1998 and Mariano and Murasawa, 2003). More recently, Giannone, Reichlin and Small (2008) have developed a framework for real-time nowcasting (and forecasting) on the basis of a large and unbalanced dataset. Their model provided a timely and up to date estimation of the state of the economy. Prior to that, and using a different approach, Evans (2005) incorporated daily information to update the real-time estimates of GDP and inflation in a nowcasting exercise.

In Brazil, with the exception to the work of Contador (1977) and Contador and Ferraz (1999), research on coincident and leading indices of economic activity is fairly young and most of the literature dates from the 2000s. Chauvet (2001) and Picchetti and Toledo (2002) use common-factor models to generate a monthly coincident indicator of economic activity. Chauvet (2002) uses a two-state Markov Chain characterising a recession or an expansion to propose a chronology for Brazilian business cycles. On a broader study, Duarte, Issler and Spacov (2004) evaluated three candidates for composite coincident indices: The Conference Board's (TCB) index; Spacov's (2000) index, and Issler and Vahid's (2006) index. Using quadratic loss, the dating of these three indices was compared with that of a monthly proxy of Brazilian GDP, suggesting that the Brazilian coincident index should use the methodology put forth by TCB. Similarly, Hollauer, Issler and Notini (2009) found that the TCB index perform best when compared to the methods of Mariano and Murasawa (2003) and of Stock and Watson (1989, 1993) when industrial-sector coincident indicators were evaluated.

Unfortunately, part of this recent research effort in Brazil came to a halt because of the recent redesign of the official employment survey conducted by IBGE – Monthly Employment Survey (*Pesquisa Mensal do Emprego*) – which provides monthly Brazilian data on employment and labour income. Indeed, the change in the survey design in 2002 was so drastic that it eliminates long-span time-series on employment and income, which are crucial series for business-cycle research using TCB and NBER oriented methods.

The first goal of this paper is to resume business-cycle research in Brazil using these methods, which proved to be valuable after the empirical results of Duarte, Issler and Spacov (2004) and Hollauer, Issler and Notini (2009). Indeed, one of the main challenges of Brazilian business-cycle research is to back-cast currently available income and employment series to be able to form a long enough coincident index using TCB method – which component series are industrial production, sales, income and employment. Here, we devote a great deal of effort to reconstruct the employment and income series using a novel state-space representation, which is explained at some length. Our proposed back-casting approach is based on the interpolation method originally proposed by Bernanke, Gertler and Watson (1997) and later improved by Mönch and Uhlig (2005). It is a very flexible setup that allows the estimation of a wide range of models in the time-series literature. As usual, estimation of the unobserved components in these models is performed employing the Kalman filter.

Once we obtain a long enough span of the usual series used in TCB method, we compute a new composite coincident index of Brazilian economic activity. Its dating of recessions is compared with those of Duarte, Issler and Spacov (2004) and with those implied by the monthly GDP estimate computed by Issler and Notini (2008). It is also compared with the dating properties of alternative techniques delivering a long-span coincident index – like chaining new and old employment and income series, for example. This comparison shows that not only our proposed technique has superior results to

chaining old and new series but also that our dating of recessions is in line with past experience.

Our last contribution is regarding the construction of leading indices of economic activity to track the composite coincident index proposed here. Although coincident indices have been relatively well studied in Brazil, leading indices have not. In constructing leading indices we take into account three interesting and novel features in Brazilian business-cycle research: i) we compare leading and coincident indices in sample and out-of-sample, where recursive methods are used to mimic the actual tracking problem encountered in practice; ii) we consider using Granger (1969) causality tests, as well as novel alternative criteria in choosing candidate series to be included in leading composite indices; and, iii) we investigate the ability of survey-based time series to lead our composite index.

Although comparisons are based on a variety of features of the dating properties of these different indices, our decision to validate the current composite index is mostly based on a variant of the Quadratic Probability Score (QPS) quadratic-loss statistic proposed for that purpose by Diebold and Rudebusch (1989).

Empirical results obtained here are compared with the previous literature on Brazil. In evaluating different results and techniques used in constructing coincident and leading indices, we borrow from the almost century-long debate on this issue that has been present in the US economy, and a similar half-century or older debate in Europe.

One of the aspects of the methodology employed in this paper is the heavy use of statistical and econometric tools in devising different measures of economic activity. This is done in detriment of the use of a theoretical-based approach. Indeed, there is an old macroeconomics debate which started after Koopmans' (1947) critique of the lack of theory behind the NBER business cycle methodology – labelled “measurement without theory”.² As stressed by Auerbach (1981), if the success of a specific approach to economic analysis can be measured by its longevity and continued use under a variety of environments, then the use of measurement of economic activity must be paramount.

Still, Koopmans compares the early empirical NBER methodology to the “Kepler stage” of science, whereas the “Newton stage” develops theoretical models explaining these measured phenomena. He seems to prefer the latter as if nothing is gained by the former. We disagree with this view. Measurement establishes “stylised facts” if done properly, which then establishes what theoretical models should aim to explain. Indeed, the main paragraph in the classic book by Burns and Mitchell (1946) is subsequently cited by most modern theoretical work in economics, *e.g.*, Lucas (1977). Also, generating the business-cycle regularities found by Burns and Mitchell and their followers now constitutes a basic goal of the recent theory of business cycles. The core of this literature and its main shortcomings is surveyed in Kocherlakota (2010), for example. Indeed, science benefits from both measurement and theory, since they complement each other in giving a broad view of a specific phenomenon. This shows that the motivation to use statistical methods is neither an end on itself nor because they are available, but simply because they will finally help in uncovering and/or answering an interesting economic question.

This article is organised as follows: Section 2 contains a brief review of the Conference Board method. Section 3 presents the Kalman filter model. Section 4 presents the data and the main results. Section 5 concludes.

2. The methodology of TCB

The ideas behind the TCB method are twofold: simplicity and robustness. Simplicity is used because they weight information in coincident and leading indices equally, once one controls for the fact that different signals carry different information depending on their variance. One simple way to treat every series equally in this context is to standardise them, treating the standardised series equally. Robustness comes into play here, since standardising is a way to robustify the series used in the econometric analysis.

The coincident series is an equally-weighted linear combination of four coincident series (income (I_t), output (Y_t), employment (N_t), and sales (S_t)), once we control for the fact that the growth rate of these series have different variances. Hence, in its bear essentials,³ the coincident indicator uses weights constructed as:

$$\Delta \ln(CI_t) = \frac{1}{4} \left[\frac{\Delta \ln(I_t)}{\sigma_{\Delta \ln(I)}} + \frac{\Delta \ln(Y_t)}{\sigma_{\Delta \ln(Y)}} + \frac{\Delta \ln(N_t)}{\sigma_{\Delta \ln(N)}} + \frac{\Delta \ln(S_t)}{\sigma_{\Delta \ln(S)}} \right], \quad (1)$$

where $\sigma_{\Delta \ln(I)}$, $\sigma_{\Delta \ln(Y)}$, $\sigma_{\Delta \ln(N)}$, and $\sigma_{\Delta \ln(S)}$ are respectively the standard deviations of income, output, employment, and sales growth. It is straightforward to construct the level series $\ln(CI_t)$ or CI_t once we possess $\Delta \ln(CI_t)$.

The leading series are usually chosen because they have turning points that happen before those of the level series $\ln(CI_t)$ or CI_t . To determine that, we first need a definition of “turning points” and of “before”. In this literature, turning points are usually determined using an accepted algorithm for turning points or local minima and maxima of a time series – the Bry-Boschan algorithm, Bry and Boschan (1971). With turning points of the target variable and of the potential leading series in hand, all we have to determine is whether those of the potential leading series precede those of the target series. Leading series are those that, on average, downturn or upturn prior to the target series. Once we determine the candidates of leading series, all we have to do is to combine them. Again, the TCB methodology uses simplicity and robustness: all leading series are combined using a procedure similar to (1).

3. Back-casting using the Kalman filter

In this section, we give a brief review of the Kalman filter model applied to back-cast two of our coincident series – employment and income. A detailed description of this technique can be found in Harvey (1989) or in Hamilton (1994). After that, we discuss how to back-cast series using a novel state-space representation.

Our starting point in using the Kalman filter to back-cast employment and income is the interpolation technique of Bernanke, Gertler and Watson (1997) and the extensions made by Mönch and Uhlig (2005). In both papers, the Kalman filter is used to interpolate GDP from quarterly to monthly frequency, where monthly auxiliary series help in estimating monthly GDP. It is assumed that unobserved monthly GDP (labelled as y_t^+ here) follows an AR (p) process explained by “pre-determined” regressors x_t and an AR (1) error term. The term x_t contains co-variates, which should have a high correlation with the series being interpolated: much of the contemporaneous behavior of the interpolated series comes from them. Also in x_t are deterministic series such as a constant and/or

seasonal dummies, which, together with these co-variates, explain the behavior of y_t^+ . The model for y_t^+ is:

$$\begin{aligned} (1 - \phi_1 L - \dots - \phi_p L^p) y_t^+ &= x_t \beta + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t. \end{aligned} \quad (2)$$

Observed quarterly GDP (labelled as y_t here) is:

$$y_t = \sum_{i=0}^2 y_{t-i}^+, \quad t = 3, 6, 9, 12, \dots \quad (3)$$

$$y_t = 0, \quad \text{otherwise.} \quad (4)$$

Hence, quarterly GDP, which we can only observe on months $t = 3, 6, 9, 12$, etc., is the sum of the corresponding monthly GDPs in that quarter.⁴ Otherwise, it is just set to a fictional value of zero. Notice that setting $y_t = 0$ for the months we do not observe GDP is a way of making quarterly GDP observable at the monthly frequency (see Appendix in Mönch and Uhlig, 2005). In the Kalman-filter literature for mixed frequency models (see, for example, Giannone, Reichlin and Small, 2008) a fictional value is usually assumed for missing observations (zero is the most frequent choice). The crucial step is to impose that the fictional observed data has a very large variance, so that the zero value is discounted and overwritten by the Kalman-filter technique. This is exactly how Mönch and Uhlig proceed.

A second issue is how to initialise the values of β , ρ , and the variance of u_t in the Kalman-filter procedure. Mönch and Uhlig aggregate the covariates in x_t from high (monthly) to low (quarterly) frequency, and run an OLS regression of y_t on its lags and x_t [regression (2)] at the quarterly frequency. This yields estimates of β and the ϕ 's, as well as an estimate for the variance of u_t . With the latter, an estimate of ρ is obtained from an OLS regression of u_t on u_{t-1} .

If we assume that the polynomial $(1 - \phi_1 L - \dots - \phi_p L^p)$ is of order one, i.e., $p = 1$, with coefficient ϕ , the state-space form of Mönch and Uhlig's problem is the following:

$$\xi_t = \begin{pmatrix} y_t^+ \\ y_{t-1}^+ \\ y_{t-2}^+ \\ u_t \end{pmatrix} = \begin{pmatrix} \phi & 0 & 0 & \rho \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1}^+ \\ y_{t-2}^+ \\ y_{t-3}^+ \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} x_t \beta \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \varepsilon_t \end{pmatrix} \quad (5)$$

$$y_t = H_t' \xi_t, \quad (6)$$

where (5) and (6) are respectively the state and the observation equations and the matrix H_t' is time-varying, with the following format:

$$H'_t = \begin{cases} \begin{bmatrix} 1 & 1 & 1 & 0 \end{bmatrix}, & t = 3, 6, 9, 12, \dots \\ \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, & \text{otherwise.} \end{cases} \quad (7)$$

One interesting feature of the approach in Mönch and Uhlig (2005) is that it encompasses several data interpolation models, summarised in Table 1.

Table 1. **Resulting model as a function of ϕ and ρ in (5)**

Model	ϕ	ρ
Static model in levels with IID residuals	0	0
Static model in levels with AR(1) residuals (Chow and Lin, 1971)	0	free
Static model in 1st differences with IID residuals (Fernandez, 1981)	0	1
Dynamic model in levels with IID residuals (Mitchell <i>et al.</i> , 2005)	free	0
Dynamic model in 1st differences with IID residuals	free	1
Dynamic model in levels with AR(1) residuals	free	free

To assess the quality of interpolation, Bernanke, Gertler, and Watson (1997) propose the use two R^2 measures of fit. Denoting by $\widehat{y}_{t|T}^+$ the smoothed estimate of monthly GDP, and by $\widehat{u}_{t|T}$ the same estimate of the error term u_t , they consider:

$$R_{\text{level}}^2 = \frac{\text{VAR}(\widehat{y}_{t|T}^+)}{\text{VAR}(\widehat{y}_{t|T}^+) + \text{VAR}(\widehat{u}_{t|T})}, \text{ and,}$$

$$R_{\text{diff}}^2 = \frac{\text{VAR}(\Delta \widehat{y}_{t|T}^+)}{\text{VAR}(\Delta \widehat{y}_{t|T}^+) + \text{VAR}(\Delta \widehat{u}_{t|T})}.$$

Bernanke, Gertler, and Watson (1997) and Mönch and Uhlig (2005) claim that it is more informative to report the R^2 in first differences since the same statistic in levels will always be close to unity.

We now adapt the state-space representation in (5) and (6) to the problem of back-casting a series for which we observe part of its realisations but not all, and where both the target variable as well as the covariates are in the same frequency (monthly, in our case). In some sense, this is very close to the problem worked out in Bernanke, Gertler and Watson, and Mönch and Uhlig. There, they only observe quarterly GDP for some but not all months of the year. Their solution was to set to zero the missing observations and to impose that the fictional observed data has a very large variance, so that this zero value is discounted and overwritten by the Kalman-filter technique. We follow their solution here, setting to zero all the missing observations to be back-cast. We also impose that the fictional data has a very large variance. Also, we initialise the values of β , ρ , and the variance of u_t in the Kalman-filter procedure by means of an OLS regression for the overlapping period, where we observe the variable being back-cast as well as the covariates used in back-casting it.

Differently from Bernanke, Gertler and Watson, and Mönch and Uhlig, here there is no need to aggregate from high to low frequency. We explain further our method below.

