



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 2014/2

A Joint Publication of OECD and CIRET

The *OECD Journal* collection comprises four periodicals: *OECD Journal: Economic Studies*, *OECD Journal: Financial Market Trends*, *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. Volume 2014 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-v2014-2-en>

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

© OECD 2015

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

Tracking economic activity in the euro area: Multivariate direct filter approach by Ginters Buss	5
What's behind the survey values? An analysis of individual forecasters' behaviour by Maritta Paloviita and Matti Viren	25
On the robustness of balance statistics with respect to nonresponse by Christian Seiler.	45
Cohesion within the euro area and the US: A wavelet-based view by António Rua and Artur Silva Lopes	63

Editors

Editor-in-Chief

Michael Graff, *ETH Zurich, Switzerland*

Associate Editor

Pierre-Alain Pionnier, *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*

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*

Ulrich 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*

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

Giuseppe Parigi, *Bank of Italy, Rome, Italy*

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 Vértés, *GKI Economic Research Co., Budapest, Hungary*

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

Tracking economic activity in the euro area: Multivariate direct filter approach

by
Ginters Buss*

The paper applies the multivariate direct filter approach on selected business and consumer confidence indicators, and share price data to construct a real-time indicator tracking the medium-to-long-run component of the quarterly growth of the euro area gross domestic product. Results show that the created indicator behaves similarly to another established indicator, Eurocoin, but slightly leads it after mid 2009. The new indicator is also compared to the Markit Euro Area Composite Purchasing Managers Index and is found to be leading it by about one month as well as being smoother. Overall, the multivariate direct filter approach is found to have merit in tracking business cycle developments, however, the increasing number of free coefficients is an issue for the filter to be applied to detailed datasets.

Keywords: real-time signal extraction, business cycles, multivariate time series

JEL classification: C13, C32, E32

* The author thanks Marc Wildi and anonymous referees for valuable feedback on an earlier draft. All remaining errors are the author's own. This report is released to inform interested parties of research and to encourage discussion. The views expressed in this paper are those of the author and do not necessarily reflect the views of the Bank of Latvia. Address for correspondence: Latvijas Banka, K. Valdemara 2A, Riga, LV-1050, Latvia; ginters.buss@gmail.com.

1. Introduction

The paper applies the recently developed multivariate direct filter approach (MDFA) (Wildi, 2011) on selected business and consumer confidence indicators, and share price data to construct a real-time indicator tracking the medium-to-long-run component of the quarterly growth of the euro area gross domestic product (EA GDP).

The demand for real-time macroeconomic indicators exists mainly because many key macroeconomic variables are released with a considerable lag and are subsequently revised in later releases. For example, the first rough estimate of the main macroeconomic aggregate, GDP, called flash GDP, is released 45 days after the reference period in the European Union (EU), including euro area (EA), and is then revised substantially. The first official release of GDP in EU is published 65 days after the reference period, and this first release too, is revised in subsequent releases. However, economists are keen to have timely information on new developments in the economy. Therefore, efforts are made in using appropriate information to capture the big picture of the overall economy in a more timely manner. For that matter, several indicators are available that try to map timely disaggregate information on the aggregate series, like the GDP. In particular, there are several established real-time indicators tracking economic situation in the EA, two of which are discussed below.

The (New) Eurocoin (Altissimo et al., 2010) is a replacement of its predecessor, (Old) Eurocoin (Altissimo et al., 2003), since May 2009. This indicator tracks the medium-to-long-run component in the quarterly growth of the EA GDP and is the result of a dynamic factor methodology (Forni et al., 2000, 2005) applied to about 145 series.

The monthly Markit flash EA Purchasing Managers Index¹ (PMI) is one of the closely watched indices for the EA economy because it is timely, unrevised and has straightforward methodology. It is based on surveys of companies.

This is the first paper known to the author that implements MDFA to track a trend-cycle and to implement this approach to the euro area dataset. Another paper that implements this approach is Wildi and Sturm (2008) however they use a bandpass specification and apply it to the US dataset. Wildi (2011) develops the filtration methodology implemented in this paper but does not apply it to the data. Recently, Gyomai and Wildi (2012) applied a refined version of the filter (Wildi, 2012) to track the bandpass of the GDP of several countries but not the euro area.

This paper applies MDFA on selected survey and stock price data to create a real-time indicator for the EA GDP, and compares its performance to the Eurocoin and PMI. The filter output for the EA dataset is found to be about coincident with Eurocoin and to lead the PMI by about one month.

Overall, MDFA is found to have merit in tracking business cycle developments using a few explanatory variables, however, the increasing number of the free coefficients of the filter is an issue for the filter to be applied on detailed datasets. See Wildi (2012) for the

recent refinement of the methodology attempting to deal with this issue, and Buss (2012) for its application to many variables.

This paper is structured as follows: Section 2 reviews the methodology of the direct filter approach; Section 3 describes the design of the indicator; Section 4 produces the indicator and compares with the alternatives; Section 5 concludes; and, Annex A discusses robustness checks.

2. Background on the direct filter approach

The direct filter approach is concerned with estimating a signal, such as a trend cycle or a business cycle, in real time. Consider weakly stationary zero-mean input series $\{x_t\}$, $t = 1, \dots, T$, with an absolutely continuous spectral distribution function $H(\omega) = \int_{-\pi}^{\pi} h(\omega') d\omega'$, where $\omega \in [-\pi, \pi]$ denotes the frequency, and $h(\omega)$ is the spectral density of the series, $h(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c(\tau) \exp(-ij\omega)$, where $c(\tau)$ is the autocovariance function, $c(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{T-|\tau|} x_t x_{t+|\tau|}$, $|\tau| = 0, \dots, T-1$, $|\cdot|$ denotes the absolute value, and i is the imaginary number defined as $i^2 = -1$. Denote $\{y_t\}$ as the ideal output of a symmetric, possibly bi-infinite filter, $\Gamma(B) = \sum_{j=-\infty}^{\infty} \gamma_j B^j$, applied on an input series $\{x_t\}$:

$$y_t = \sum_{j=-\infty}^{\infty} \gamma_j x_{t-j} \quad [1]$$

where B is the lag or backshift operator. The filter in [1] is called the ideal filter and the filter output, $\{y_t\}$, is called the ideal filter output. Time series samples have finite length in practice and, therefore, the ideal filter, in general, is infeasible to apply. A practitioner might use a finite length symmetric filter as an approximation to [1] in the middle of the time series but even the symmetric approximation is infeasible to apply at the very end of the sample, that is, in real time. A real-time estimate of $\{y_t\}$, given a finite input $\{x_1, \dots, x_T\}$, can be written as:

$$\hat{y}_T = \sum_{j=0}^{T-1} b_j x_{T-j} \quad [2]$$

It is a well-known fact that the filter in [2] generally possesses a non trivial time shift (defined as the phase, see definition below, divided by the frequency), that is, its output is lagging in time. In order to define the phase and, in general, the effect of a filter, denote the generally complex transfer functions of filters in [1] and [2] by $\Gamma(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j \exp(-ij\omega)$ and $\hat{\Gamma}(\omega) = \sum_{j=0}^{T-1} b_j \exp(-ij\omega)$, respectively. The number $\Gamma(\omega)$ can be represented in polar co-ordinates as $\Gamma(\omega) = A(\omega) \exp[-i\phi(\omega)]$, where $A(\omega) = |\Gamma(\omega)|$ is the amplitude, and $\phi(\omega) = -\arg[\Gamma(\omega)]$ is the phase. The effect of a filter – as represented by its transfer function – can be summarised by the effects of the amplitude and phase functions. The amplitude function, $A(\omega)$, can be interpreted as the weight – amplification or damping – attributed by the filter to the input series at frequency ω . The phase function, $\phi(\omega)$, can be interpreted as a shift function of the amplified or damped signal at frequency ω . This paper is concerned with obtaining a one-sided filter output which would be as good as possible in real time, approximation to the ideal output. For a weakly stationary input series $\{x_t\}$, the mean squared filter error (MSFE) can be expressed as the mean squared difference between the ideal output and the real-time estimate:

$$\int_{-\pi}^{\pi} |\hat{\Gamma}(\omega) - \Gamma(\omega)|^2 dH(\omega) = E[(y_T - \hat{y}_T)^2] \quad [3]$$