Suppose we possess a total of $t = 1, 2, \dots, T^*, \dots, T$, observations on x_t . However, for series y_t^+ , we only possess data from $t = T^* + 1, \dots, T$, with missing values from $t = 1, 2, \dots, T^*$. This is exactly our setup for income and employment in this paper. If we set the order of the polynomial $(1 - \phi_1 L - \dots - \phi_p L^p)$ to unity, i.e., $p = 1$, with coefficient ϕ , recalling that now we need not impose the time-aggregation restriction in (7), the state-space form of our problem collapses to the following:

$$\xi_t = \begin{pmatrix} y_t^+ \\ u_t \end{pmatrix} = \begin{pmatrix} \phi & \rho \\ 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1}^+ \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} x_t \beta \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_t \end{pmatrix} \quad (8)$$

$$y_t = H_t' \xi_t, \quad (9)$$

where (8) and (9) are respectively the state and the observation equations and the matrix H_t is time-varying, with the following format:

$$H_t = \begin{cases} \begin{bmatrix} 1 & 0 \end{bmatrix}, & t = T^* + 1, \dots, T \\ \begin{bmatrix} 0 & 0 \end{bmatrix}, & \text{otherwise.} \end{cases} \quad (10)$$

Notice that our back-cast problem is simpler than the interpolation problem, since we do not have to impose high- to low-frequency constraints in the data which are present in (5). The key to our problem lies in the choice for H_t in (10). Here, we make the latent variable y_t^+ identical to y_t for the periods in which the latter is observed (notice that there is no error term in (9)). This has two consequences. First, the algorithm will forecast y_t^+ to be identical to y_t for $t = T^* + 1, \dots, T$. Second, it will use the available data on employment (income) to estimate a model and will use this model to forecast the latent variable in the periods in which it is not observable, i.e., from $t = 1, 2, \dots, T^*$. Under correct specification, this model can produce the optimal forecasts of the latent variable consistent with all available future information. That will be simply given by the smoothed forecast of y_t^+ , i.e., by $\widehat{y_{tT}^+}$.

4. Empirical results

4.1. Data

An important part of this paper is the choice of the variables to be included in the coincident indicator. We follow the recent Brazilian experience: Duarte, Issler and Spacov (2004) and Spacov (2001). For output, labelled Y_t , we use industrial production, computed by Instituto Brasileiro de Geografia e Estatística (IBGE), and available from 1980:1. There is no long-span sales series in Brazil, we therefore follow Duarte, Issler and Spacov (2004) and use total production of corrugated paper in Brazil as a proxy for sales, labelled S_t , which is computed by Associação Brasileira de Papelão Ondulado (ABPO). Employment, labelled N_t , is given by the total number of persons – 10 years old or older – that have a job. It is extracted from the *Monthly Employment Survey* computed by IBGE. Income is proxied by the labour income series, labelled I_t , extracted from this same *Survey*.

The last two series – employment and income – are only available from 2003 on, due to a drastic redesign of the *Monthly Employment Survey*. Here, we back-cast these series using a state-space representation estimated using the Kalman filter.

4.2. The coincident series

As stressed above, one of the original contributions of this paper is to back-cast two of the coincident series for the Brazilian economy – income and employment. We used the techniques described in the previous section to back-cast them. In the current *Monthly Employment Survey*, income is available from 2002:2 on, while employment is available from 2002:3 on.

Back-casting was conducted in two steps. First we selected the co-variate series, which could potentially explain the variations of income or employment. These co-variables were then used in the state-space regression, which was estimated using the framework described above – based on the algorithm by Mönch and Uhlig (2005). Our setup allows for several different dynamic models to be estimated, all described in Table 1, depending on different values for the parameters ϕ and ρ .

We tested seven series as auxiliary regressors in the back-casting procedure, all available for the period 1980:1 to 2007:11. They are: industrial production, output in the process industry, corrugated paper production, car production, steel production, cement production, energy production, and the monthly real GDP series estimated by Issler and Notini (2008). The dependent variables and all co-variables entered in levels in the state space representation, which is estimated in all the six different versions described in Table 1. In addition to the co-variables listed above, our models also include eleven seasonal dummies. Therefore, there is no seasonal adjustment prior to back-casting.

In Table 2, we present the R_{diff}^2 measure of fit for each model described in Table 1.

Table 2. **Employment and income resulting R_{diff}^2 for each model**

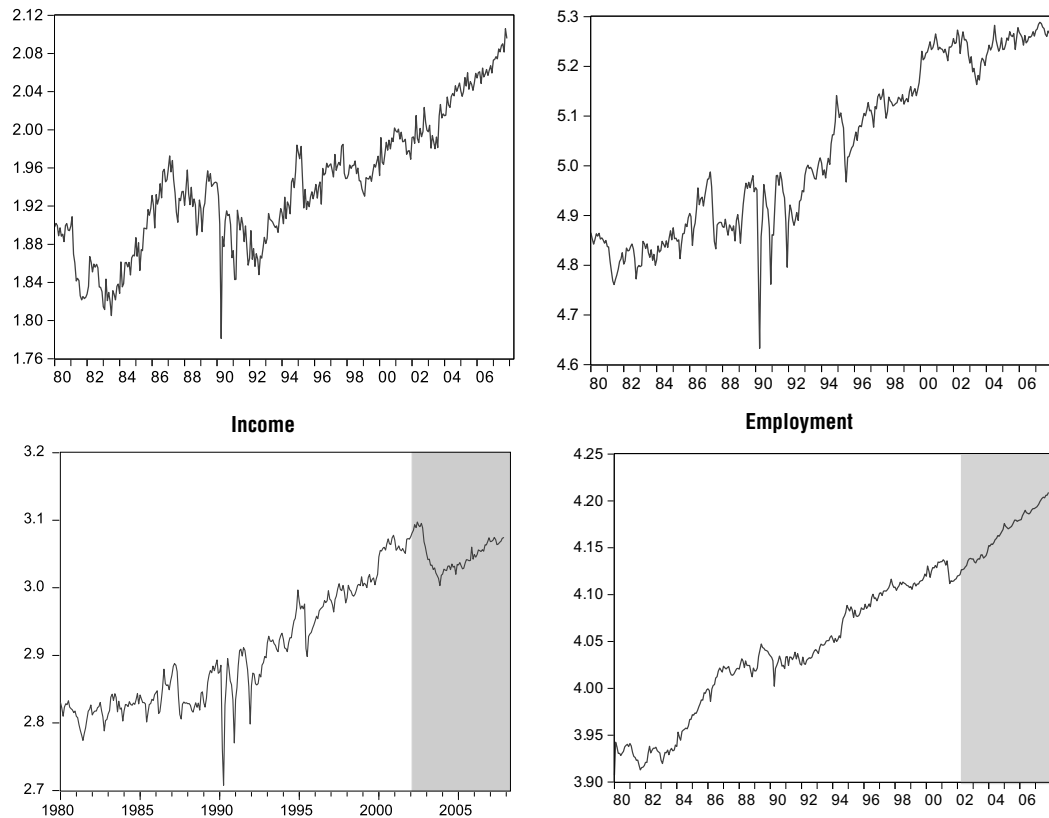
Model	Employment	Income
Static model in levels with IID residuals	0.4979	0.1134
Static model in levels with $AR(1)$ residuals	0.4729	0.0425
Static model in 1st differences with IID residuals	0.0072	0.0000
Dynamic model in levels with IID residuals	0.0597	0.0827
Dynamic model in 1st differences with IID residuals	0.0000	0.0000
Dynamic model in levels with $AR(1)$ residuals	0.0000	0.0048

Our final choice of auxiliary variables and models were as follows. For employment (in logarithms) we choose only the monthly GDP series and energy production (in logarithms) as co-variables. For income (in logarithms), we selected only the paper production series and cement production (in logarithms) as auxiliary variables. In both cases, the model with the highest R_{level}^2 and R_{diff}^2 was the static model with i.i.d. errors.

Once the back-casting procedure was implemented, all coincident series were seasonally adjusted using the X-12 procedure.⁵ The results are plotted below including the results of the back-cast series. For income and employment, the shaded areas in the graphs below depict the actual sample in which we observe them.

All four coincident series were tested for unit roots. We used three different tests. On a preliminary basis, we used the Augmented Dickey-Fuller (ADF) test. Initial results were later examined in light of the results of the Phillips and Perron (1988) test and the stationarity test proposed by Kwiatkowski et al. (1992). All four coincident series showed signs of unit roots in testing and therefore were transformed into first differences (logs) prior to combination into a composite index.

Figure 1. **Coincident series – in log and seasonally adjusted**
Industrial production **Sales**



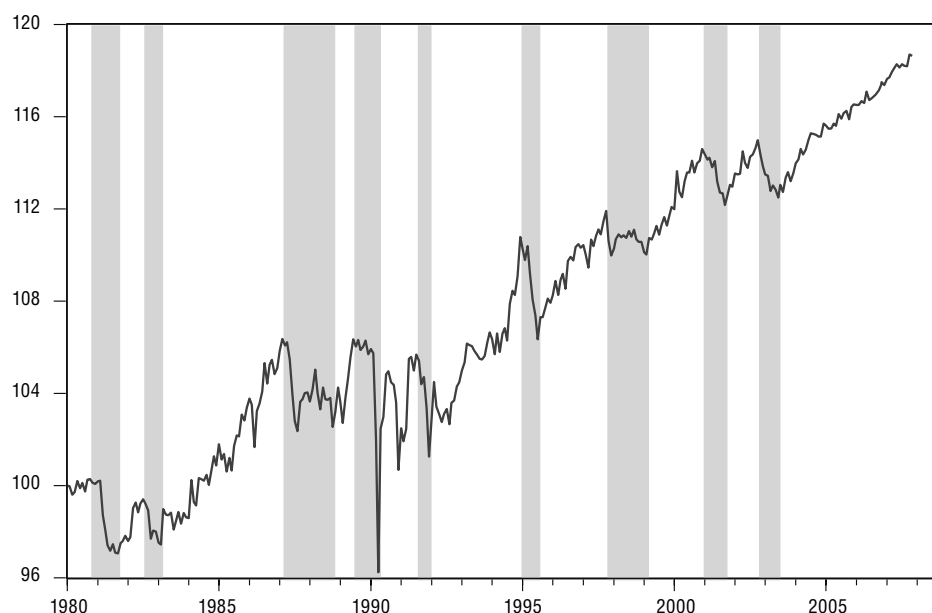
Note: Shaded areas are the actual sample.

4.3. TCB's Coincident Index – $TCB - CI_t$

Using (1), we constructed a coincident index consistent with TCB's method, labelled $TCB - CI_t$, and plotted in Figure 2. Next, we compare the turning-point dating of this index with that of two other indices: a monthly estimate of Brazilian GDP computed by Issler and Notini (2008) and the composite index previously proposed by Duarte, Issler and Spacov (2004), available until 2002:11. The latter also uses the TCB technique.

The turning points of these three composite indices were then compared using the Bry and Boschan (1971), and the Mönch and Uhlig (2005) dating algorithms, the latter being a slightly modified version of the former. Results in Table 3 show that the current dating using the TCB method yields results closer to the dating in Duarte, Issler and Spacov than to the dating of Brazilian monthly GDP. The most striking differences appear in the dating of the 1991 recession. The dating of Duarte, Issler and Spacov and of GDP encompass two recession episodes into one as compared to the dating of $TCB - CI_t$. It is also noteworthy that GDP misses the two last recessions as dated by $TCB - CI_t$ and by Duarte, Issler and Spacov.⁶

Given the results in Table 3, we can compute how frequent Brazilian recessions are. From 1980-2007:11 we have a total of nine recessions. On average, we observe in this period one recession at approximately every 3 years and 3 months, which is substantially more frequent than the US historical average of one recession about every five years. Recessions in Brazil also last longer than US recessions: while ours last about 12 months, on average,

Figure 2. **Coincident index – Shaded Bry-Boschan turning points**Table 3. **Turning-point comparisons using Bry-Boschan dating**

Peak dates			Through dates		
$TCB - CI_t$	Duarte <i>et al.</i>	Brazilian GDP	$TCB - CI_t$	Duarte <i>et al.</i>	Brazilian GDP
1980:10	NA		1981:09	NA	1981:11
1982:07		1982:6	1983:02	1983:10	1983:02
1987:02	1987:04	1988:3	1988:10	1989:02	1988:10
1989:06	1989:08	1989:6	1990:04		
1991:07			1991:12	1991:03	1991:12
1994:12	1995:03	1994:12	1995:07	1995:09	1995:07
1997:10	1997:10	1997:10	1999:02	1999:02	1999:01
2000:12			2001:09		
2002:10	2002:4		2003:06		

Notes: The analysis in Duarte *et al.* (2004) starts in 1982:05, therefore could not have dated the recession of 1980. Brazilian GDP dating uses the monthly series constructed by Issler and Notini (2008).

US recessions last typically from six months to one year, on average (in our sample period here – 1980:1 to 2007:11 – US recessions lasted, on average, nine months). Indeed, Duarte, Issler and Spacov (2004) make the point that this behavior may be due to hardships that the Brazilian economy has endured in the post-1980 era, where GDP growth declined from about 7% a year in real terms prior to 1980 to about 2.2% a year after 1980.

Table 4 lists Brazilian recessions from 1980:1 to 2007:11 when the dating of turning points is made using the modified Bry and Boschan technique proposed by Mönch and Uhlig (2005). The latter takes into account asymmetry differences in peak and through dating, which may be at work to explain the difference in dating between the Bry and Boschan and the Mönch and Uhlig method. Here, the dating of peaks in $TCB - CI_t$ is identical to that in Brazilian GDP, whereas the dating of troughs is almost identical.

Taking into account the overall results of the dating exercise shows that the back-casting of income and employment proposed in this paper has the following properties:

Table 4. **Turning-point comparisons using Mönch and Uhlig dating**

Peak dates			Through dates		
$TCB - Cl_t$	Duarte <i>et al.</i>	Brazilian GDP	$TCB - Cl_t$	Duarte <i>et al.</i>	Brazilian GDP
1980:10	n.a	1980:10	1981:09	n.a.	1981:11
1982:07		1982:07	1983:02		1983:02
1987:02	1987:04	1987:02	1988:10	1989:02	1988:10
1989:06	1989:08	1989:06	1990:04		1990:04
1991:07		1991:07	1991:12	1991:12	1991:12
1994:12	1994:12	1994:12	1995:07	1995:9	1995:07
1997:10	1997:10	1997:10	1999:02	1999:02	1999:01
2000:12	2000:12	2000:12	2001:09	2001:9	2001:09
2002:10		2002:10	2003:06		2003:03

Notes: The analysis in Duarte *et al.* (2004) starts in 1982:05, therefore could not have dated the recession of 1980. Brazilian GDP dating uses the monthly series constructed by Issler and Notini (2008).

i) generates sensible results for those series in the back-cast period; ii) generates a sensible composite coincident index of economic activity. The latter is able to approximate reasonably well the turning points of monthly GDP and those of the TCB index using the retired income and employment series in Duarte, Issler and Spacov (2004). Of course, there are more similarities in turning-point dating when dating uses the technique proposed by Mönch and Uhlig.

Because of these results, we believe that the strategy we chose in this paper to construct a long span time-series for the Brazilian coincident indicator was the best possible. Despite of our belief, we have also considered an alternative exercise: chaining the current employment and income series with their respective series retired by IBGE. The results of the latter, in terms of dating Brazilian cycles, were inferior to that of the back-cast series. For example, in terms of the Quadratic Probability Score (QPS), which measures the in-sample frequency of erroneous dating, when we take the GDP dating as a benchmark, our technique produced $QPS = 0.26$, while the chained series produced $QPS = 0.34$. Similar results were obtained when we consider only recession periods as dated by monthly GDP. What probably explains the superior dating of our chosen technique was that the redesign of the *Monthly Employment Survey* was drastic. Hence, chaining was done using two completely different series.⁷ Indeed, we observe also an increase in “incorrect” dating when chained-series coincident indicators are evaluated *vis-à-vis* the dating in Duarte, Issler and Spacov.

Finally, we link our dating of Brazilian recessions with economic episodes in Brazil and/or abroad, trying also to identify in each recession, the most important factor for its occurrence and the transmission mechanism from factor to the four series in the coincident index. The recession in 1980-81 can be linked to the increase in interest rates by the FED in the early 1980s – which was later responsible for the emerging-market (mostly Latin America) debt crisis. The 1982-83 recession is related to the Latin American debt crisis, where international credit to these economies came to a halt after the Mexican moratorium in 1982. In these two recessions, there was an external factor at work. In 1980-81, this external factor was the increase of foreign interest rates, which reduced domestic demand and economic activity. In 1982-83, due to the Mexican Moratorium, there was a halt to international lending to Brazil as well as capital flight abroad, both working towards reducing economic activity domestically.

The next three recession were “home made”: 1987, 1989, 1991. In them, the Brazilian government could not curb high inflation with many unsuccessful economic plans, which had in common the absence of long-term fiscal discipline. In all of them, a sudden transitory contraction of the money supply was observed, which could not be sustained in the long run due to the lack of fiscal discipline.

After the Real Plan, in July 1994, inflation was finally controlled and most recessions are again related to events abroad, which generated capital flight and thus prompted a sudden rise in domestic interest rates as a reaction to keep foreign capital in domestic markets. The Mexican crisis in 1994 affected Brazil and other emerging markets. In 1997-98, the Asian, the Russian, and the Brazilian crises had similar effects here and in other emerging markets. On these three occasions domestic interest rates risen to very high levels, leading to a reduction in domestic economic activity. In 2000-01 our local energy-supply crisis (electrical energy) was responsible to an economy-wide crisis, while in 2002-03 international markets were nervous with Brazilian high indebttness and a newly elected “left-wing” president.

4.4. The composite leading indicator

Here we propose and evaluate the tracking performance of ten different leading indices *vis-à-vis* the coincident index proposed in the previous section. This is done in two different ways. The first is a typical in-sample analysis and comprises the whole period 1980:1 to 2007:11. Such an exercise is of limited practical importance, since it does not take into account the usual lags in data releases and the fact that dating algorithms, such as Bry-Boschan, requires at least six months ahead of data to detect a recession today. In order to track closely their future use as leading indicators, we also devised an out-of-sample exercise where leading indices were recursively computed using the most recent observations available and were forecast six months ahead in order to date the current state of the economy (recession vs. expansion). The latter is then confronted to the realisations of the future coincident index using the QPS statistic. In this second exercise, we use the period 1980:1 to 1994:12 as a starting point to compute weights of the series in the leading indices. From 1995:1 to 2007:11, these leading indices have their tracking performance evaluated *vis-à-vis* the coincident index.

4.4.1. Data properties and evaluation methods

The selection of a leading indicator index involves three steps: i) select an appropriate indicator as a measure of economic activity to be targeted, also called a reference series; ii) select appropriate economic and financial indicators as predictors of the turning points of the reference series; and, iii) combine the selected leading series in order to construct a composite leading index. We search for series to be in the leading indices that satisfy the following conditions: a) are observable at a monthly frequency for the period 1980-2007:11; b) which have timely data releases and small revisions regarding final data figures.

Recent research has shown that business-tendency survey data are particularly suitable for business cycle monitoring and forecasting. Business tendency surveys are conducted in all OECD member countries and have proved to be a cost-effective means of generating timely information on short-term economic fluctuations. In Brazil, Getulio Vargas Foundation (FGV) is a pioneering institution that computes surveys of economic activity. These include a survey of consumer expectations and another on business

expectations on industrial production and related series: inflow of new orders, level of book orders, stocks of finished goods, etc.⁸

From the FGV survey series and other Brazilian databases (IBGE, IPEADATA, and the Central Bank), we selected 44 series that are candidates of being leading series in leading indices. Our choice was guided by the international experience (Stock and Watson, 1989, 1993) and also by local experience (Duarte, Issler and Spacov, 2004).

All leading nominal series were deflated to reflect their purchasing power as of March, 2008. The deflator used was the Brazilian General Price Index “IGP-DI” – calculated by FGV. All series denominated in foreign currency were converted into Brazilian Reais at the prevailing exchange rate and subsequently deflated. All series were logged, unless logs could not be taken of the original series (potentially zero or negative figures). All series were also seasonally adjusted prior to the analysis using the X-12 procedure, whenever a seasonal pattern in them was detected.

With the exception of the survey-tendency series, all leading series were tested for unit roots. Survey series are bounded series by construction, which is theoretically inconsistent with the presence of unit roots in them. To test for unit roots we used the Augmented Dickey-Fuller (ADF) test, the Phillips and Perron (1988) test, and the stationarity test proposed by Kwiatkowski *et al.* (1992). All series with a unit root were transformed into first differences (logs) prior to forming a composite index.⁹

In order to measure the quality with which a leading series correctly anticipate the “state of the economy” implied by the coincident series (recession or expansion), we use a criterion originally proposed by Diebold and Rudebusch (1990), and later employed by Zhang and Zhuang (2002) and Gallardo and Pedersen (2007). The Quadratic Probability Score, labelled as QPS (h), is given by:

$$QPS(h) = \frac{\sum_{t=1}^T (P_t - R_t)^2}{T} \quad (11)$$

where P_t denotes the predicted state outcomes from a candidate leading indicator and R_t denotes the observed state outcome of the reference series. Both are equal to one for a recession and zero otherwise; T is the total number of sample observations, while h is the horizon in which the leading series potentially predicts the reference series. By construction, the value of QPS (h) ranges between zero and one, with zero indicating a perfect fit for the the “state of the economy” of the reference series h periods in advance.

Next, we describe the basic criteria used to select the leading series that will compose our index. First, for each series, we calculate the optimum (minimum) QPS (h) value, denoted by QPS (h^*), where h^* is the resulting optimum lag. To be a leading series candidate, the series must have $h^* > 0$ in QPS (h^*). This means that the series is leading – not lagging or coincident to – the reference series. Second, we apply Granger (1969) causality tests in order to examine whether the leading series precedes the reference series. We expect that a leading series Granger-causes the reference series but is not Granger-caused by it.

In the Appendix, Table A1. shows the QPS (h^*), h^* , and Granger-causality test results. The majority of the potential leading series do not Granger cause the coincident series. The exceptions are some FGV survey series, in addition to the Central Bank’s basic interest rate (SELIC) and the Brazilian Stock Market Index (IBOVESPA). From them, IBOVESPA

shows promise, since its $QPS(h^*) = 24.5\%$, and $h^* = 5$. This means that, when we take the IBOVESPA index, with a lag of five months vis-à-vis period t , it correctly predicts 75.5% of the “state of the economy” as measured by the peak and through behaviour of our composite index. A slightly worse result is observed to the survey series on the production of real-estate inputs – $QPS(h^*) = 25.1\%$, and $h^* = 1$.

The $QPS(.)$ statistic has only three series with values between 10% and 20% – intermediate-good production, consumer-good production, and inventories, and a few between 20 and 30%. The intersection of the two criteria above – “Granger causality” and “low $QPS(h^*)$ ” – only has the IBOVESPA index and the production of real-estate inputs. Across all potential leading series the mean lag is 3, but the median and modal lag are 1. There are several interesting series which have $h^* > 1$ and a relatively low $QPS(h^*)$: the FGV survey series on inventories ($QPS(h^*) = 17.6\%$), the IBOVESPA index ($QPS(h^*) = 24.5\%$), as well as a myriad of other FGV survey series.

4.4.2. In-sample analysis

Next, we present 10 *ad-hoc* criteria to select series to be in a leading index. The basic idea is that we want h^* to be high, $QPS(h^*)$ to be low, and that a leading series Granger-causes the reference series but is not Granger-caused by it. We also investigate whether a series that has low h^* , with low $QPS(h^*)$, would also have a relatively low $QPS(h)$ for higher values of h . The 10 criteria we chose were:

1. Select all series possessing $QPS(h)$ less than 0.4 and positive optimum lag.
2. Select all series possessing $QPS(h)$ less than 0.4.
3. Select all series that satisfied the Granger causality test criterion.
4. Select all series in the intersection between the first and third criterion.
5. Select all series in the intersection between the second and third criterion.
6. Select all Survey series that satisfied the Granger causality test criterion.
7. Select the five series in Table A1. that have the lowest $QPS(h)$ value.
8. Select the series for which h^* is between two and seven months and $QPS(h^*) < 0.3$.
9. Select the series for which h^* is between two and seven months.
10. Select survey series for which $QPS(h^*) < 0.3$.

Given these criteria, we computed 10 different composite leading indices of economic activity, labelled $LI_{i,t}$, $i = 1, 2, \dots, 10$. We chose to combine leading series into the composite index using a counterpart of equation (1) – equal weights on standardised growth rates of the leading series.¹⁰

Table 5 lists the values of QPS for each criterion listed above, computed for the optimum lag, i.e., $QPS(h^*)$.

From the results in Table 5, $LI_{7,t}$ stands out as a candidate of composite index. All leading series in it have their $QPS(h^*)$ between 11.04% and 23.28%. Despite that, the composite index has a $QPS(h^*) = 10.15\%$ – lower than the smallest $QPS(h^*)$ of the series in it. Although the $QPS(h^*)$ statistic shows that $LI_{7,t}$ appropriately dates the state of the economy about 90% of the time, its optimal h is rather small – just one month, on average. If we compute by how many months the leading series $LI_{7,t}$ leads $TCB - CI_t$ in peaks and troughs we get the following results: by 2.5 months for peaks and by 0.37 months for troughs. Thus it does a much better job in anticipating peaks.

Table 5. **Leading indices: QPS (h^*) computed using Mönch and Uhlig's method**

Series	Description	QPS (h^*)	h^*
$LI_{1,t}$	$h^* > 0$ and $QPS < 0.4$	0.1910	1
$LI_{2,t}$	$QPS < 0.4$	0.1612	1
$LI_{3,t}$	Series that Granger cause the Coincident Index	0.2478	1
$LI_{4,t}$	Granger cause, $h^* > 0$ and $QPS < 0.4$	0.2358	3
$LI_{5,t}$	Granger cause and $QPS < 0.4$	0.2328	1
$LI_{6,t}$	Granger cause (FGV survey series only)	0.2657	1
$LI_{7,t}$	The five series which have the lowest QPS	0.1015	1
$LI_{8,t}$	h^* between 2 and 7 and $QPS \leq 0.3$	0.2507	4
$LI_{9,t}$	h^* between 2 and 7 (FGV survey series only)	0.2866	3
$LI_{10,t}$	$QPS \leq 0.3$ (FGV survey series only)	0.2507	3

There are other leading indices with a higher h^* in Table 5, but they all have a QPS (h^*) statistic that is at least two and a half times those of that of $LI_{7,t}$. If we constrain the QPS (h^*) statistic to be below 20%, there are two other composite indices that also fare relatively well: indices $LI_{2,t}$ and $LI_{1,t}$, which can be considered as an alternative to $LI_{7,t}$, especially if we take into account the fact that $LI_{7,t}$ does not perform well in terms of through prediction. For example, while $LI_{7,t}$ predicts throughs with a very small average lead of 0.37 months, $LI_{1,t}$ leads $TCB - CI_t$ by 2.25 months on average.

4.4.3. Out-of-sample analysis

In the out-of-sample exercise, we used the period 1980:1 to 1995:1 as a starting point to compute weights of the series in the leading indices. For 1995:1, and each candidate leading index, we estimated an AR(1) model to forecast all leading indices six months ahead in order to date whether 1995:1 was either a recession or an expansion period.¹¹ We re-balanced the weights for 1995:2 using data until then, and repeated the forecasting exercise dating whether 1995:2 was a recession or an expansion period. This recursion ended in 2007:11. As a result, for each leading index, we have a sample of states for the economy dated from 1995:1 to 2007:11. These were confronted to the realizations (observed states of the economy) of the coincident index computed in the previous section. The distance between each leading index and the coincident index was computed using the QPS statistic, h^* periods ahead. Table 6 contains the QPS (h^*) results for the 10 candidate leading indices of this exercise.

In Table 6, using the AR(1) model as a benchmark to forecast leading indices out of sample, we find QPS (h^*) statistics which are higher than those in Table 5 (in-sample exercise). This is expected, since in the previous exercise there was no forecasting uncertainty in predicting the state of the economy. Also, the sample is not the same in both exercises: 1980:1-2007:11 versus 1995:1-2007:11. Trying to reproduce the results in the previous section for 1995:1-2007:11, we simply computed non-recursive weights for the latter -last three columns of Table 6. To see the effect of forecasting on final results we computed recursive weights with no forecast.