A finite sample approximation of MSFE, [3], is:

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \xi_k |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 S(\omega_k) \quad [4]$$

where $\omega_k = k2\pi/T$, $[T/2]$ is the greatest integer smaller or equal to $T/2$, and the term ξ_k is defined as:

$$\xi_k = \begin{cases} 1 & \text{for } |k| \neq T/2 \\ 1/2 & \text{otherwise} \end{cases} \quad [5]$$

Brockwell and Davis (1987, Ch. 10) explain the reason to use ξ_k ; as without it the inverse discrete Fourier transform does not replicate the sample autocovariance perfectly. $S(\omega_k)$ in [4] is defined to be an estimate of the unknown spectral density of $\{x_t\}$, which can be any spectral estimate. As discussed in Wildi (2008), consistency of $S(\omega_k)$ is not required because the goal is not to estimate $dH(\omega)$ but the filter mean squared error, [3]. Therefore, this paper uses the periodogram, $I_{Tx}(\omega_k)$ – which is sufficient for our purposes, and bears an analogy to a sufficient statistic, defined as:

$$S(\omega_k) := I_{Tx}(\omega_k) = \frac{1}{2\pi T} \left| \sum_{t=1}^T x_t \exp(-it\omega_k) \right|^2 \quad [6]$$

Minimising expression [4] yields the real-time filter output optimally approximated to the ideal output in regards to the mean squared error. The expression [4] is a generalised problem that encompasses the problems of: Baxter and King (1999) where the feasible filter is assumed to be a symmetric bandpass and the spectral estimate, $S(\omega_k)$ be the spectrum of the white noise; Christiano and Fitzgerald (2003), where the default specification of the feasible filter is assumed to be a bandpass and the spectral estimate, $S(\omega_k)$ be the spectrum of the random walk; or, the Hodrick-Prescott filter, which can be interpreted to be a Butterworth-type filter, assuming the input following an ARIMA(0,2,2) process, and its amplitude function being of a particular truncated bell-shaped form, (see King and Rebelo, 1993; Maravall and Rio, 2001). Yet, Wildi (2008) proposes a generalised version of [4] which this paper finds to be useful in creating a real-time indicator, and which is described in the Section 2.1.

2.1. Univariate direct filter approach

Wildi (2008) proposes a customised version of [4], a part of which is implemented in this paper. The discrete version of MSFE, [4], is rewritten as:

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \xi_k |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 I_{Tx}(\omega_k) W(\omega_k) \quad [7]$$

which is identical to [4] for $W(\omega_k) = 1$. However, a more general version of $W(\omega_k) = W(\omega_k, expw, cut)$ can be written as:

$$W(\omega_k, expw, cut) = \begin{cases} 1 & \text{if } |\omega_k| < cut \\ (1 + |\omega_k| - cut)^{expw} & \text{otherwise} \end{cases} \quad [8]$$

which collapses to unity for $expw = 0$, in which case the classical mean squared optimisation, [4], is obtained. Parameter cut (for cut-off frequency) marks the transition between passband and rightmost stopband, and positive values of $expw$ (for exponent weight) emphasise high-frequency components in the rightmost stopband, thus, making the filter output smoother than the one obtained by minimising [4] for positive $expw$. By choosing $expw > 0$, the additional weighting function $W(\omega_k)$ grows strictly monotonically in the stopband. Thus, it assigns more weight to the fit of $\Gamma(\omega_k)$ by $\hat{\Gamma}_x(\omega_k)$ towards higher frequencies. Since the ideal trend filter vanishes in the stopband, the additional weighting scheme forces $\hat{\Gamma}_x(\omega_k)$ to approach $\Gamma(\omega_k)$, which vanishes in the stopband. As a result, high-

frequency components are damped more vigorously by the resulting real-time filter: the output, in general, will be smoother.

Univariate filters can be useful in case the target variable of interest, x_t , is timely; however, it is usually not the case for macroeconomic series. For example, one of the key macroeconomic variables, GDP, is released with a delay and is revised in subsequent releases. Therefore, a practitioner might be interested in using other economic variables with shorter delays in creating benchmark indicators for GDP. Therefore, the Section 2.2 examines the multivariate filter.

2.2. Multivariate direct filter approach

The above univariate customised filter has been generalised to a multivariate filter in Wildi (2011). Rewrite the univariate minimisation problem, [7], with the discrete Fourier transform (DFT), $\Xi_{Tx}(\omega_k)$:

$$\begin{aligned} & \frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \xi_k |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 I_{Tx}(\omega_k) W(\omega_k) \\ & = \frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \xi_k |\Gamma(\omega_k) \Xi_{Tx}(\omega_k) - \hat{\Gamma}(\omega_k) \Xi_{Tx}(\omega_k)|^2 W(\omega_k) \end{aligned} \quad [9]$$

where,

$$\Xi_{Tx}(\omega_k) = \sqrt{\frac{1}{2\pi T}} \sum_{t=1}^T x_t \exp(-it\omega_k) \quad [10]$$

Assume there are m additional explanatory variables z_n , $n = 1, \dots, m$ that might help improve the real-time estimate of $\{y_T\}$ obtained with a univariate filter. Then, the second expression in the modulus on the second line of [9], $\hat{\Gamma}_X(\omega_k) \Xi_{Tx}(\omega_k)$, is replaced by:

$$\hat{\Gamma}_X(\omega_k) \Xi_{Tx}(\omega_k) + \sum_{n=1}^m \hat{\Gamma}_{z_n}(\omega_k) \Xi_{Txz_n}(\omega_k) \quad [11]$$

where,

$$\hat{\Gamma}_X(\omega_k) = \left(\sum_{j=0}^L b_{xj} \exp(-ij\omega_k) \right) \quad [12]$$

$$\hat{\Gamma}_{z_n}(\omega_k) = \left[\sum_{j=0}^L b_{z_nj} \exp(-ij\omega_k) \right] \quad [13]$$

are the one-sided transfer functions applied to the explanatory variables, $\Xi_{Tx}(\omega_k)$, $\Xi_{Txz_n}(\omega_k)$ are the corresponding DFTs, and $L \in \{1, 2, \dots, T-1\}$ is the filter length. Then, the multivariate version of [9] can be written as:

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \xi_k \left| \Gamma(\omega_k) - \hat{\Gamma}_X(\omega_k) \Xi_{Tx}(\omega_k) - \sum_{n=1}^m \hat{\Gamma}_{z_n}(\omega_k) \Xi_{Txz_n}(\omega_k) \right|^2 W(\omega_k) \quad [14]$$

The paper uses the filter obtained by minimising [14] (without $\hat{\Gamma}_X(\omega_k)$ since the lagged target series is not included in the set of the explanatory variables, see discussion below) with respect to filter parameters and subject to the amplitude constraint at frequency zero:

$$\sum_{j=0}^L b_{z_nj} = 1/m \quad [15]$$

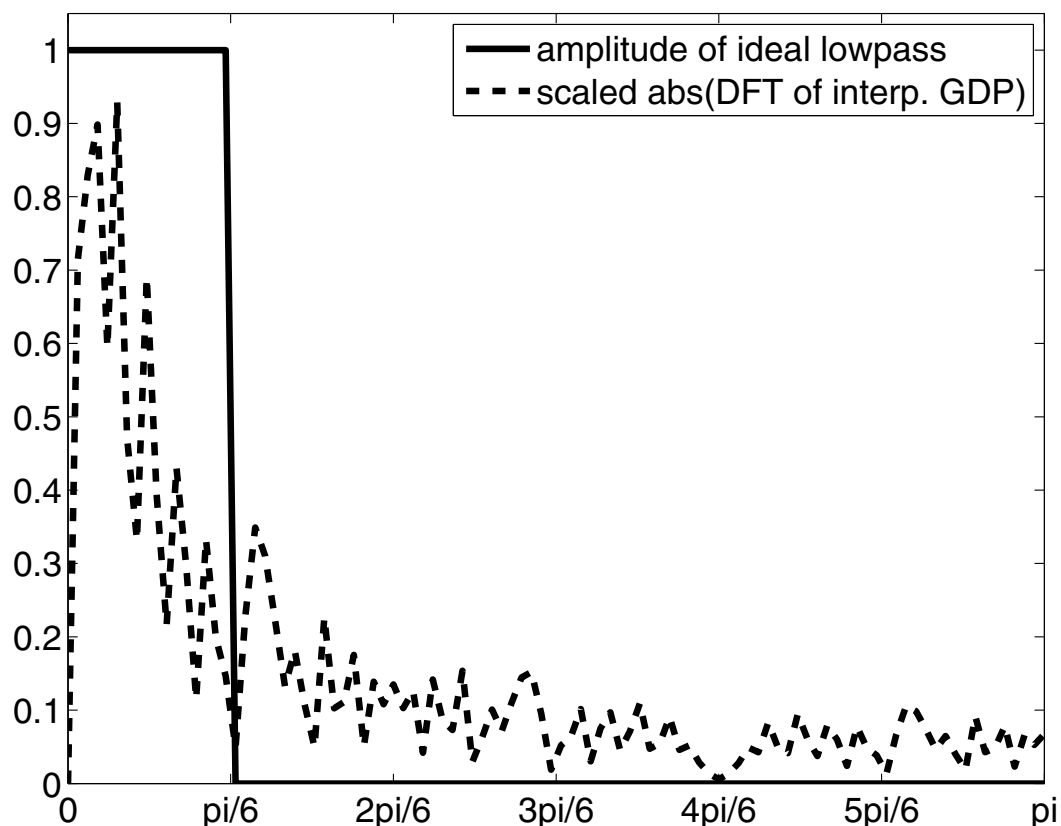
that imposes a constraint at frequency zero according to $\hat{\Gamma}_{z_n}(0) = 1/m$ for all $n = 1, \dots, m$, such that the sum of the amplitudes over the cross-sectional dimension of the filter, m , yields unity. Such a constraint ensures that the scale of the filter output matches the scale of the target signal. See Wildi (2011) for further details.

3. Indicator design

3.1. The target

The target is the medium-to-long-run fluctuations in the quarterly growth of the seasonally adjusted (SA) EA GDP in chain-linked prices. Specifically, the filter is set to target the ideal lowpass of the quarterly growth of SA EA GDP with the cut-off wave length of 12 months (see Figure 1).

Figure 1. **Amplitude of the ideal lowpass filter**



Note: The scaled absolute value of discrete Fourier transform (DFT) of differenced interpolated GDP is zero in frequency zero because the GDP series is demeaned.

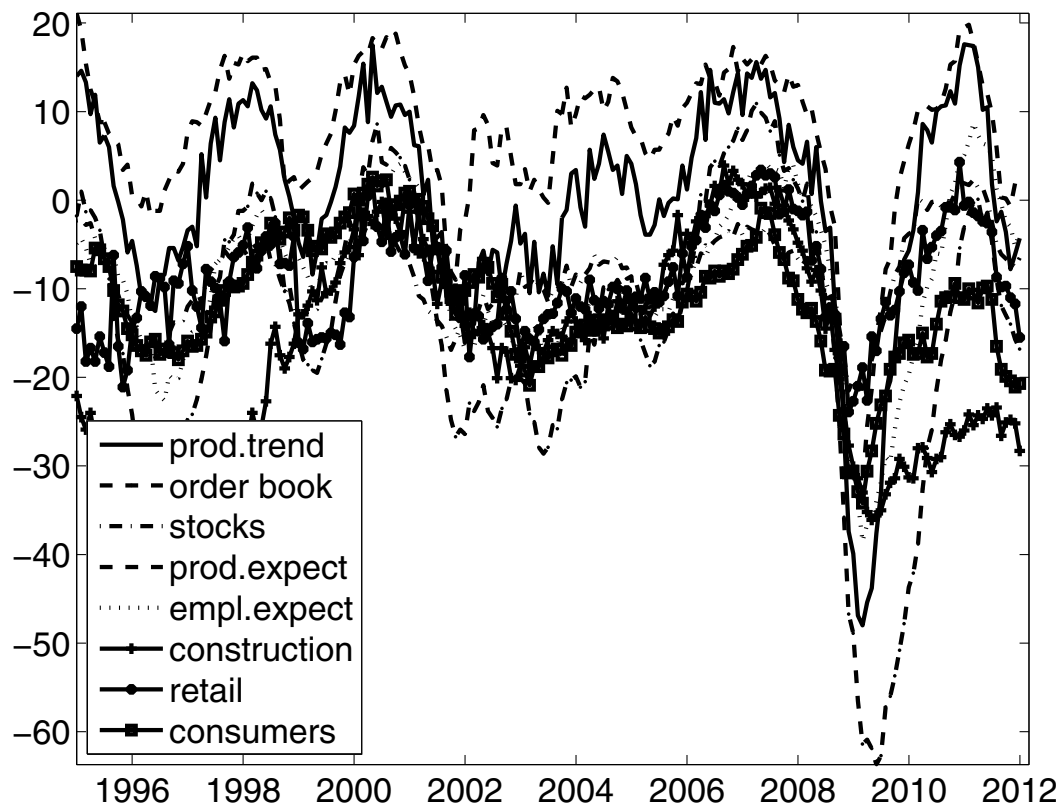
The typically defined minimum length of a business cycle is 18 months (Burns and Mitchell, 1946; Baxter and King, 1999; Christiano and Fitzgerald, 2003). However, passing higher-frequency content through the filter allows for both faster turning point detection and a closer scale fit. The quarterly GDP data are taken from 1995Q1 to 2011Q4. The GDP series is linearly interpolated to monthly frequency (by drawing a straight line between the two nearest observed monthly values in levels), logged, regularly differenced and demeaned before its spectral content enters the filter. Alternatively, the interpolation and differencing can be performed in frequency domain but the results show no obvious improvement with the latter, compared to such data transformation in time domain. Linear interpolation of quarterly GDP has little or no effect on the real-time performances of the filter because the interpolation affects the monthly frequencies, but the target of the filter (trend cycle with a cut-off length of one year) eliminates the information on the monthly frequencies.

3.2. Explanatory variables

Selected monthly confidence indicators published by the Directorate-General for Economic and Financial Affairs of the European Commission (DG ECFIN) and the United States (US) share price index published by Eurostat² are used as explanatory variables for the EA GDP. DG ECFIN data are usually published at the end of the reference month, except for December, for which data are published in early January. DG ECFIN business and consumer surveys data are almost unrevised – this applies both to the seasonally unadjusted and the seasonally adjusted data, as the latter is the product of the seasonal adjustment software Dainties that does not revise history as new data come in.³ The above-mentioned considerations make DG ECFIN data suitable for real-time filtration. The selected business and consumers confidence data are: production trend observed in recent months (industry), assessment of order-book levels (industry), assessment of stocks of finished products (industry), production expectations for the months ahead (industry), employment expectations for the months ahead (industry), confidence indicator in construction, confidence indicator in retail trade and consumers confidence indicator.

Other survey data have been tested and found not to add substantial quality to the indicator. SA DG ECFIN data are used (see Figure 2).

Figure 2. DG ECFIN survey data



Monthly US share price index as published by Eurostat is used as an additional explanatory variable due to the potential economic spillovers from the US. It is published about one week after the reference month. Since the real-time indicator, as presented in

this paper, is set to be published at the end of reference month, the US share price index enters the filter with a lag of one month. Thus, there is a room for potential improvement in this aspect, as financial data generally are available on the go.

The reason for limiting the set of explanatory variables to the above is the restrictive nature of the filter to deal with richer datasets due to the increasing number of free coefficients of the filter that worsens its out-of-sample performance, see discussion in Section 3.3.

Although the GDP series itself could be filtered, it is chosen not to do so as the GDP series is substantially lagged, happens to be revised substantially and appears not to be particularly useful for the concurrent signal extraction. Therefore, only the spectral content of the GDP series is fed to the filter and the monthly variables are set to target that spectral content.

All explanatory variables are taken from January 1995 to January 2012, regularly differenced and standardised to zero mean and unit variance. The list of data and their transformations is shown in Table 1.