When using the AR(1) model to forecast the state of the economy, the smallest QPS (h^*) statistic is attained by $LI_{1,t}$, followed by $LI_{6,t}$, with $LI_{2,t}$ and $LI_{8,t}$ tied in 3rd place. The exercise with recursive weights and no forecast produces a reduction of the all QPS (h^*) statistics of roughly 20%, whereas using non-recursive weights reduces it by about 40%. Finally, Table 6 presents confidence bands for all QPS (h^*) statistics, showing that, with three exceptions

Table 6. **QPS (h^*) for leading indices – recursive analysis results**

Series	Rec. weights, AR(1)		Rec. weights, no forecast		Non-recursive weights	
	QPS (h^*)	95% conf. int.	QPS (h^*)	95% conf. int.	QPS (h^*)	95% conf. int.
$LI_{1,t}$	0.3247	(0.203, 0.446)	0.2727	(0.160, 0.385)	0.1883	(0.092, 0.284)
$LI_{2,t}$	0.3506	(0.224, 0.476)	0.2792	(0.164, 0.394)	0.1688	(0.078, 0.259)
$LI_{3,t}$	0.4026	(0.269, 0.535)	0.3117	(0.184, 0.439)	0.2402	(0.122, 0.358)
$LI_{4,t}$	0.3571	(0.232, 0.481)	0.2792	(0.165, 0.393)	0.1233	(0.043, 0.203)
$LI_{5,t}$	0.3896	(0.255, 0.523)	0.3312	(0.199, 0.463)	0.2272	(0.111, 0.342)
$LI_{6,t}$	0.3333	(0.219, 0.446)	0.2876	(0.170, 0.404)	0.1710	(0.072, 0.269)
$LI_{7,t}$	0.3831	(0.268, 0.498)	0.2338	(0.120, 0.347)	0.1753	(0.083, 0.266)
$LI_{8,t}$	0.3506	(0.235, 0.465)	0.2727	(0.157, 0.388)	0.2565	(0.138, 0.374)
$LI_{9,t}$	0.4575	(0.325, 0.589)	0.3922	(0.263, 0.521)	0.2941	(0.173, 0.415)
$LI_{10,t}$	0.3571	(0.230, 0.483)	0.2857	(0.167, 0.404)	0.2565	(0.138, 0.374)

Notes: In the first four columns, weights of the series in each leading index are computed recursively and the AR(1) forecast model is used to predict 6-months ahead for each leading index. This allows dating whether or not there is a recession in period t . The next three columns lists results when weights of the series in each leading index are computed recursively but we do not forecast 6-months ahead for each leading index, rather using their realisations 6-months ahead to date whether or not there is a recession in period t . The final three columns lists results when weights are computed for the whole sample and no forecast is used to detect whether or not there is a recession in period t . Robust SE are computed in constructing 95% confidence bands for QPS($\cdot$), where the procedure proposed by Newey and West (1987) was employed. Values for h^* are given in Table 5.

($LI_{3,t}$, $LI_{5,t}$, and $LI_{9,t}$), 50% is not in the 95% confidence interval. This is an important benchmark in this case, since it is equivalent to tossing up a coin to decide on the state of the economy.

All and all, considering the whole evidence in Section 4.4, we choose $LI_{1,t}$ to be our composite leading index of economic activity. First and foremost, our choice is supported by a QPS value of 32.47% in the out-of-sample exercise, whereas its QPS statistic reduces to 27.27% if we do not have to forecast this leading index six months into the future. From a practical point-of-view, if someone uses this leading index to predict the state of the economy in the near future (h^* periods ahead) they will make an accurate prediction of the state of the Brazilian economy in roughly 70% of the time. Second, with much smaller weight, we also took into account its relatively good results for the in-sample analysis in the previous section.

The performance of $LI_{1,t}$ can be attributed to some of its components: series highly correlated to the international economic activity (e.g., terms of trade, exports quantum and prices, etc.), which we previously identified as key factors determining many recessions in Brazil. Moreover, it also contains some key financial variables such as stock-market index in Brazil (IBOVESPA) and M1, as well as some important credit variables (SPC).

Finally, an alternative to use $LI_{1,t}$ as a leading index could be the use of $LI_{6,t}$ or $LI_{2,t}$ or $LI_{8,t}$. The first includes only series that Granger cause the coincident index but are not caused by it, which perhaps suggests that we should pay more attention to Granger-causality tests in choosing series to be in a composite leading index.

5. Conclusion

This paper has three original contributions. First, by back-casting current employment and income series for Brazil, we allow business-cycle research in Brazil to resume using methods inspired in TCB and NBER, which proved valuable after the comparison made by Duarte, Issler and Spacov (2004). Indeed, the main challenge of Brazilian business-cycle

research was to be able to form a long enough coincident index with the usual series used in the TCB method – industrial production, sales, income and employment. Here, we devoted a great deal of effort in reconstructing employment and income using a novel flexible state-space representation based on the interpolation method of Bernanke, Gertler and Watson (1997) and Mönch and Uhlig (2005).

Once we obtained a long enough span of the usual series used in the TCB method, we compute a new composite coincident index of Brazilian economic activity. Its dating of recessions is compared with those in Duarte, Issler and Spacov and with those implied by the monthly GDP estimate computed by Issler and Notini (2008).

Our last contribution is to propose a composite leading index of economic activity to track our composite coincident index. This is an important topic here, since Brazilian research had focused mainly on the construction of coincident indices. After a wide empirical search, we settled for a composite index ($LI_{1,t}$) that predicts correctly the state of the economy (expansion vs. recession) about 70% of the time. This is achieved in a recursive exercise, where we forecast the next six months of the leading index in order to date whether or not the economy was in recession today.

Notes

1. Regarding Brazilian data, the evidence in Duarte, Issler and Spacov (2004) and in Hollauer, Issler and Notini (2009) concurs with this positive assessment for TCB index.
2. See also Hoover (1988), Biddle (1994), and Simkins (1999).
3. As is well known, the TCB coincident series is computed recursively. Also, it does not use instantaneous growth rates of the coincident series in it. Equation (1) serves merely to illustrate the method in a simplistic way.
4. Note that the aggregation of monthly GDP can also be made averaging the y_t^+ , i.e., as $y_t = \frac{1}{3} \sum_{i=0}^2 y_{t-i}^+$.
5. The seasonal adjustment method adopted was the X-12 Reg-ARIMA 0.3 method (US Census Bureau, 2007). Indeed, changing the seasonal adjustment method changes very little the final results in this paper.
6. This behavior – GDP missing the last two recessions – vanishes if one uses the modified Bry-Boschan dating method proposed in Mönch and Uhlig (2005) to date all three indices.
7. Another alternative would be to only use industrial production and sales to construct the composite index up to 2002:2, and then use the four usual series from 2002:3 onwards. This procedure would probably induce structural changes in mean and variance of the composite index after 2002:3.
8. The FGV survey series are computed on a quarterly frequency up to September 2005. From then on, surveys were then conducted on a monthly basis. Therefore, there is the need to interpolate the data on quarterly frequency to have a homogeneous series on a monthly basis. Our interpolation method was, again, Mönch and Uhlig's (2005).
9. ADF unit-root test results other test results are available upon request.
10. Only series that had a unit root in testing were first differenced. Stationary series were used in levels.
11. We have done the same procedure with a ARMA (1,1) model and the with the best ARMA(p,q) model chosen by the AIC. Results using the AR(1) are superior in almost all cases and differences in results are of no practical importance.

References

- Auerbach, Alan J. (1981), The index of leading indicators: "Measurement without theory", twenty years late, *NBER Working Paper*, No. 761.
- Bernanke B., M. Gertler and M. Watson (1997), "Systematic Monetary Policy and the Effects of Oil Price Shocks", *Brookings Papers on Economic Activity*, 1997(1), pp. 91-157.
- Biddle, J.E., Unpublished interview with Moses Abramovitz, June 1994, "Wesley Clair Mitchell", In: Davis, John B., D. Wade Hands, and Uskali Mäki, eds., *The Handbook of Economic Methodology*, Cheltenham, United Kingdom: Edward Elgar Publishing, 1998, pp. 313-315.
- Bry, G. and C. Boschan (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, New York: National Bureau of Economic Research.
- Burns, A.F. and W.C. Mitchell (1946), *Measuring Business Cycles*, New York: National Bureau of Economic Research.
- Chauvet, M. (1998), "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching", *International Economic Review*, No. 39, pp. 969-996.
- Chauvet, M. (2001), "A Monthly Indicator of Brazilian PIB", *Brazilian Review of Econometrics*, No. 21, pp. 1-48.
- Chauvet, M. (2002), "The Brazilian Business Cycle and Growth Cycles", *Revista Brasileira de Economia*, No. 56, pp. 75-106.
- Chow, G.C. and A. Lin (1971), "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series", *The Review of Economics and Statistics*, No.53(4), pp. 372-375.
- Contador, R.C. (1977), "Ciclos Econômicos e Indicadores de Atividade", Rio de Janeiro, INPES/IPEA, 237 p.
- Contador, C. and C. Ferraz (1999), "Previsão com Indicadores Antecedentes", Rio de Janeiro: Silcon.
- Diebold F.X. and G.D. Rudebusch (1989), "Scoring the leading indicators", *The Journal of Business*, Vol. 62, No. 3, pp. 596-616, July.
- Diebold F.X. and G.D. Rudebusch (1990), "A Nonparametric Investigation of Duration Dependence in the American Business Cycle", *The Journal of Political Economy*, Vol. 98, No. 3 (June, 1990), pp. 596-616.
- Duarte, A.J.M., J.V. Issler and A. Spacov, "Coincident Indices of Economic Activity and a Chronology of Brazilian Recessions", (in Portuguese), *Pesquisa e Planejamento Econômico*, v. 34(1), pp. 1-37, 2004.
- Evans, C.L. 2005, "Comment on "Exploring interactions between real activity and the financial stance" by Jacobson T., J. Linde and K. Roszbach", *Journal of Financial Stability*, Vol. 1(3), pp. 433-434.
- Fernandez, R. (1981), "A Methodological Note on the Estimation of Time Series", *Review of Economics and Statistics*, No. 63, pp. 471-478.
- Gallardo, M. and M. Pedersen (2007), "Indicadores líderes compuestos. Resumen de metodologías de referencia para construir um indicador regional em América Latina", *Serie Estudios estadísticos y prospectivos*, CEPAL, 2007.
- Granger, C.W.J. (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, Vol. 37, pp. 424-438.
- Giannone D., L. Reichlin and D. Small (2008), "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, Elsevier, Vol. 55(4), pp. 665-676.
- Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.
- Harvey, A. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge: Cambridge University Press.
- Hollauer, G., J. Issler and H. Notini (2009), "Novo Indicador Coincidente para a Atividade Industrial Brasileira", *Economia Aplicada*, Vol. 13, No. 1, pp. 5-27.
- Hoover, K.D. (1994), "Econometrics as Observation: The Lucas Critique and the Nature of Econometric Inference", *The Journal of Economic Methodology*, 1:1, June 1994, 65-80.
- Issler, J.V. and H.H. Notini (2008), "Estimating Brazilian Monthly Real PIB: a Kalman Filter Approach", Paper presented in CIRET 2008, mimeo., Graduate School of Economics, Getulio Vargas Foundation.
- Issler, J.V. and A.D. Spacov (2000), "Usando Correlações Canônicas para Identificar Indicadores Antecedentes e Coincidentes da Atividade Econômica no Brasil", mimeo, Relatório de Pesquisa para o Ministério da Fazenda.

- Issler, J.V. and F. Vahid (2006), "The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity", *Journal of Econometrics*, Vol. 132, No. 1, pp. 281-303.
- Koopman, T.C. (1947), "Measurement without Theory", 1947, *Review of Economics and Statistics*, Vol. 29(3), pp. 161-172.
- 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: How Sure Are We That Economic Time Series Have a Unit Root?", *Journal of Econometrics*, Vol. 54, pp. 159-178.
- Kocherlakota, Narayana (2010), "Modern Macroeconomic Models as Tools for Economic Policy", *Banking and Policy Issues Magazine*, Federal Reserve Bank of Minneapolis, Retrieved 2010-07-23.
- Lucas, R.E., Jr. (1977), "Understanding Business Cycles", *Carnegie-Rochester Conference Series on Public Policy*, Vol. 5, pp. 7-29, Amsterdam: North Holland.
- Mariano, R. and Y. Murasawa (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series", *Journal of Applied Econometrics*, Vol. 18, pp. 427-43.
- Mitchell, J., R.J. Smith, M.R. Weale, S. Wright and E.L. Salazar (2005), "An Indicator of Monthly PIB and an Early Estimate of Quarterly PIB Growth", *The Economic Journal*, Vol. 115, F108-F129.
- Mönch, E. and H. Uhlig (2005), "Towards a Monthly Business Cycle Chronology for the Euro Area", *Journal of Business Cycle Measurement and Analysis*, Vol. 2005(1), Paper No. 2.
- Phillips, P. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, Vol. 75, pp. 335-46.
- Picchetti, P. and C. Toledo (2002), "Estimating and Interpreting a Common Stochastic Component for the Brazilian Industrial Production Index", *Revista Brasileira de Economia*, Vol. 56, pp. 107-20.
- Simkins, S. (1999), "Measurement and Theory In Macroeconomics", *Working Paper*, North Carolina A&T State University.
- Spacov, A.D. (2000), "Índices Antecedentes e Coincidentes da Atividade Econômica Brasileira: uma Aplicação da Análise de Correlação Canônica", mimeo, Master's Dissertation, EPGE-FGV.
- Stock, J. and M. Watson (1988a), "A New Approach to Leading Economic Indicators", mimeo, Harvard University, Kennedy School of Government.
- Stock, J. and M. Watson (1988b), "A Probability Model of The Coincident Economic Indicators", *NBER Working Paper*, No. 2772.
- Stock, J. and M. Watson (1989), "New Indexes of Coincident and Leading Economics Indicators", *NBER Macroeconomics Annual*, pp. 351-95.
- Stock, J. and M. Watson (1993a), "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience", in *New Research on Business Cycles, Indicators and Forecasting*, J. Stock and M. Watson, eds., Chicago: University of Chicago Press.
- Stock, J. and M. Watson (1993b), *New Research on Business Cycles, Indicators and Forecasting*, J. Stock and M. Watson, eds., Chicago: University of Chicago Press.
- Zhang, W. and J. Zhuang (2002), *Leading Indicators of Business cycles in Malaysia and the Philippines*, ERD Working Paper, No. 32.

Appendix

Table A1. **Leading series QPS(h^*) and Granger causality**

Leading	Optimum h^*	Min QPS-QPS(h^*)	Granger causes
BASE_R	1	0.4567	B
DEMGLOB	3	0.2806	B
DEMPREV	4	0.2866	B
EXP_R	7	0.3642	N
EXP_QUANTUM	12	0.3104	N
HPP	1	0.4299	B
HTP	1	0.3940	N
IMP_R	1	0.3761	N
IPA_R	11	0.5164	N
M1_R	1	0.3373	C
NUCL_BC	1	0.3463	C
NUCL_BK	1	0.3134	C
NUCL_MC	1	0.2507	C
PO	1	0.4090	N
PRODAUTO	1	0.2149	B
PROD_BC	1	0.1254	B
PROD_BCND	1	0.2328	N
PROD_BI	1	0.1104	N
PROD_BK	1	0.3463	N
PRODINDT	1	0.3104	N
ESTOQUES	2	0.1761	B
IBOV_R	5	0.2448	C
ICMS_R	1	0.3134	B
INPC_R	5	0.4776	N
NUCL_BR	1	0.2746	B
NUCIFESP	1	0.2478	B
PROD_BCD	1	0.2746	N
PROD_CAM	1	0.2627	N
PRODONI	1	0.4179	N
PRODPREV	3	0.2507	N
PRODVEI	1	0.3224	N
SAL_R	1	0.3761	N
NUCL_BI	1	0.3224	B
CAMBIO_R	12	0.6000	B
EXP_PRECOS	3	0.3881	N
IMP_PRECOS	10	0.5045	N
SELIC_R	11	0.5821	C
TTROCA	2	0.3642	N
DEMEXT	6	0.3164	N
DEMINT	2	0.2746	B
DEMPREVEXT	4	0.3463	N
DEMPREVINT	4	0.2716	B

Table A1. **Leading series QPS(h*) and Granger causality (cont.)**

Leading	Optimum h^*	Min QPS-QPS(h^*)	Granger causes
IMP_QUANTUM	1	0.3343	N
SPC	1	0.2537	B

Notes: i) Statistics QPS(h^*) and h^* are computed in accordance with the description of the equation (11) in the text. ii) in the Granger causality test, the symbol C means that the leading series Granger cause at least three out of four series that make up the coincident index with the reciprocal is not true. The symbol B means bi-directional causality in the Granger causality test. The symbol N indicates that the leading series not Granger cause the coincident series. The level of significance was set at 5% in these tests and the number of lags tested was set at 3, 6, 12. To compute the results of the Granger test it was considered the existence of causality in at least one of these lags.

Table A2. **Leading series**

Series name	Description	Source	Series Name	Description	Source
BASE_R	Monetary base	Bacen	NUCI_BR	Survey of manufacturing industry	FGV
SELIC_R	Selic interest rate	Bacen	NUCI_BC	Survey of consumer goods industry	FGV
M1_R	M1 money stock	Bacen	NUCI_BK	Survey of capital goods industry	FGV
IBOV_R	Ibovespa index	Bovespa	NUCI_MC	Survey of construction materials industry	FGV
EXP_PRECOS	Exports prices	Funcex	NUCI_BI	Survey of intermediaries goods industry	FGV
EXP_QUANTUM	Quantum of exports	Funcex	DEMINT	Survey of industry – Level of internal demand	FGV
EXP_R	Exports (FOB)	Funcex	DEMEXT	Survey of industry – Level of external demand	FGV
TTROCA	Terms of trade	Funcex	DEMPREVINT	Survey of industry – Internal demand forecast	FGV
IMP_PRECOS	Imports prices	Funcex	DEMPREVEXT	Survey of industry – External demand forecast	FGV
IMP_QUANTUM	Quantum of imports	Funcex	DEMGLOB	Survey of industry – Level of global demand	FGV
IMP_R	Imports (FOB)	Funcex	DEMPREV	Survey of industry – Global demand forecast	FGV
CAMBIO_R	Exchange rate	Bacen	EMPPREV	Survey of industry – Employment forecast	FGV
NUCIFESP	Manufacturing industry	Fiesp	ESTOQUES	Survey of industry – Level of inventories	FGV
PROD_BC	Production – Consumer goods	IBGE/PIM	PRODPREV	Survey of industry – Production forecast	FGV
PROD_BCD	Production – Consumer durable	IBGE/PIM	FALENCIAS	Bankruptcy – Sao Paulo capital	SP State
PROD_BCND	Production – Consumption and non durable	IBGE/PIM	IPA_R	Ratio: PPI to general price index	FGV
PROD_BI	Production – Intermediate goods	IBGE/PIM	SPC	Credit rating checks	ACSP
PROD_BK	Production – Capital goods	IBGE/PIM	INPC_R	National Consumer Price Index	IBGE
PRODINDT	Industrial production – processing industry	IBGE/PIM	ICMS R	ICMS taxes	Confaz
PRODONI	Production – buses	IBGE/PIM	HTP	Hours worked in production – industry	Fiesp
PRODVEI	Production – vehicles	Anfavea	HPP	Hours paid – industry	Fiesp
PRODAUTO	Production – motors	Anfavea	PO	Staff employed – industry	Fiesp
PRODCAM	Production – trucks	Anfavea	SAL_R	Nominal Salary – industry	Fiesp

Table A3. **Leading series in leading indices 1-10**

- LI_1 : DEMGLOB, DEMPREV, EXP_R, EXP_QUANTUM, IMP_R, M1_R, PRODAUTO, PROD_BC, PROD_BCND, PROD_BI, PROD_BK, PRODINDT, ESTOQUES, IBOV_R, ICMS_R, PROD_BCD, PRODPREV, EXP_PRECOS, TTROCA, DEMEXT, DEMINT, DEMPREVEXT, DEMPREVINT, SPC.
- LI_2 : DEMGLOB, DEMPREV, EXP_R, EXP_QUANTUM, IMP_R, M1_R, NUCI_BC, NUCI_BK, NUCI_MC, PRODAUTO, PROD_BC, PROD_BCND, PROD_BI, PROD_BK, PRODINDT, ESTOQUES, IBOV_R, ICMS_R, NUCI_BR, NUCI_BI, EXP_PRECOS, TTROCA, DEMEXT, DEMINT, DEMPREVEXT, DEMPREVINT, IMP_QUANTUM, SPC.
- LI_3 : BASE_R, M1_R, NUCI_BC, NUCI_BK, NUCI_MC, IBOV_R, NUCI_BI, CAMBIO_R, SELIC_R, DEMEXT.
- LI_4 : M1_R, IBOV_R, DEMEXT.
- LI_5 : M1_R, NUCI_BC, NUCI_BK, NUCI_MC, IBOV_R, NUCI_BI, DEMEXT.
- LI_6 : DEMGLOB, DEMPREV, IBOV_R, PRODPREV, DEMPREVINT.
- LI_7 : PROD_BI, PROD_BC, ESTOQUES, PRODAUTO, PROD_BCND.
- LI_8 : DEMGLOB, DEMPREV, IBOV_R, PRODPREV, DEMEXT, DEMPREVINT.
- LI_9 : DEMGLOB, DEMPREV, PRODPREV, DEMEXT, DEMPREVINT, DEMPREVEXT.
- LI_{10} : DEMGLOB, DEMPREV, ESTOQUES, NUCI_MC, PRODPREV, DEMINT, DEMPREVINT, NUCI_BR.

Heuristic model selection for leading indicators in Russia and Germany

by

Ivan Savin*, ** and Peter Winker***

Business tendency survey indicators are widely recognised 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.

JEL classification: C52, C61, E37

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

* DFG Research Training Program “The Economics of Innovative Change”, Friedrich Schiller University Jena and the Max Planck Institute of Economics.

** Corresponding author: Bachstrasse 18k Room 216, D-07743 Jena, Germany. Tel.: +49-3641-943275, Fax: +49-3641-943202, Ivan.Savin@uni-jena.de.

*** Department of Economics, Justus Liebig University Giessen.

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 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 industrial production (IP) or real GDP, which include BTSs, outperform forecasts of univariate time series models (Hüfner and Schröder, 2002; Kholodilin and Siliverstovs, 2006). Similar evidence can be found for other OECD countries (Ozyldirim *et al.*, 2010; Carstensen *et al.*, 2011). For Russia, in contrast, there is less evidence on the usefulness of BTSs in forecasting IP or other aggregate measures of output (one exception is presented by Nilsson and Brunet, 2006). Possible reasons for that may be a relatively short history of the 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 over optimistic 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 a 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, 2012) allowing for “holes” in multivariate lag structures.

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 analysed BTS-indicators on their forecasting performance during a recession period,¹ the second contribution is to determine whether BTSs in both countries provide useful information before and during the recent crisis. Third, we illustrate the more flexible subset selection strategy based on GA (and the resulting advantages) in comparison to the standard approach for short time series data.²

The rest of this paper is structured as follows. Section 2 presents the database for both countries and the VAR modelling framework used. Results of model specification and forecasting with the standard (*“take all up to the p -th lag”*) approach are reported in Section 3. Section 4 presents the heuristic algorithm for subset model selection and compares its results with the standard procedure. Section 5 concludes with a short summary and an outlook to further research.

2. Data and modelling framework

2.1. Data description

The statistical indicators used in the following are obtained through BTSs conducted in Germany and Russia. The surveys are collected on a monthly basis and are usually available before the end of the respective month.³

The business cycle indicators for Germany are derived from the Ifo business cycle test conducted first in an experimental phase in 1948. Since only the early 1960s, also the diffusion index is calculated (Strigel, 1965). It is published on a regular base by the Ifo Institute (Munich) since the early 1970s (Strigel, 1997). The sample includes about 7 000 enterprises 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; Mitnik and Zadrozny, 2005).

For Russia we consider 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 necessary due to the limited data availability with regard to the BTSs. From IET the data is 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 1 200 enterprises from all major industries with a response rate of about 65% (Tsukhlo, 2005). The coverage of the Rosstat is even larger (over 4 000 companies) from all major industries and all Russian regions.⁶

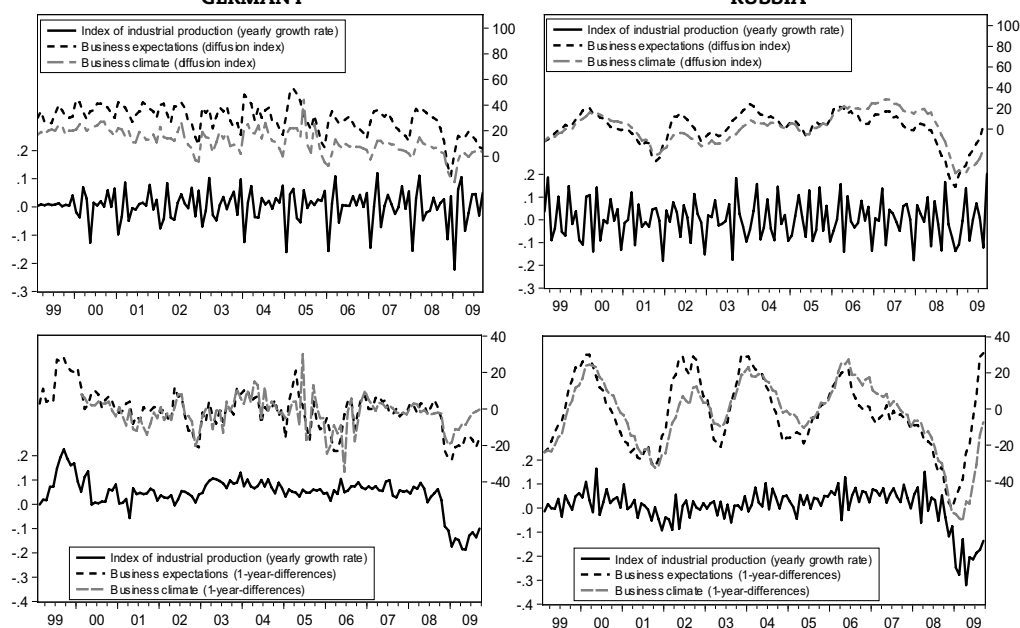
Because of the limited data availability for Russia we consider the period 02/1999-09/2009, leaving one or two years for an out-of-sample forecast analysis between 11/2006 and 09/2009. Thus, we consider the forecasting performance of our models both prior and during the crisis.⁷ By restricting our data 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 performance 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.

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 also refer to the business climate indicator in our analysis. It 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, Goldrian [2007]). The upper plot of 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). It can be observed that BTSs in both countries were predicting the drop in IP for 2008. Considering yearly growth rates (lower plot of Figure 1) one can conclude that for Germany leading indicators started to predict a drop in IP even earlier. Furthermore, it is remarkable that the business expectations indicator for Russia is almost always higher than the business climate index (upper left plot), whereas for Germany there is no systematic margin between the two indicators.

Figure 1. **BTSs and IP 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 analyse this component and its effect on the dynamic adjustment process.⁹

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.¹⁰ This results in a more flexible model accounting for the change in the seasonal structure for Russian BTSs. As regards to IP indices for Germany, we use the data provided by the Federal Statistical Office with base year 2005. The index accounts for just under 80% of total production in the manufacturing sector and covers enterprises also in former East Germany.¹¹ For Russia we consider the data provided by Rosstat with base year 2002 and also a broad coverage of firms.

Using monthly IP indices (instead of quarterly GDP) has certain advantages. First, available at monthly frequency IP requires no aggregation of the BTSs and hence, no loss of information. Second, the number of observations for IP is higher increasing the available 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 monthly 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 $\hat{Y}_{t+h}$ made at time t is the average of the variable Y in the period between 1 and t :

$$\hat{Y}_{t+h} = \frac{1}{t} \sum_{\eta=1}^t Y_{\eta}. \quad (2)$$

This benchmark might not be hard to beat, but allows to compare the relative performance of alternative models. This is due to a similar performance of NF before and during the recession, which is not the case for, e.g., a simple random walk forecast. As alternative benchmarks, we consider AR(p)-processes to identify 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. Therefore, 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 summarise results of unit root tests. Although the results are not perfectly conclusive, they allow the assumption that both variables behave as stationary ones.

The lag length selection for the VAR models is based on the Schwarz (SIC) and the Akaike information criteria (AIC). The latter IC is asymptotically inconsistent (Shibata, 1976), but penalises free parameters less strongly than SIC, which is considered as an advantage for small sample sizes.

2.3. Goodness of forecasts

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

$$\text{RMSFE} = \sqrt{\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} e_t^2}, \quad (3)$$

where e_t is the difference between the actual IP growth rate (y_t) and its forecast ($\hat{y}_t$); and T_1 T_2 indicate the first and the last period in the out-of-sample forecast, respectively. The smaller the RMSFE, the better.

The second accuracy measure is “Theil’s U” (TU) (Theil, 1966):

$$TU = \sqrt{\sum_{t=T_1}^{T_2} e_t^2} / \sqrt{\sum_{t=T_1}^{T_2} u_t^2}. \quad (4)$$

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

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 VAR models. To investigate dynamics of the prediction errors, the so called “rolling windows” method is used (for more information see Swanson, 1998). Doing this we either fix the estimation period, so that its length is the same for different overlapping 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 is 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 model fit ($\hat{\Sigma}$) 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 A.2.