Table 1. **The list of data used for the construction of the indicator**

Variable	Source	Transformation
Gross domestic product, chain-linked, SA	Eurostat	$\Delta \log$, lin. interp.
Production trend observed in recent month (industry), SA	DG ECFIN	Δ
Assessment of order-book levels (industry), SA	DG ECFIN	Δ
Assessment of stocks of finished products (industry), SA	DG ECFIN	Δ
Employment expectations for the months ahead (industry), SA	DG ECFIN	Δ
Production expectations for the months ahead (industry), SA	DG ECFIN	Δ
Confidence indicator in construction, SA	DG ECFIN	Δ
Confidence indicator in retail trade, SA	DG ECFIN	Δ
Consumers confidence indicator, SA	DG ECFIN	Δ
The US share price index	Eurostat	$\Delta \log$

Notes: 1) Δ denotes the first difference operator, that is $\Delta x_t = x_t - x_{t-1}$; 2) lin. interp. denotes linear interpolation defined in the main text.

3.3. Filter dimension

The filter dimension is defined by its length, L , and its cross-sectional dimension, that is the number of explanatory variables, m . A higher dimension of the filter leads to a higher number of free coefficients of the filter, which may lead to worse out-of-sample performance. Thus, the filter is tested for various dimension sizes. Unreported results of the author show that, given the available sample length, the cross-sectional dimension of one to four variables show comparable in-sample and out-of-sample performances but using more than four variables manifests the signs of overfitting, that is the out-of-sample performance seemingly deteriorates compared to the in-sample results. Therefore, the cross-sectional dimension of the filter is fixed to two variables.

For the same reason, the filter length is set to depend on the number of input variables (negatively) and the sample length (positively). The particular two-variable filter is set to be 38 to 48 observations long, depending on the in-sample length.

3.4. Noise suppression

Extra suppression of high-frequency content in the stopband is implemented with a positive $expw$ parameter. Particularly, $expw = 1/2$. As a result, the filter output is smoother than the one of the classical mean squared filter error problem, see discussion in Section 4.1.

3.5. Cross-sectional aggregation

Given that the considered set of explanatory variables contains more variables than the selected filter cross-sectional dimension, all possible combinations of variables are entered in the filter, such that the final filter output is combined 2-variable filter outputs weighted by weights proportional to inverse estimated mean squared filter error (EMSFE). Alternatively, one could use equal weighting but the resulting output might be more sensitive to the inclusion of irrelevant variables. As another alternative, principal components (Stock and Watson, 2002) could be used to shrink the dimension of the dataset to a few factors but unreported results of the author show no evident improvement in this regard. One explanation for this result might be the small size of the data set. However, the author suspects that there is more than that; the spectral content of the unobserved principal components might be subject to variation over different subsamples which might lead to inferior real-time performance.

The disadvantage of the cross-sectional aggregation procedure is that it does not take into account the full cross-sectional information available. The advantage of this procedure is that the errors of individual filters are partially cancelled out. This procedure resembles that of model combination in the time series forecasting literature.

3.6. Output re-scaling and adding the mean back

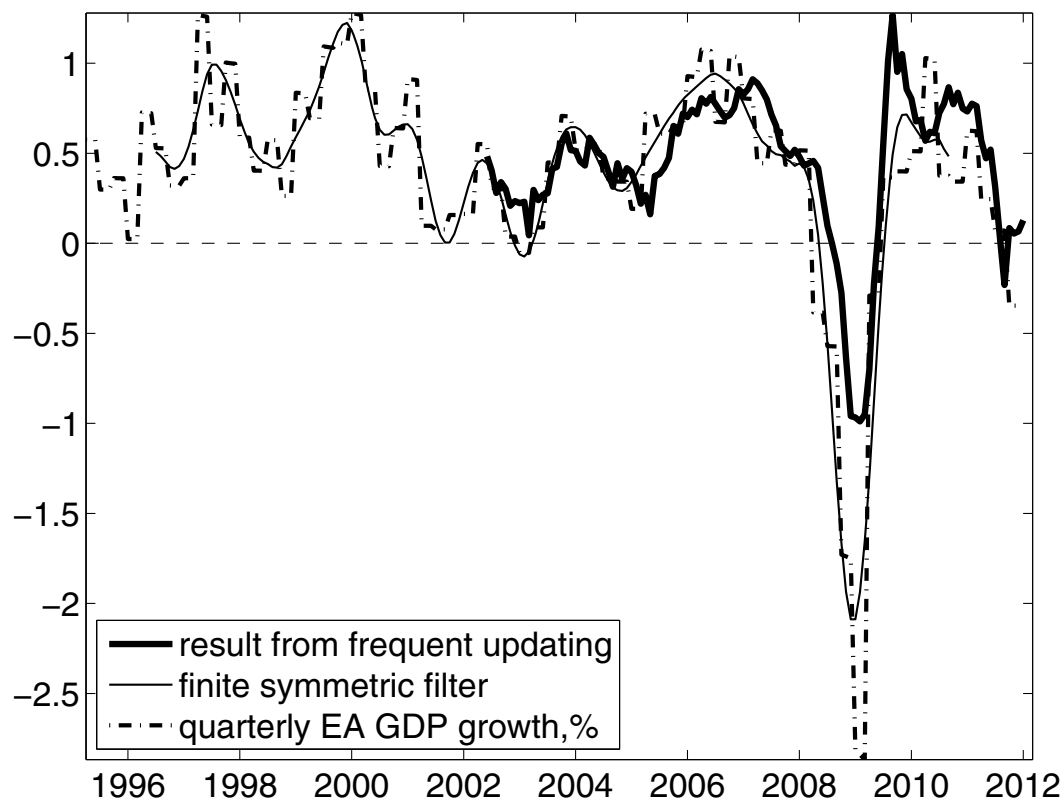
Filter output obtained thus far is then rescaled to the variance of the output of a 31 month long finite symmetric filter for the time span available at the particular real-time estimation moment. Finally, the mean of the output of the finite symmetric filter is added to the filter output. One could use the mean of the GDP series instead, but as it is more volatile, the output of the finite symmetric filter is preferred.

3.7. Averaging along time dimension

At this point, one can produce the real-time indicator. However, a couple of issues emerge. *First*, the GDP series is revised. Thus, some of its observations at the end of series, gap, are left out of sample before its spectral content is entered into the filter. Denote the whole sample length by len . Then, the maximum in-sample period spans observations $1:len - gap$. Setting gap to nine or 15 months does not make much of a difference but insures against GDP revisions at the end of series and ensures that the target is nearer the final GDP as opposed to the first-release. The indicator presented below is the result of setting gap to nine months.

Second, if filter coefficients are updated every time new data become available but the historical values of the indicator are not updated, then the noise generated by the changing estimated latent level of the target signal can suppress the estimated signal. Figure 3 shows an indicator resulting from such filter coefficient re-estimation every quarter. Although the filter output is timely, it is not smooth. The peak correlation (0.872) between the filter output and the GDP series is found to be at a one month lag with respect to GDP, and the second highest correlation (0.862) located at a zero months lag with respect to GDP (Table 2).

In this light, one might choose an in-sample span yielding the most satisfactory out-of-sample performance and fix coefficients. In this case, the real-time estimate is smooth. A real-time recession indicator for the US economy, USRI,⁴ is based on the latter approach. Yet, another issue that emerges is that sooner or later the filter coefficients may need an update due to possible structural changes in the economy.

Figure 3. **Filter output resulting from frequent filter coefficient update**

Note: The symmetric filter is 31 months long.

Table 2. **Dynamic correlations of indicators with GDP growth rates**

Indicators	Dynamic correlation at lag:		
	-2	-1	0
Eurocoin	0.880*	0.885**	0.853
Markit PMI	0.891**	0.880*	0.824
MDFA, $expw = 0$	0.824	0.882**	0.874*
MDFA, $expw = 0.5$	0.841	0.877**	0.850*
MDFA, $expw = 1$	0.846*	0.856**	0.811
MDFA in Figure 3	0.804	0.872**	0.862*

Notes: 1) The correlations are calculated from the longest common sample such that all the reported numbers are comparable; The sample of the GDP series spans from May 2007 to November 2011. 2) ** marks the peak correlation; * marks the second highest correlation.