In the following, the resulting VAR models are compared with regard to their out-of-sample performance. In particular, we use a 1-step-ahead forecast that employs only actual observations. First, we consider results with increasing estimation windows (IEW) for an evaluation period of one and two years (Table 1), respectively. 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. These models have an RMSFE smaller than the benchmarks. Similarly, the VAR forecasts for Russia specified by SIC are more accurate than the ones based on AIC. The improvement is of a similar scale as in Germany.¹⁴

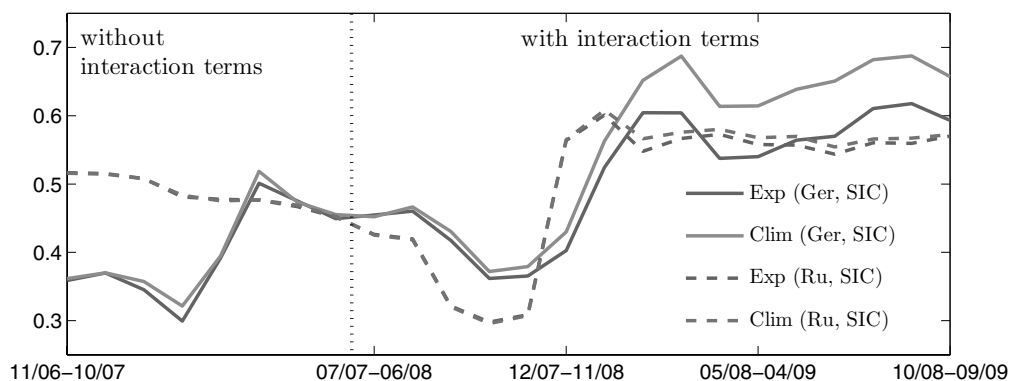
Table 1. **Forecasting performance (RMSFE) of the VAR models**

Model specification		Germany	Russia	Germany	Russia
		10/2007-09/2009		10/2008-09/2009	
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

Next, to analyse goodness of the resulting forecasts for other periods, we consider TU relative to the NF forecast for overlapping windows within 11/2006-09/2009. Results for SIC¹⁵ with IEW are given in Figure 2 for one-year periods (for two-year periods and FEW see

Figures A.1-3). On the x-axis the forecasted period is given, while the y-axis indicates values of Theil's U.¹⁶ While prior to the end of 2008 (when the crisis has become pronounced also in IP indices) the values of Theil's U vary approximately between 0.3-0.5, the relative performance of the VAR models decreases dramatically afterwards.

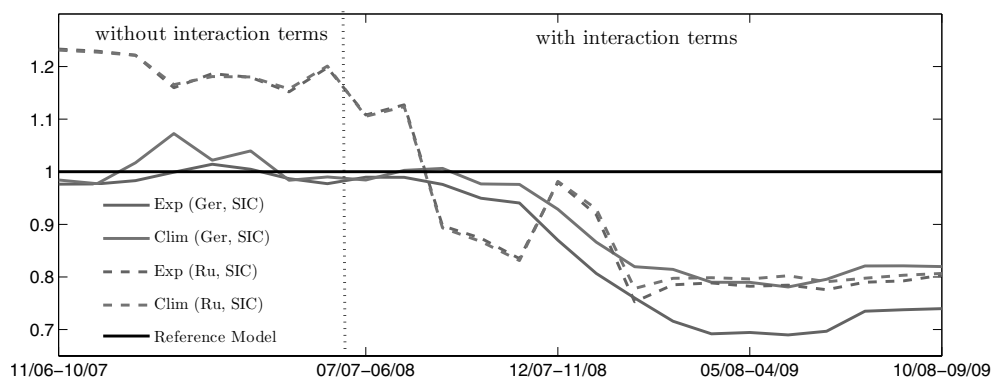
Figure 2. **One-year forecast accuracy compared to NF with IEW (TU)**



Note: We denote in our VAR models Germany with "Ger" and Russia with "Ru", while "Exp" stands for business expectation and "Clim" stands for business climate.

The question, however, is whether the BTS indicators are useful in producing those forecasts (both prior and during the recession). To this end, the VAR models need to be compared with a different benchmark. The AR(p)-processes are most suitable for this comparison, because the contribution of the BTSs as additional indicators is expected to become evident. Nonetheless, as one can see from 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 AIC¹⁷) 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. For example, 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 compared to AR with IEW (TU)**



Note: We denote in our VAR models Germany with "Ger" and Russia with "Ru", while "Exp" stands for business expectation and "Clim" stands for business climate.

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.

4. Heuristic algorithm and resulting models

In the standard bivariate VAR model with 13 lags and seasonal dummies in our study, the rich lag structure requires $k = 76$ parameters (52 lags and 24 dummies) to be estimated. For Russia this number is even higher ($k = 106$) due to the shift dummies and interaction terms. Having only 128 observations (including a period for an out-of-sample forecast), this leaves only a very small number of degrees of freedom. 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 motivating us to use an alternative subset selection strategy allowing for “holes” in multivariate lag structures.

Due to the discrete search space and 2^{m^2p} potential submodels we can neither apply standard optimisation methods nor enumerate all possible solutions for high or even medium values of m and p (as, *e.g.*, 2 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). 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) and later also applied to VAR models (Hsu *et al.*, 2007). Due to the shrinkage parameter the method might exhibit a substantial estimation bias (Hsu *et al.*, 2007; Savin, forthcoming).

To tackle the complex integer optimisation problem and, at the same time, 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 population-based heuristics that operate on a set of solutions (population).¹⁸ 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.

Realising the limitations of IC in our study (in particular, strong overfitting feature¹⁹), in the following we consider RMSFE as an objective criterion for GA. Thus, 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 candidate 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

RMSFE estimate 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 (2012). 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 $G = 100$ is sufficient for one restart (10 restarts for each setting).²⁰

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)²¹ and adjusted R^2 (for the IP growth rate’s equation) (Table 2).

Table 2. **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	0.8397
		SIC	32	0.8661		
	Business climate	AIC	40	0.8694	32	0.8462
		SIC	32	0.8539		
Russia	Business expectations	AIC	62	0.7981	32	0.7751
		SIC	54	0.7868		
	Business climate	AIC	62	0.7947	34	0.7727
		SIC	54	0.7860		

Note: See Section 4.1 for explanation.

As one can see, for Germany GA include an intermediate number of parameters between AIC and SIC, whereas for Russia heuristics come up to a smaller subset than any IC demonstrating in both cases only a slightly lower in-sample fit than the standard lag selection strategy. Thus, omitting certain lags does not produce any substantial loss in the in-sample fit.

4.2. Static forecasts and significance testing

The resulting out-of-sample fit is our main focus due to two reasons. First, a good out-of-sample performance 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.

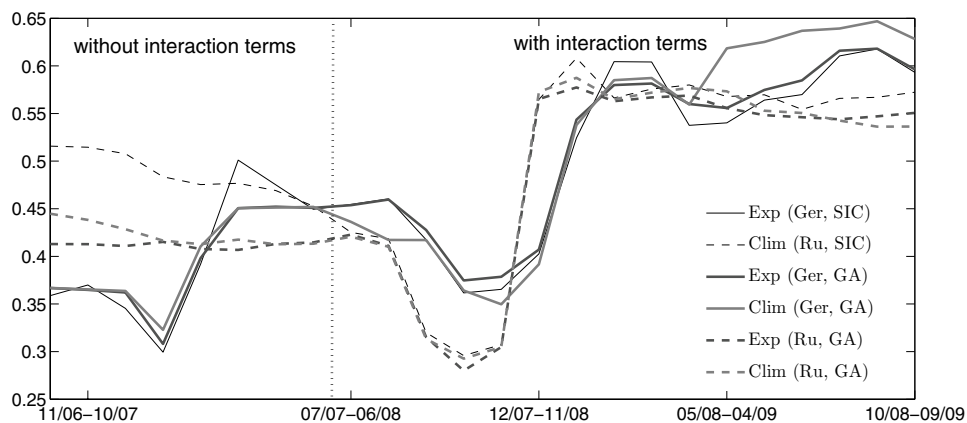
To compare the forecasting performance of our VAR models specified based on the two alternative strategies, consider the resulting TUs (with NF as a benchmark) for one- and two-year out-of-sample periods (Table 3).

Table 3. **Forecasting performance of the VAR models**

Model specification			Standard	GA	Standard	GA
			10/2007-09/2009		10/2008-09/2009	
Germany	Business expectations	AIC	0.5489	0.5075	0.6826	0.5964
		SIC	0.4957		0.5932	
	Business climate	AIC	0.5276	0.5027	0.6459	0.6282
		SIC	0.5395		0.6573	
Russia	Business expectations	AIC	0.6886	0.4834	0.7047	0.5506
		SIC	0.4943		0.5701	
	Business climate	AIC	0.5205	0.4802	0.6096	0.5363
		SIC	0.4956		0.5724	

Note: See Section 4.2 for explanation.

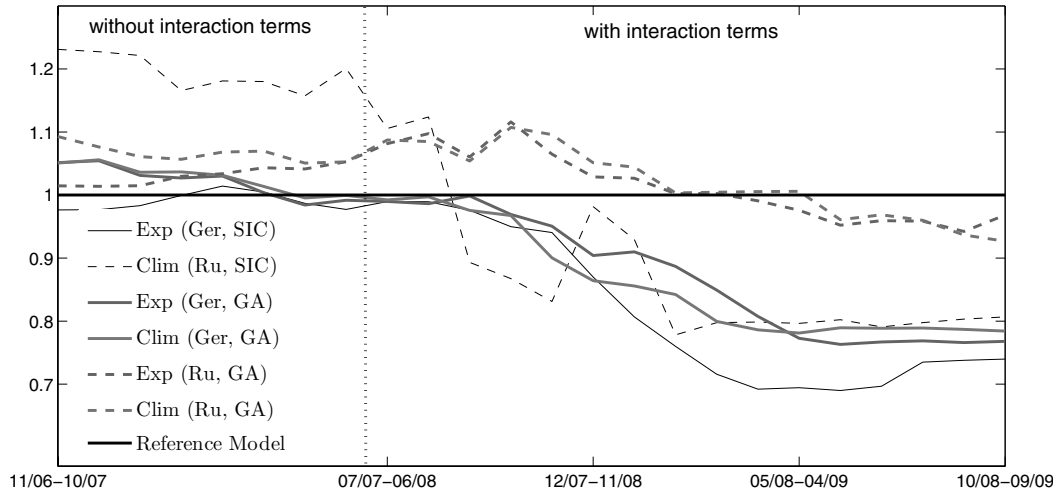
One can see that the GA forecasts based on RMSFE both for one- and two-year periods always outperform the forecasts based on the standard model selection approach via AIC. Moreover, they either outperform the standard strategy based on SIC (*e.g.*, forecasts for Russia and business climate for Germany) or do not substantially differ 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 4).²² 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 4. **One-year forecast accuracy compared to NF with IEW (TU) via GA**

Note: We denote in our VAR models Germany with “Ger” and Russia with “Ru”, while “Exp” stands for business expectation and “Clim” stands for business climate.

Similarly, one can consider the resulting forecasting accuracy of our models in relation to the AR-benchmark (Figure 5).²³ 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

Figure 5. **One-year forecast accuracy compared to AR with IEW (TU) via GA**

Note: We denote in our VAR models Germany with “Ger” and Russia with “Ru”, while “Exp” stands for business expectation and “Clim” stands for business climate.

Giacomini and White (2006) test (GW test henceforth) of conditional predictive ability.²⁴ 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 non-nested 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 generalised approach to existing studies in this field (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 difference of the loss function for the forecast based on the VAR model (e_{t+1}^2) and the benchmark (u_{t+1}^2) for a 1-step-ahead forecast, the null-hypothesis of the test is $E[h_t \Delta L_{m,t+1} | F_t] = 0$, where F_t is an information set available at time t . Then, the test statistic is:

$$GW_{m,n}^1 = n \left(n^{-1} \sum_{t=T_1}^{T_2-1} h_t \Delta L_{m,t+1} \right)' \hat{\Omega}_n^{-1} \left(n^{-1} \sum_{t=T_1}^{T_2-1} h_t \Delta L_{m,t+1} \right) \stackrel{\alpha}{\sim} \chi_q^2, \quad (5)$$

where $n = (T_2 - 1) - (T_1 - 1) = T_2 - T_1$ is the number of out-of-sample forecasts, m 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 $\begin{bmatrix} 1 & \Delta L_{m,t} \end{bmatrix}$.²⁵

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 11/2008.²⁶ 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 2.

While for Germany constructed VAR models demonstrate significantly better forecasting performance only in the latter subsample (during the recession), resulting VAR

Table 4. **Conditional predictive ability test results**¹

Model specification			“All up to p -th lag”		GA	
			Prior	During	Prior	During
Germany	Business expectations	AIC	0.8810	0.8490	1.0010	0.7602*
		SIC	0.9530	0.7415		
	Business climate	AIC	0.8677	0.8034*	1.0210	0.7735**
		SIC	0.9788	0.8204		
Russia	Business expectations	AIC	1.1599	1.0241**	1.0634**	0.9541
		SIC	1.1755*	0.7989		
	Business climate	AIC	1.2335*	0.8791	1.1177*	0.9061
		SIC	1.1714*	0.8021*		

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

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 heuristic approach consistently demonstrates better forecasting performance (Figures 4 and 5), its results can be considered as more justified alternative.

Differences in data publication

As it was mentioned earlier, BTSs are particularly useful because they are published more promptly than IP indices. However, once one is interested in real time forecasts, it is important to control for publication delay of the time series used. In our study, IP indices are published with a delay of 35 and 15 days for Germany and Russia, respectively. BTSs in the two countries, in contrast, are published more synchronously (at the end of the respective month). This implies for the models we build (see equation 1) that the first lag of IP (Y_{t-1}) for Germany is not available for the time the forecast is made (for Russia no additional restriction in data availability is applicable). To account for this important difference we construct VAR and AR models using standard and GA strategies and excluding Y_{t-1} , and repeat the exercise made before. Results are displayed in Figure 6 and Table 5.

Clearly, the supremacy of the VAR models over the AR benchmark becomes less stable when imposing the restriction that the first IP lag should not be included (GA, however, still consistently outperforms the standard model selection strategy). For example, shifting the out-of-sample period by just one month may shift the results in favour of either an AR or a VAR model (Figure 6). As a result, the GW test provides less support for the contribution of the German leading indicators to the forecasting performance (results only significant at the 10% level for business expectations and not significant for business climate). This finding can be explained by a fairly unstable contribution of the first IP lag to forecasting performance in such a volatile period as the recent crisis. Thus, comparing Figures 5 and 6 we see that depending on a particular out-of-sample period exclusion of this information improves the AR benchmark's relative performance to a certain extent, sometimes making it even a dominant strategy.

Figure 6. **One-year forecast accuracy compared to AR with IEW (TU) via GA accounting for the restriction in data availability**

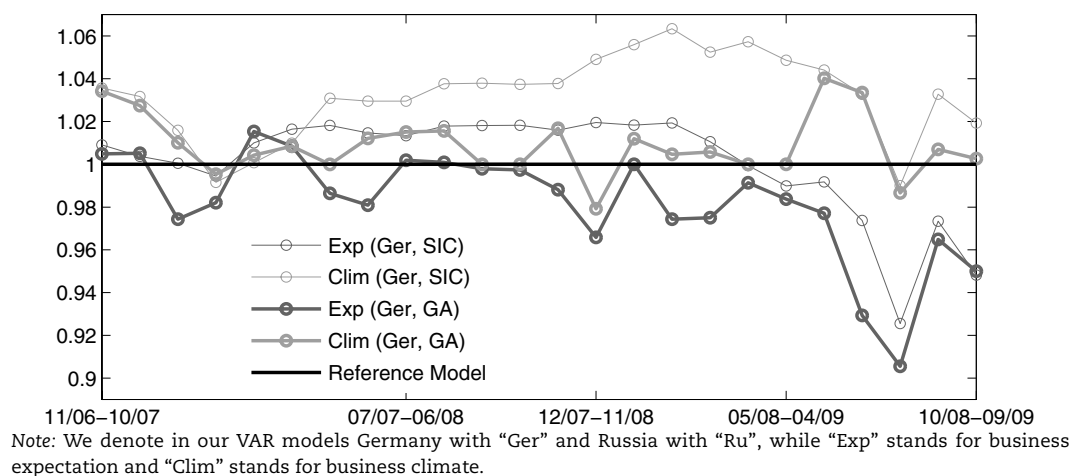


Table 5. **Conditional predictive ability test results accounting for the restriction in data availability¹**

Model specification			"All up to p -th lag"		GA	
			Prior	During	Prior	During
Germany	Business expectations	AIC	0.8803	0.7806*	1.0000	0.8584*
		SIC	0.9368	0.8262		
	Business climate	AIC	0.8441	0.8094	1.0283	1.0164
		SIC	0.9753	0.8653		

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

This latter finding provides an important argument in favour of an earlier publication of real economic data by statistical offices (as Destatis in our case) as this is crucial for the quality of the forecasts to be produced. In particular, while Russian BTS do not significantly contribute to the VAR models' performance, for Germany the evidence is dependent on the data availability issue: in the scenario when the IP value lagged one period is not available for Germany the improvement is marginal, while if all data would have become available at the time of the forecast, resulting forecasts could be improved significantly.

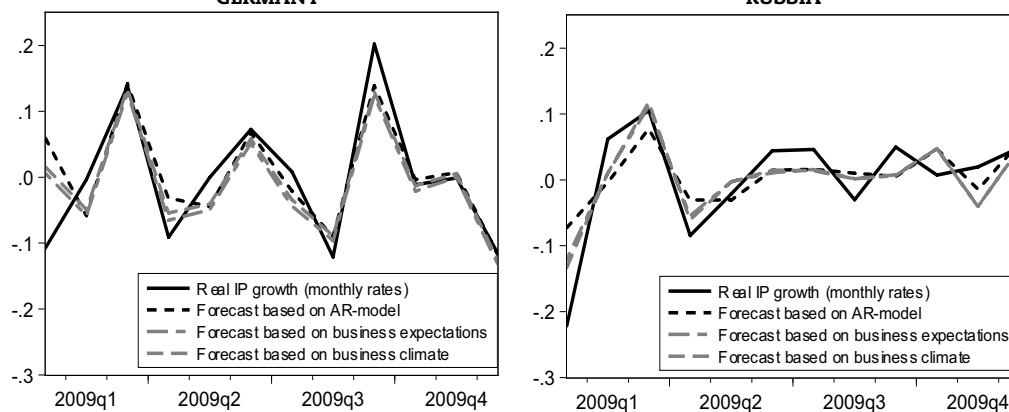
4.3. Dynamic forecast

To provide an additional analysis of the two model selection strategies, we also consider a dynamic h -step-ahead forecast.

Consider first the standard model selection approach. We construct dynamic forecasts for the entire year 2009 (01/2009-12/2009) using the rest of the sample for estimation. The resulting forecasts for all three alternatives (AR models, VAR models with either business expectations or with business climate) via SIC are presented in Figure 7.

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)²⁷ using the best unbiased model specification alternatives according to Table 1 (VAR with business expectations via SIC for both countries), one obtains results as in Table 6.

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

Figure 7. **Prediction dynamics for the IP growth rate for Germany and Russia**

growth rate is much closer to the actual rate for Germany 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.

In contrast, employing the subset models obtained via GA for AR and VAR (with business expectations) models, forecasts constructed for IP in Russia exhibit a slightly improved approximation of the actual growth rates. This holds for both AR and VAR models, while 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 also slightly improved in comparison to the standard strategy.

Table 6. **Prediction accuracy for the IP growth rate in 2009**

		%			
		NF forecasts	AR forecasts	VAR forecasts	Actual rates
Germany	"All up to p -th lag"		-6.13	-17.51	
	GA	-14.45	-6.13	-16.59	-15.99
Russia	"All up to p -th lag"		-4.16	-5.68	
	GA	+0.11	-7.52	-7.42	-10.84

5. Conclusion and outlook

While there are many studies providing evidence that leading indicators improve univariate time series models forecasting of 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 analysing business tendency surveys (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 increasing and fixed estimation windows, we demonstrate that while for Russia there is no significant difference between VAR and AR forecasts during the crisis (prior the crisis the difference is significant, but negative), for Germany the evidence is dependent on the data availability issue: in the scenario taking the restricted data availability into account, the improvement is limited, while if all data would have become available at the time of the forecast, resulting forecasts could be improved significantly. These results are confirmed by the conditional predictive ability test. The latter fact provides an important argument in favour of an earlier publication of real economic data by statistical offices (such as Destatis in our case) as this is crucial for the quality of the forecasts to be produced.

Also, we show that these results can be obtained employing GA, but not the standard “all up to p -th lag” strategy. This seems to be a result of the very limited data sample, where a comprehensive model selection strategy is required to optimise multivariate lag structures sparing some degrees of freedom, 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. As this problem is not extraneous for Germany either (due to, *e.g.* its reunification) and considering the fact that this model selection technique has not been applied in the context for Germany before, the study is seen to be informative for both countries. For future study remains their comparison with other alternatives as, *e.g.*, Bayesian model averaging or shrinkage estimators. Furthermore, in a few years it might be 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. We also thank two anonymous referees for suggestions that led to substantial improvements in the manuscript. All shortcomings are our responsibility.

Financial support from the German Academic Exchange Service (DAAD), the EU Commission through MRTN-CT-2006-034270 COMISEF and the German Science Foundation (DFG RTG 1411) is gratefully acknowledged.

Notes

1. The only exception for Germany is presented by Drechsel and Scheufele (2010).
2. The latter focus is seen to be important for both countries 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 and neither of them used Russian or German BTS data before.
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 a change in classification systems no observations are available.
6. See the Federal State Statistics Service website, www.gks.ru.

7. Note that starting from the fourth quarter of 2008, GDP and IP growth rates for Russia and Germany become negative.
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. Alternatively, we also consider year-on-year growth rates both for IP and BTSs without including seasonal dummies and trying in this way to relax the degrees of freedom problem. In general, however, this does not change any major conclusion (contribution of BTSs before and during the crisis, comparison between standard and GA model selection strategies) stated below. Results are available upon request.
10. Significance of the interaction terms for the corresponding VAR models is confirmed via the F-test at the 1% significance level.
11. See the Federal Statistical Office's web site, www.destatis.de.
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 $d = 12$ and 27 is the number of exogenous variables for Germany and Russia, respectively.
14. We also tested the Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969) to ensure that the produced forecasts are not biased. 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 A.3 and provide some evidence that the respective hypotheses can not be rejected. Similar results are also obtained for other out-of-sample experiments.
15. Results for AIC are available upon request.
16. All figures containing results on rolling windows (Figures 2-5) present TUs in relation either to the NF or AR benchmark. See respective discussion for further explanation.
17. These results are also available upon request, and for two-year periods and FEW, they can be found in Figures A.4-A.6. For the two-year forecasts it is also evident that the interaction terms improve the forecasting accuracy of Russian VAR models.
18. In Savin and Winker (2012) this method is compared with a different heuristic (Threshold Accepting) and is found as more efficient for discrete subset selection problems.
19. See (Savin and Winker, 2010) for an extensive discussion.
20. The corresponding CPU time for 100 generations using Matlab 7.11 on a Pentium IV 3.3 GHz requires approximately 1 600s.
21. This includes both endogenous and exogenous variables.
22. Results of the standard model selection strategy for business expectations via SIC (the best alternative identified in Section 3) are plotted for comparative reasons.
23. In this case both AR and VAR model structures are optimised via GA.
24. So far, available literature for Germany (Kholodilin and Siliverstovs, 2006, Drechsel and Scheufele, 2010) indicates that the difference between AR and VAR models in forecasting performance is significant only to some extent.
25. Setting the conditioning set in this way, we expect past loss differences to predict their future dynamics, i.e. to indicate the more accurate model (Giacomini and White, 2006).
26. Alternatively, we also try different starting periods shifting 11/2008 by one month back and forth. The results remain robust and are available upon request.
27. This methodology is also used by Thompson Reuters.

References

- Alberto, I., et al. (2010), "Variable selection in STAR models with neighbourhood effects using genetic algorithms", *Journal of Forecasting*, Vol. 29(8), pp. 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*, Vol. 22(5), pp. 661-692.

- 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*, Vol. 65(332), pp. 1509-1526.
- Carstensen, K., K. Wohlrabe and C. Ziegler (2011), "Predictive ability of business cycle indicators under test. A case study for the Euro Area industrial production", *Journal of Economics and Statistics*, Vol. 231(1), pp. 82-106.
- Chamberlin, G. (2007), "Forecasting GDP using external data sources", *Economic and Labour Market Review*, Vol. 1(8), pp. 18-23.
- Claveria, O., E. Pons and R. Ramos (2007), "Business and consumer expectations and macroeconomic forecasts", *International Journal of Forecasting*, Vol. 23(1), pp. 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*, Vol. 74(366), pp. 427-431.
- Diebold, F.X. and R.S. Mariano (1995), "Comparing predictive accuracy", *Journal of Business & Economic Statistics*, Vol. 13(3), pp. 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*, Vol. 223(5), pp. 556-570.
- Gatu, C., et al. (2008), "An efficient branch-and-bound strategy for subset vector autoregressive model selection", *Journal of Economic Dynamics and Control*, Vol. 32, pp. 1949-1963.
- Giacomini, R. and H. White (2006), "Tests of conditional predictive ability", *Econometrica*, Vol. 74(6), pp. 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*, Vol. 4, pp. 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 and Data Analysis*, Vol. 52(7), pp. 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*, Vol. 222(3), pp. 316-336.
- Kapetanios, G. (2007), "Variable selection in regression models using nonstandard optimization of information criteria", *Computational Statistics and Data Analysis*, No. 52(1), pp. 4-15.
- 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*, Vol. 226(3), pp. 234-259.
- Kwiatkowski, D., et al. (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root", *Journal of Econometrics*, Vol. 54(1-3), pp. 159-178.
- Langenmantel, E. (1999), "Das ifo-Geschäftsklima als Indikator für die Prognose des Bruttoinl and produktes", *IFO Schnelldienst*, Vol. 16/17, pp. 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*, Vol. 23(3), pp. 667-685.
- Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis. Corr. 2nd printing*, Springer, Berlin, Heidelberg, New York.
- Meyer, M. and P. Winker (2005), "Using HP filtered data for econometric analysis: some evidence from Monte Carlo simulations", *Allgemeines Statistisches Archiv*, Vol. 89(3), pp. 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.
- Mitnik, 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*, Vol. 47(2), pp. 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*, www.oecd.org/dataoecd/29/61/31837055.pdf.
- Ozyldirim, 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*, Vol. 29(1-2), pp. 6-28.
- Savin, I. and P. Winker (2010), "Heuristic model selection for leading indicators in Russia and Germany", *Technical Report 01-2011, MAGKS Joint Discussion Paper Series in Economics*.
- Savin, I. and P. Winker (2012), "Heuristic optimization methods for dynamic panel data model selection, Application on the Russian innovative performance", *Computational Economics*, Vol. 39(4), pp. 337-363.
- Savin, I. (forthcoming), "A comparative study of the Lasso-type and heuristic model selection methods", *Journal of Economics and Statistics*.
- Shibata, R. (1976), "Selection of the order of an autoregressive model by Akaike's information criterion", *Biometrika*, Vol. 63(1), pp. 117-126.
- Strigel, W. (1965), "Die Konjunkturmfragen des ifo-Instituts für Wirtschaftsforschung", *Allgemeines Statistisches Archiv*, Vol. 49, pp. 129-163.
- Strigel, W. (1997), "Seit 25 Jahren: Das ifo Geschäftsklima als Konjunkturindikator", *ifo Schnelldienst*, Vol. 27, pp. 3-4.
- Swanson, N.R. (1998), "Money and output viewed through a rolling window", *Journal of Monetary Economics*, Vol. 41(3), pp. 455-474.
- Theil, H. (1966), *Applied Economic Forecasting*, Rand McNally, Chicago.
- Tsukhlo, S.V. (2005), *The IET surveys, organisational problems and analysis opportunities (in Russian)*. Institute for the Economy in Transition, www.iet.ru/files/persona/tsukhlo/council_24.03.05.pdf.
- West, K.D. (1996), "Asymptotic inference about predictive ability", *Econometrica*, Vol. 64(5), pp. 1067-1084.
- Winker, P. (1995), "Identification of multivariate AR-models by threshold accepting", *Computational Statistics & Data Analysis*, Vol. 20, pp. 295-307.
- Winker, P. (2000), "Optimized multivariate lag structure selection", *Computational Economics*, Vol. 16, pp. 87-103.

Annex A

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).¹ 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 summarised in columns 3 and 6 in Table A.1.

Table A.1. **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 the IP growth rates² 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 columns 4 and 7 in Table A.1). As a result, we cannot reject the null-hypothesis for business expectations in Russia, but due to the low power of the ADF-test we presume the variable to be stationary, *e.g.*, owing to its constraint on the interval $[-100; 100]$.

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 A.2) with its *p*-values in parentheses.

Table A.2. **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)

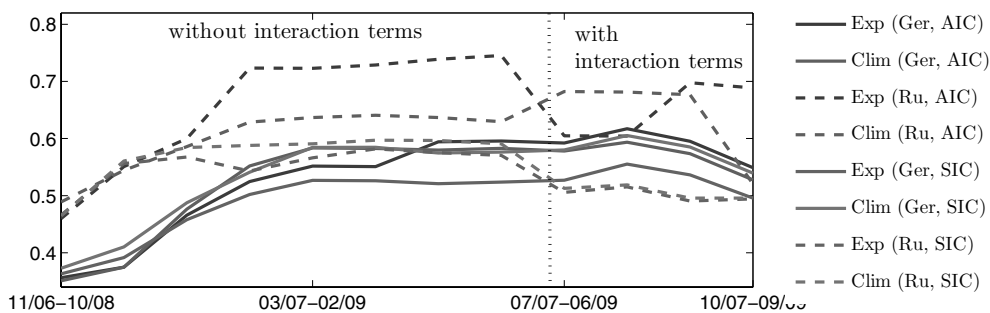
It becomes obvious that the models for the German data do not exhibit residual autocorrelation at the 5% significance level when choosing the lag length suggested by SIC. This also holds true for the Russian VAR model with the business expectations, but not for the business climate. This result might 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 A.2. 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.