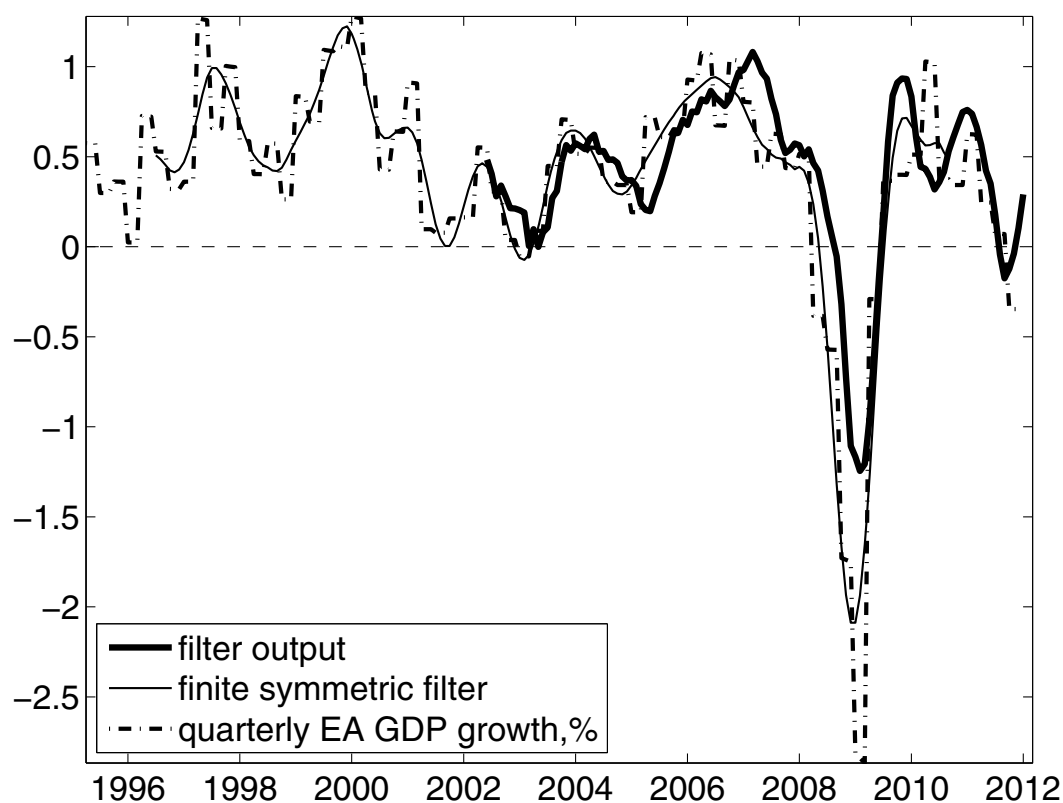
The above issues are solved with the following *ad hoc* filter averaging procedure along the time dimension. Update the filter coefficients every time the new GDP data become available. The final indicator is the average of all filter outputs created up to the estimation period. Thus, iteratively estimate filter coefficients for reduced in-sample length, $1:len - gap - 3$, $1:len - gap - 6$, etc., until some minimum in-sample length, but estimate filter outputs for the whole sample span, len . Then, take the arithmetic average⁵ of all historically estimated indicators. More indicators are averaged as more data come in but the maximum aggregation time span is set to five years in order to take into account possible structural changes. This updating mechanism overcomes the problem induced by

the level of target series and at the same time robustifies signal detection, and incorporates new information as new data become available. As Section 4 shows, this *ad hoc* procedure does not considerably change the timeliness or the scale of the filter output, as compared to that in Figure 3, but reduces the short-term fluctuations.

4. Results

The above mechanism is performed for the last ten years. Out-of-sample time span is limited by the filter and sample lengths. The resulting simulated real-time out-of-sample filter output is shown in Figure 4, along with the quarterly growth of the EA GDP and the output of a 31-months long finite symmetric trend cycle filter.

Figure 4. **Filter output compared to finite symmetric filter and EA GDP growth**



Note: The symmetric filter is 31 months long.

Figure 4 shows that the new indicator is slightly lagging in a historic perspective, and that its scope is comparable to that of the output of the finite symmetric filter except for the great recession episode. A visual inspection suggests that the indicator gets slightly faster after the great recession episode, when it appears to be coincident with GDP growth. Methodologically, the indicator is constructed as a coincident indicator. Furthermore, the indicator gets slightly smoother over time. The improvement in the smoothness over time could be explained by the increasing data length available to the filter. Moreover, the smoothness performance is partly due to the filter averaging along the time dimension. The first observation of the indicator in early 2002 is the result of a single cross-sectionally aggregated filter output; the next indicator values are averages of increasing number of

filter outputs along the time dimension until five years pass, when only the filter outputs over the last five years are averaged to account for the effect of different phases of a business cycle, as well as possible structural changes.

Quantitatively, the filter output cannot be compared to the ideal filter output since the latter is unavailable for finite samples, and its finite approximation might be imprecise for such a short sample span. Therefore, this paper calculates dynamic correlations between the filter output and the GDP series. The dynamic correlations capture well both the scale fit and the time shift. The peak correlation (0.877) between the filter output in Figure 4 and the GDP series is found to be at a one month lag with respect to GDP, and the second highest correlation (0.850) located at a zero months lag with respect to GDP (see Table 2).

4.1. Comparison with Eurocoin and Markit EA PMI

4.1.1. Eurocoin

Eurocoin (Altissimo et al., 2010) is an established coincident indicator for EA GDP. According to its website,⁶ Eurocoin is a smooth real time estimate of the quarterly growth of EA GDP, released at the end of reference month and is not revised since May 2009. It is the result of the generalised dynamic factor model (Forni et al., 2000, 2005; Forni and Lippi, 2001) applied to 2010 (about 145 series), according to Altissimo et al. In contrast, the new indicator is a result of a multivariate filter applied on only selected nine explanatory variables. Eurocoin and the new indicator have a few features in common, as well – both target a trend cycle in the quarterly growth of EA GDP as defined by a lowpass with the cut-off frequency of one year, and both are designed to be coincident indicators. As such, Eurocoin is a decent benchmark for the new indicator.

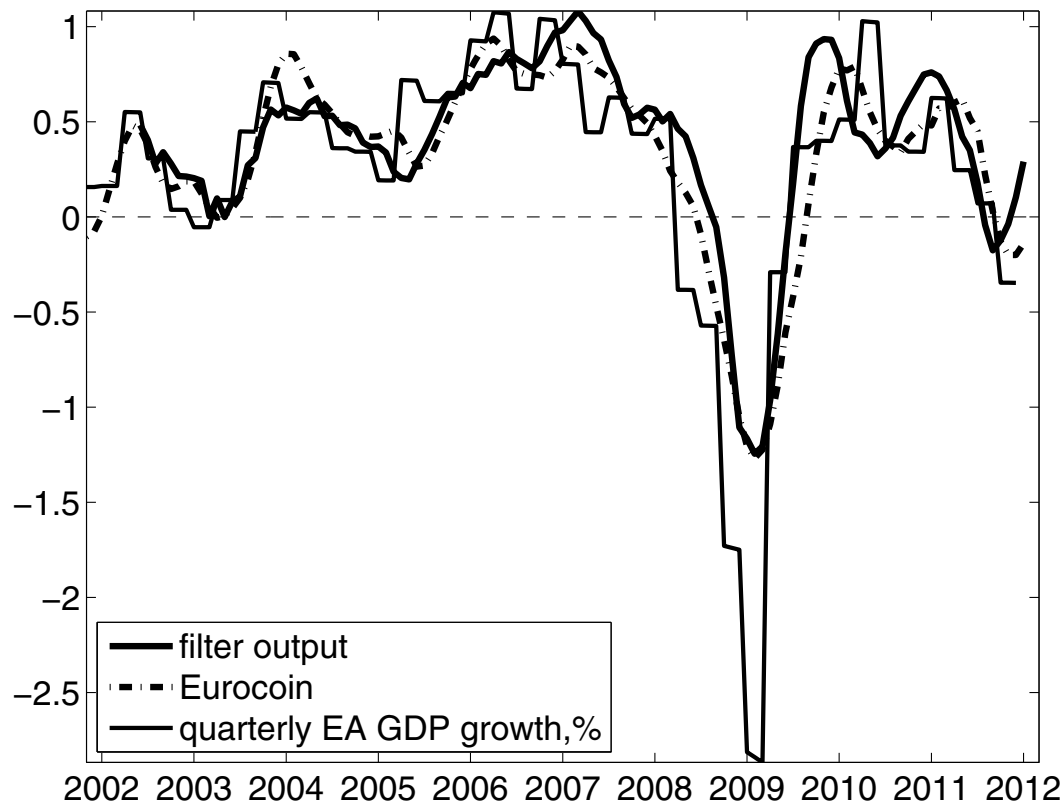
Figure 5 shows Eurocoin, latest vintage of the quarterly growth of EA GDP, and the filter output for the last ten years.

Figure 5 shows that the scope of both indicators is similar over the considered period. The new indicator is about coincident with Eurocoin during 2002-04, but is less smooth due to the small in-sample period. The new indicator appears to be faster than Eurocoin during 2005 and coincident, but with worse level of fit during 2007-08. Eurocoin gets slightly after 2009. Given that the true out of sample period for Eurocoin is only since May 2009, and that its pseudo real-time values are calculated on the last vintage data for the period before May 2009,⁷ the new indicator might indicate that the real-time performance of Eurocoin is slightly worse than its pseudo real-time estimates suggest, and that the new indicator slightly leads Eurocoin during the true real time performance of both indicators. Quantitatively, the peak correlation (0.885) between the Eurocoin and the GDP series is found to have a one month lag with respect to GDP, and the second highest correlation (0.880) at a two month lag with respect to GDP (see Table 2). Thus, the new indicator leads the Eurocoin on average by zero to one month.

4.1.2. Markit EA PMI

Markit EA PMI is revealed a few days after the end of reference month and is advertised as being closely correlated with the quarterly growth of EA GDP. PMI is one of the closely watched indices by economists due to its early release, simple design and economic relevance. The PMI data for the last five years are collected from Bloomberg⁸ and plotted against the new indicator. Since PMI is of different magnitude than the GDP growth, all variables are normalised to zero mean and unit variance for an easier comparison. Figure 6 shows that the new indicator is

Figure 5. Filter output compared to Eurocoin



about coincident with, but is smoother than PMI. Quantitatively, the peak correlation (0.891) between PMI and GDP is found to be at a two months lag with respect to GDP, and the second highest correlation (0.880) located at a one month lag with respect to GDP (see Table 2). Thus, the new indicator leads the PMI on average by about one month.

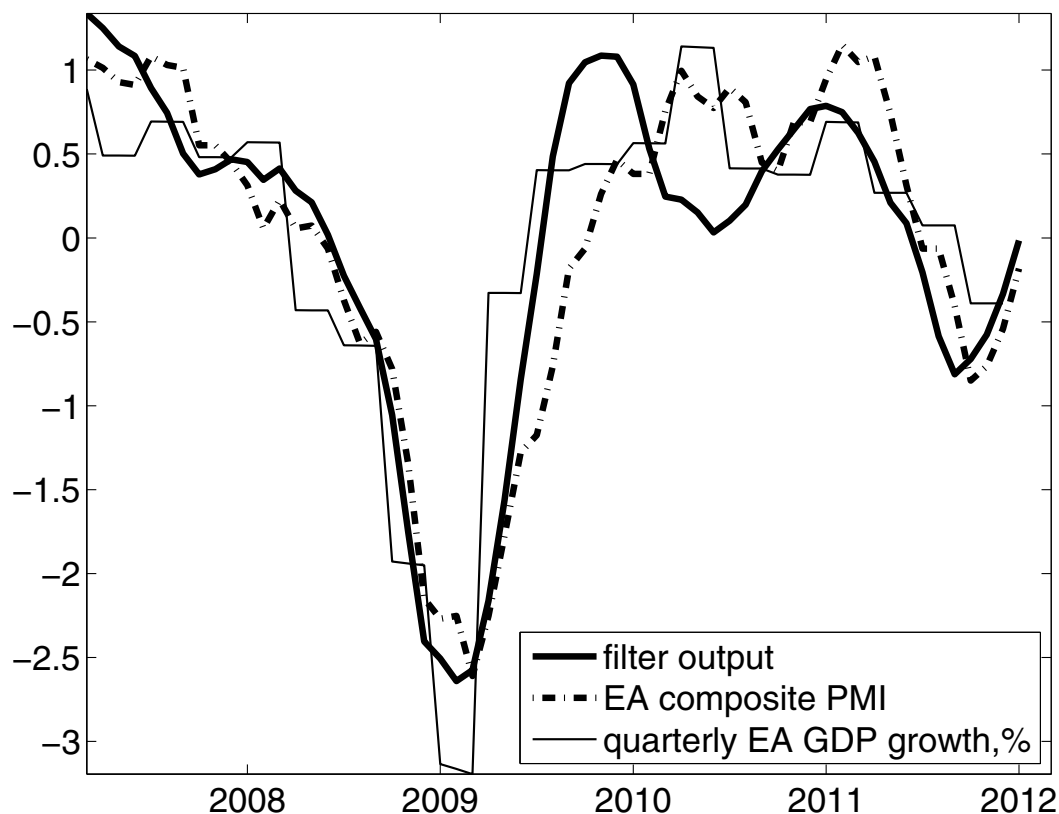
5. Conclusions

The paper applies the multivariate direct filter approach on selected business and consumer confidence indicators and share price data to construct a real-time indicator tracking the medium-to-long-run component of the quarterly growth of EA GDP.

The results show that the constructed indicator behaves similarly to another established indicator, Eurocoin, but leads it slightly after mid-2009. The new indicator is also compared to the Markit euro area composite purchasing managers index and is found to be leading it on average by about one month as well as being smoother.

The main conclusion of the paper is that the multivariate direct filter approach has its merits but it is not suitable to process rich datasets due to the increasing number of free coefficients and thus the deteriorating out-of-sample performance. In this regard, either the dimension of the dataset should be reduced prior to applying the filter, or the regularised version of the filter (Wildi, 2012) should be considered.

Figure 6. Filter output compared to Markit EA PMI



Notes: All series are normalised to zero mean and unit variance.

The apparent dip in the filter output right after 2010 is due to the rescaling; see Figure 5 for actual level.

Notes

1. See www.markiteconomics.com/Survey/Page.mvc/PressReleases, accessed on 26 Nov. 2012.
2. <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>, accessed on 26 Nov. 2012.
3. For details, see European Commission Directorate-General for Economic and Financial Affairs (2007), The Joint Harmonised EU Programme of Business and Consumer Surveys, User Guide, available at: http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf, accessed on 26 Nov. 2012.
4. See www.idp.zhaw.ch/usri, accessed on 26 Nov. 2012.
5. A weighted average proportional to inverse EMSFE also was tested and found not to yield superior results.
6. See <http://eurocoin.bancaditalia.it/>, accessed on 26 Nov. 2012.
7. See the note in http://eurocoin.cepr.org/files/file/Ecoin_realtime_99Feb12.xls, accessed on 26 Nov. 2012. An excerpt from the linked text: "Technically we Should Call this a 'Pseudo-Realtime Exercise' Since we Use the 'Final' Data in the Estimation and Not the First Releases Since they Are Not Available for All Data Series".
8. www.bloomberg.com, accessed on 26 Nov. 2012.