Forecasts' bias analysis

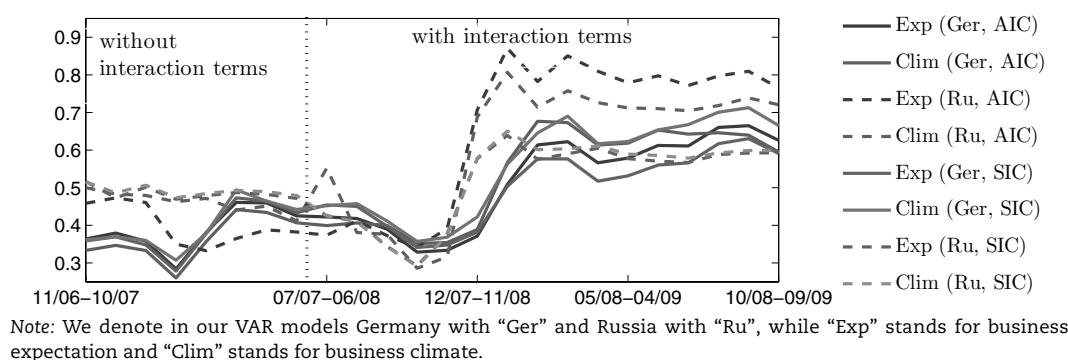
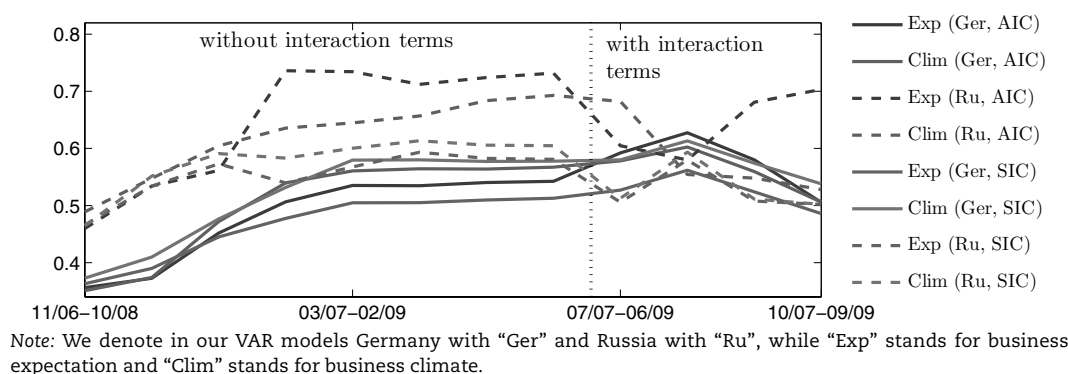
Table A.3. **Forecasts' bias analysis**

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

Forecast accuracy in relation to NF

Figure A.1. **Two-year forecast accuracy in relation to NF with IEW**

Note: We denote in our VAR models Germany with "Ger" and Russia with "Ru", while "Exp" stands for business expectation and "Clim" stands for business climate.

Figure A.2. **One-year forecast accuracy in relation to NF with FEW**Figure A.3. **Two-year forecast accuracy in relation to NF with FEW**

Forecast accuracy in relation to AR

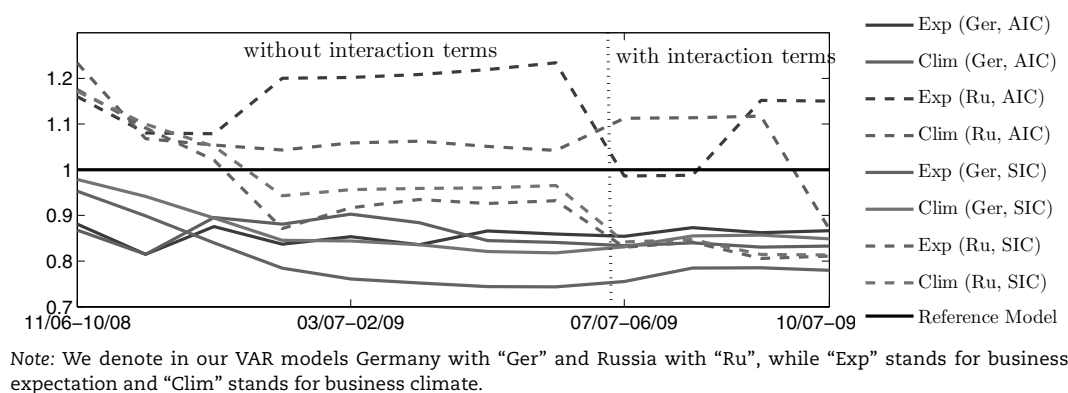
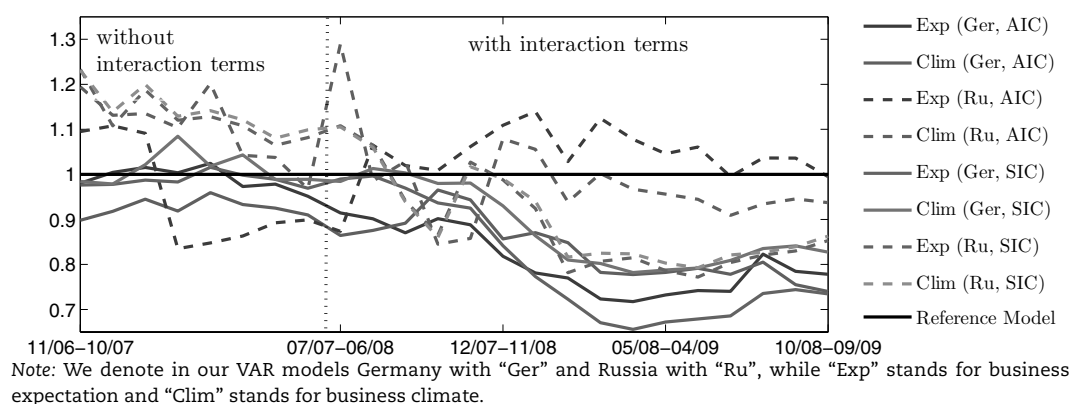
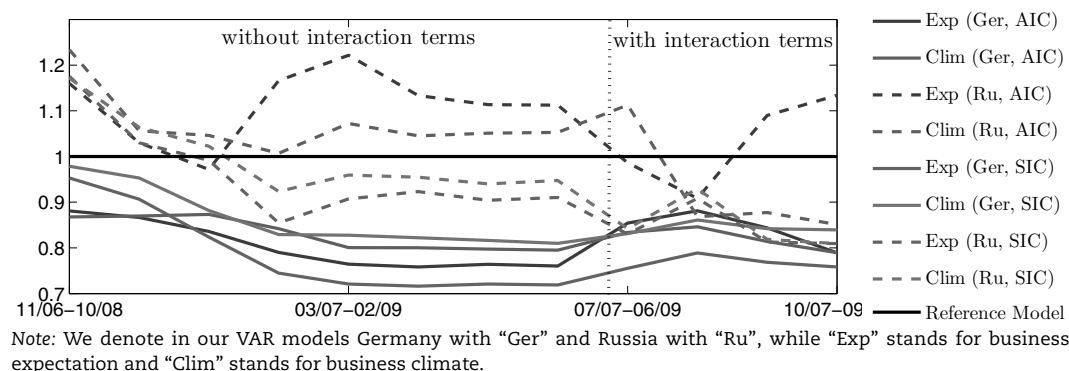
Figure A.4. **Two-year forecast accuracy in relation to AR with IEW**

Figure A.5. **One-year forecast accuracy in relation to AR with FEW**Figure A.6. **Two-year forecast accuracy in relation to AR with FEW**

Notes

1. The latter test is implemented using the JMulti free statistical software package available at www.jmulti.de.
2. 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 upon request).

Referee acknowledgement

The Editorial Board of the *Journal of Business Cycle Measurement and Analysis* gratefully acknowledges the following individuals for reviewing papers submitted to this journal in 2011 and 2012. Without their devoted and honorary effort, this publication would not be feasible.

Aguiar-Conraria, Luis Francisco	Heij, Christiaan	Piger, Jeremy
Anas, Jacques	Hervai, Saeed	Podgórska, Maria
Angeloni, Ignazio	Hulagu, Timur	Poser, Günter
Arnaud, Franck	Issler, João Victor	Próchniak, Mariusz
Artis, Michael	Jacobs, Jan	Proietti, Tommaso
Assenmacher, Kathrin	Kappler, Marcus	Radovic-Stojanovic
Bañbura, Marta	Kaya, Hüseyin	Raponi, Valentina
Bandholz, Harm	Keating, John	Rathke, Alexander
Benczúr, Péter	Keck, Alexander	Reijer, Ard den
Berge, Travis John	Knetsch, Thomas A.	Reijnders, Jan P.G.
Billio, Monica	Knoll, Christian	Richter, Christian R.
Branch, William A.	Kowalczyk, Barbara	Ríos Rull, José Víctor
Bruzda, Joanna	Lamla, Michael	Rodríguez, Gabriel
Buss, Ginters	Lenart, Lucas	Sattarova, Liliya
Caporale, Guglielmo Maria	Lenza, Michele	Saygil, Hülya
Castro, Vítor	Leproux, Solange	Scheiblecker, Marcus
Çelik, Sadullah	Lewis, John	Scheufele, Rolf
Chow, Hwee Kwan	Love, David R.F.	Schmitt, Elizabeth Dunne
Conti, Antonio Maria	Lyziak, Tomasz	Schnatz, Bernd
Cosar, Evren Erdogan	Malgarini, Marco	Scotti, Chiara
Croushore, Dean	Marczak, Martyna	Sella, Lisa
Cubadda, Gianluca	Margani, Patrizia	Seymen, Atilim
Cuche-Curti, Nicolas	Massmann, Michael	Shields, Kalvinder
Cunningham, Rose	Maurin, Laurent	Sigalla, Fiona
De Carvalho, Miguel	Mayr, Johannes	Siliverstovs, Boriss
D'Elia, Enrico	Mazzi, Gian Luigi	Silvestrini Andrea
Di Fonzo, Tommaso	McElroy, Tucker Sprague	Simonelli, Saverio
Dijk, Dick van	Meyler, Aidan	Smirnov, Sergey
Dinu, Marin	Mitchell, James	Starica, Catalin
Duarte, Augustin	Moauero, Filippo	Stark, Tom
El Dakkak, O.	Modugno, Michele	Strawczynski, Michel
Fichtner, Ferdinand	Mönch, Emanuel	Sturm, Jan-Egbert
Frale, Cecilia	Negri, Valeria	Tallman, Ellis
Fritsche, Ulrich	Nieuwenhuyze,	Terai, Akira
Friz, Roberta	Christophe van	Terasvirta, Timo
Giannone, Domenico	Norden, Simon van	Veronese, Giovanni Furio
Giedeman, Daniel	Notini, Hilton Hostalacio	Vuletin, Guillermo
Groenen, Patrick J.F.	Nunes, Luis C.	Wegener, Michael
Gyomai, György	Nymoen, Ragnar	Wynne, Mark A.
Hall, Pamela	Paradiso, Antonio	Zanetti, Attilio
Heer, Burkhard	Paries, Matthieu Darracq	
	Pelagatti, Matteo M.	

Aims and scope

The *Journal of Business Cycle Measurement and Analysis* is jointly published by the OECD and CIRET to promote the exchange of knowledge and information on theoretical and operational aspects of economic cycle research, involving both measurement and analysis.

The OECD has developed various cyclical indicators using quantitative and qualitative information from its member countries. Information about the work of OECD in the field of short-term economic statistics may be found at www.oecd.org/std.

CIRET, the Centre for International Research on Economic Tendency Surveys, is a global forum for economists and institutions, which conduct and analyse business and consumer surveys. CIRET holds international conferences every other year. Information about CIRET may be found at www.ciret.org.

The scope of the *Journal of Business Cycle Measurement and Analysis* encompasses:

- Analysis of cyclical fluctuations.
- Business cycle specification, definition and classification.
- Statistical approaches to the development of short-term economic statistics and indicators.
- Business tendency, investment and consumer surveys.
- Use of survey data or cyclical indicators for business cycle analysis.

Researchers in the field of business cycle measurement and analysis are encouraged to submit their papers to the Editor-in-Chief, Michael Graff (editor-in-chief@ciret.org). All papers undergo a double-blind review process.

Under the heading “Report” additional contributions, including topics such as sample design, data collection and validation, dissemination and evaluation of survey results, are published. Reports are reviewed by members of the Editorial Board.

Other correspondence related to the journal may also be addressed to Pierre-Alain PIONNIER, at the OECD (Pierre-alain.PIONNIER@oecd.org).

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD Journal

Journal of Business Cycle Measurement and Analysis

The *Journal of Business Cycle Measurement and Analysis* is jointly published by the OECD and the Centre for International Research on Economic Tendency Surveys (CIRET) to promote the exchange of knowledge and information on theoretical and operational aspects of economic cycle research, involving both measurement and analysis.

Measuring Capacity Utilisation in the Italian Manufacturing Sector: A Comparison between Time Series and Survey Estimates

Marco Malgarini and Antonio Paradiso

Nowcasting Irish GDP

Antonello D'Agostino, Kieran McQuinn and Derry O'Brien

Day of the Week Effect in Consumer Confidence Index: The Case of Turkey

Sadullah Çelik and Hüseyin Kaya

Constructing Coincident and Leading Indices of Economic Activity for the Brazilian Economy

João Victor Issler, Hilton Hostalacio Notini and Claudia Fontoura Rodrigues

Heuristic model selection for leading indicators in Russia and Germany

Ivan Savin and Peter Winker

Referee Acknowledgement

www.oecd.org

www.ciret.org

www.kof.ethz.ch

This issue is published on the OECD iLibrary at: <http://dx.doi.org/10.1787/19952899>

The OECD iLibrary gathers all OECD books, periodicals and statistical databases. Visit www.oecd-ilibrary.org, and do not hesitate to contact us for more information.

Volume 2012/2

OECD publishing
www.oecd.org/publishing

ISSN 1995-2880

VOLUME 2012

(2 ISSUES)

33 2012 02 1 P 9



771995 288001