References

- Altissimo, F. et al. (2010), New Eurocoin: Tracking Economic Growth in Real Time, *The Review of Economics and Statistics*, 92[4]: 1024-1034, doi: http://dx.doi.org/10.1162/REST_a_00045.
- Altissimo, F. et al. (2003), Eurocoin: A Real Time Coincident Indicator for the Euro Area Business Cycle, *Computing in Economics and Finance* 2003, 242, Society for Computational Economics.
- Arnaud, B. and E.P. Hong (2001), Comparison of Compilation Methodologies for the Composite Leading Indicators of Euro Area, OECD, available at www.oecd.org/std/leading-indicators/1907370.pdf.
- Baxter, M. and R.G. King (1999), Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series, *The Review of Economics and Statistics*, 81[4]: 575-593, doi: <http://dx.doi.org/10.1162/003465399558454>.
- Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31[3]: 307-327, doi: [http://dx.doi.org/10.1016/0304-4076\(86\)90063-1](http://dx.doi.org/10.1016/0304-4076(86)90063-1).
- Bollerslev, T. and J.M. Wooldridge (1992), Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances, *Econometric Reviews*, 11[2]: 143-172.
- Brockwell, P.J. and R.A. Davis (1987), *Time Series: Theory and Methods*, New York, Springer Verlag.
- Burns, A.F. and W.C. Mitchell (1946), *Measuring Business Cycles*, New York, NBER.
- Buss, G. (2012), Forecasting and Signal Extraction with Regularised Multivariate Direct Filter Approach, *Working Papers*, 2012/06, Latvijas Banka.
- Caporello, G., A. Maravall and F.J. Sanchez (2001), Program TSW Reference Manual, *Banco de Espana Working Papers*, 0112, Banco de Espana.
- Christiano, L.J. and T.J. Fitzgerald (2003), The Band Pass Filter, *International Economic Review*, 44[2]: 435-465, doi: <http://dx.doi.org/10.1111/1468-2354.t01-1-00076>.
- European Commission Directorate-General for Economic and Financial Affairs (2007), The Joint Harmonised EU Programme of Business and Consumer Surveys, User Guide, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf, accessed on 26 Nov. 2012.
- Forni, M. et al. (2000), The Generalized Dynamic Factor Model: Identification and Estimation, *The Review of Economics and Statistics*, 82[4]: 540-554, doi: <http://dx.doi.org/10.1162/003465300559037>.
- Forni, M. et al. (2005), The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting, *Journal of the American Statistical Association*, Vol. 100, 830-840.
- Forni, M. and M. Lippi (2011), The Generalized Dynamic Factor Model: Representation Theory, *Econometric Theory*, 17(06): 1113-1141.
- Gyomai, G. and M. Wildi (2012), OECD Composite Leading Indicators for G7 Countries: A Comparison of the Hodrick-Prescott Filter and the Multivariate Direct Filter Approach, *OECD Statistics Working Papers*, 2012/5, OECD Publishing, doi: <http://dx.doi.org/10.1787/5k4dhnzw8xr3-en>.
- Hodrick, R. and E.C. Prescott (1997), Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit, and Banking*, 29[1]: 1-16.
- Jarque, C.M. and A.K. Bera (1987), A Test for Normality of Observations and Regression Residuals, *International Statistical Review*, 55[2]: 163-173.
- King, R.G. and S.T. Rebelo (1993), Low Frequency Filtering and Real Business Cycles, *Journal of Economic Dynamics and Control*, 17(1-2): 207-231, doi: [http://dx.doi.org/10.1016/S0165-1889\(06\)80010-2](http://dx.doi.org/10.1016/S0165-1889(06)80010-2).
- Kolmogorov, A.N. (1941), Stationary Sequences in Hilbert Space (in Russian), *Bull. Moscow University*, 2[6]: 1-40.
- Massey, F.J. (1951), The Kolmogorov-Smirnov Test for Goodness of Fit, *Journal of American Statistical Association*, 46(253): 68-78.
- Maravall, A. and A. Rio (2001), Time Aggregation and the Hodrick-Prescott Filter, *Banco de Espana Working Papers*, 0108, Banco de Espana.
- Stock, J.H. and M.W. Watson (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, 20[2]: 147-162.
- Wiener, N. (1949), *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press.
- Wildi, M. (2008), *Real-Time Signal-Extraction: Beyond Maximum Likelihood Principles*, available at http://blog.zhaw.ch/idp/sefblog/uploads/Wildi_Real_Time_SE_0810010.pdf, accessed on 26 Nov. 2012.

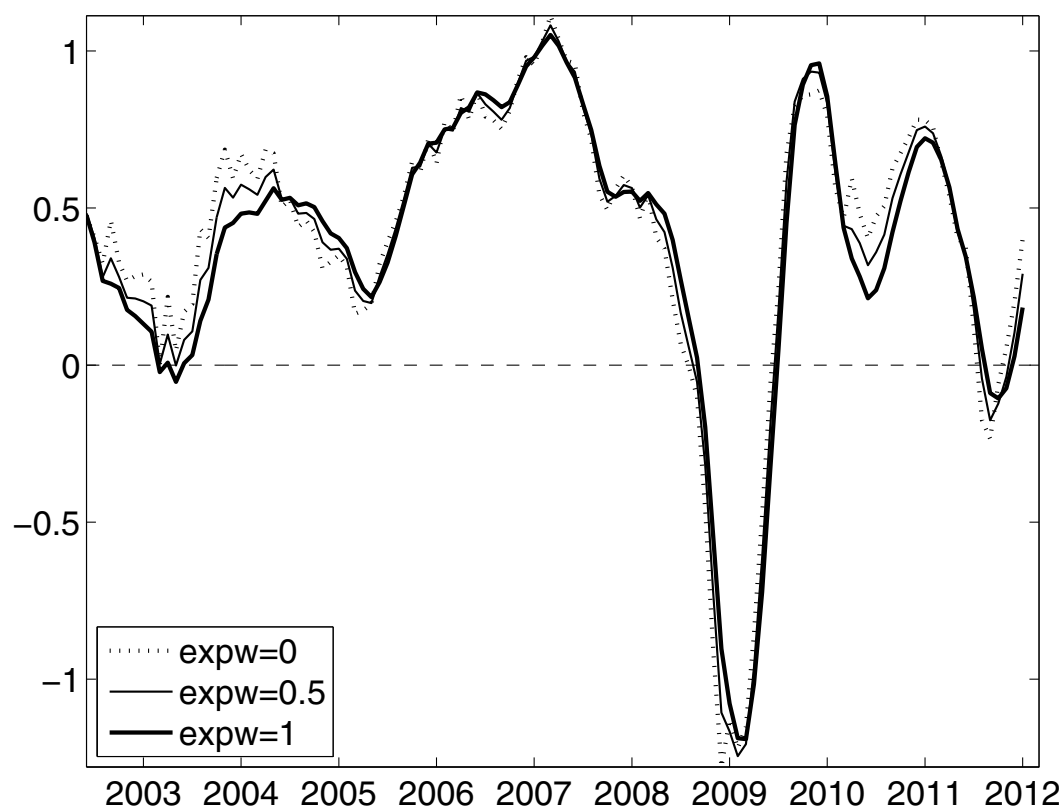
- Wildi, M. (2011), I-DFA and I-MDFA: Companion Paper to R-Code Published on SEFBlog, *IDP-Working Paper*, available at: http://blog.zhaw.ch/idp/sefblog/uploads/working_paper1.pdf, accessed on 26 Nov. 2012.
- Wildi, M. (2012), Elements of Forecasting and Signal Extraction, *IDP-Working Paper*, available at: http://blog.zhaw.ch/idp/sefblog/uploads/elements_3007.pdf, accessed on 26 Nov. 2012.
- Wildi, M. and J.E. Sturm (2008), Real-Time Filtering: Using the Multivariate DFA to Monitor the US Business Cycle, *IDP-Working Paper*, available at: www.idp.zhaw.ch/fileadmin/user_upload/engineering/_Institute_und_Zentren/IDP/forschungsschwerpunkte/FRME/sef/signalextraction/papers/IDP-WP-08Sep-02.pdf, accessed on 26 Nov. 2012.

ANNEX A

*Robustness checks***A.1. The effect of high-frequency noise suppression**

Figure A1 shows the effect of the high-frequency noise suppression in the rightmost stopband by the customised filter.

Figure A1. **The effect of high-frequency noise suppression**



Note: The final indicator presented in this paper is the middle case with $expw = 0.5$.

The dotted line corresponds to the classical mean squared optimisation problem with $W(\omega_k) = 1$, that is, there is no extra noise suppression. The other two lines correspond to the customised filter output with $expw = 0.5$ and $expw = 1$. It can be seen that the indicator without the extra noise suppression is the fastest but also the noisiest of the three. The

user of a real-time indicator might want to trade timeliness for extra smoothness such that the real-time signal would be more reliable. In that case, the filter customisation helps by suppressing high-frequency content in the stopband. Ideally, that extra noise suppression would not alter the phase function in the passband but in practice, it does to some extent, that is, the additional noise suppression in the stopband slows down the filter output somewhat. Quantitatively, the peak correlation (0.882) between the filter output with $expw = 0$ and the GDP series is found to be at a one month lag with respect to GDP, and the second highest correlation (0.874) located at a zero months lag with respect to GDP (see Table 2). As for the filter output with $expw = 0.5$, the corresponding numbers are 0.877 (at a one month lag) and 0.850 (at a zero months lag). The peak correlation (0.856) between the filter output with $expw = 1$ and the GDP series is found to be at a one month lag with respect to GDP, and the second highest correlation (0.846) located at a two months lag with respect to GDP. The presented final indicator is the middle case with $expw = 0.5$.

A.2. Other data and transformations

Other additional data (for example, other DG ECFIN data, new registered cars, industrial production, producer and consumer price indices and other financial variables) were tested and found not to yield conceptually different results; their usage is under consideration but excluded from the paper for simplicity and the lack of proper vintage data.

Both seasonally adjusted and unadjusted survey data were tested as inputs. Preliminary results have shown that the use of seasonally adjusted survey data by Dainties, as published by DG ECFIN, for filter inputs gives the most satisfactory results. Another nuance is whether to use seasonally adjusted or unadjusted GDP for spectral estimates. If the target is lowpass with cut-off to the left of the one-year frequency, both could be used without much difference in outcome. However, since the cut-off is set to the seasonal frequency, one year, seasonally adjusted GDP is used for the target spectral estimate.

A.3. Data revisions

The seasonal adjustment method used by the DG ECFIN Dainties to seasonally adjust the survey data does not induce revisions. The last month of the non-adjusted data might be slightly revised, thus inducing revisions in SA data. However, these revisions are negligible and thus are ignored.

Another source of potential data revision is revisions in GDP. However, the solution to the revised GDP series is straightforward; the GDP series itself is not filtered, instead, only its spectral content is fed to the filter, therefore, a robust filter output is obtained by feeding the filter with estimated spectra of the final/revised GDP by dropping several observations from the end of the series. That is, the described method does not use uncertain first releases, and dropping the last three or five quarters from the estimation routine gives similar results.

Therefore, the plotted indicator's performance is expected not to deteriorate with new real-time data observations entering the estimation routine.

A.4. Amplitude constraint

The amplitude constraint is imposed by setting the filter amplitude to unity at frequency zero. Strictly speaking, the amplitude constraint is not necessary for stationary data. However, unreported results of the author suggest that an unconstrained filter tends

to yield inferior real-time performance with respect to the scale fitting compared to the constrained filter, particularly so for small samples and during the great recession period.

A.5. Filter updating frequency

The default filter updating period is set to every quarter, as new GDP data become available. Filter updating is performed in order to take into account the potential structural changes. Moreover, the indicator has been checked for different, less frequent, filter updating frequency and is found to be robust, which is an expected result since the spectral content of the GDP data in the relevant passband is not expected to change rapidly.

What's behind the survey values? An analysis of individual forecasters' behaviour

by

Maritta Paloviita and Matti Viren*

This paper studies the internal consistency of professional forecasts on a micro level using three alternative data sets. The analysis is mainly based on the ECB Survey of Professional Forecasts for the euro area, but for comparison we also study the Consensus Economics survey and Survey of Professional Forecasts for the US. Forecast uncertainty is explored using two alternative measures, the conventional standard deviation of individual point forecasts and the mean uncertainty based on subjective probability distributions of survey respondents. Our analysis indicates that individual forecasters' price and real GDP expectations are positively related, but that forecasters deviate systematically from each other. We also find clear evidence that individual forecasters form expectations according to the hybrid specification of the New Keynesian Phillips curve. On a micro level, inflation uncertainty seems to be closely related to output uncertainty. However, the relationship between alternative measures of uncertainty is relatively weak.

Keywords: Forecasting, Survey data, Expectations, Phillips curve

JEL Classification: C53, E37, E31

* Maritta Paloviita is a research economist at the Bank of Finland and Matti Viren a scientific advisor at the Bank of Finland and professor of economics at University of Turku. Emails: maritta.paloviita@bof.fi and matti.viren@bof.fi. Antti Komonen, Sami Oinonen and Tarja Yrjölä have kindly derived the data and Mika Kortelainen has provided useful comments. Comments from two anonymous referees are also gratefully acknowledged.

1. Introduction

Survey data have been widely used to study inflation expectations and other macroeconomic variables (see Pesaran and Weale, 2006; and Sinclair, 2010 for basic references of survey-based studies). Recently, increased macroeconomic volatility during the financial crisis has emphasised the need to analyse forecast uncertainty. An obvious way to measure uncertainty is to use survey data, which are nowadays collected in several studies.

Typically, surveys publish only mean or median values of individual forecasts, along with corresponding conventional standard deviations as a measure of expectations disagreement. However, the ECB Survey of Professional Forecasters (ECB SPF) enables a richer analysis of forecast uncertainty, based on probability distributions of individual respondents' expectations.¹ Contrary to many other surveys, the ECB SPF survey provides both *fixed event* and *fixed horizon* forecasts for different time horizons (terminology is from Dovern et al., 2012). *Fixed event* forecast refers to a certain calendar year (for example, the next calendar year) and *fixed horizon* forecast to a certain time period ahead (e.g. four quarters ahead). In some studies, fixed horizon forecasts are approximated using weighted averages of fixed event forecasts (see e.g. Gerlach, 2007).

The main purpose of this study is to investigate the internal consistency of individual survey expectations. By internal consistency we mean that expectations and actual data behave in the same way for different variables like inflation and output growth. Thus, if these variables are positively (strongly/weakly) correlated in the actual data, we would expect that the corresponding expected values are also positively (strongly/weakly) correlated. We examine that with both individual inflation and output growth expectations and expectation errors. This is done by scrutinising simple bi-variate relationships at different forecast horizons and using the standard New-Keynesian Phillips curve as the underlying specification for both sets of variables. This comparative analysis is the main purpose of the paper. We also compare alternative measures of forecast uncertainty for different variables and scrutinise the relationships between inflation and output growth uncertainties. This is the second purpose of the paper. In addition to these analyses of different relationships, we carry out some more technical analyses: we compare fixed horizon forecasts to corresponding approximations based on weighted averages of fixed event forecasts. Finally, we examine the heterogeneity of individual expectations. In all analyses, special attention is paid to the impact of the recent financial and Euro crisis on individual expectations, especially in terms of forecast uncertainty.

The empirical analysis is mainly based on the ECB Survey of Professional Forecasts for the euro area. For comparison, we also consider the Consensus Economics survey (CF) and Survey of Professional Forecasts for the US (US SPF). Both fixed event and fixed period expectations are used in these analyses. Two alternative measures of forecast uncertainty are also employed: the standard deviation of point forecasts and the mean uncertainty based on probability distributions of survey respondents. These measures are typically used for policy purposes, but in this context relatively little attention is paid to uncertainty

measures and other distributional issues (see Schnader and Stekler, 1991; and Kolb and Stekler, 1996 for a more thorough analysis of consensus in forecasting).²

Our results suggest that surveys are useful for analysing changes in inflation and inflation uncertainty, albeit some measurement problems are involved. For instance, we find that basing forecasts on weighted averages of fixed event forecasts may not be the optimal way to approximate fixed horizon forecasts. The analyses provide evidence of systematic differences across individual forecasts. On a micro level, future price and real GDP growth expectations are positively related, and individual forecasters seem to form expectations according to the hybrid specification of the New Keynesian Phillips curve. As regards uncertainty, inflation uncertainty seems to be closely related to output uncertainty. However, the relationship between alternative uncertainty measures is relatively weak.

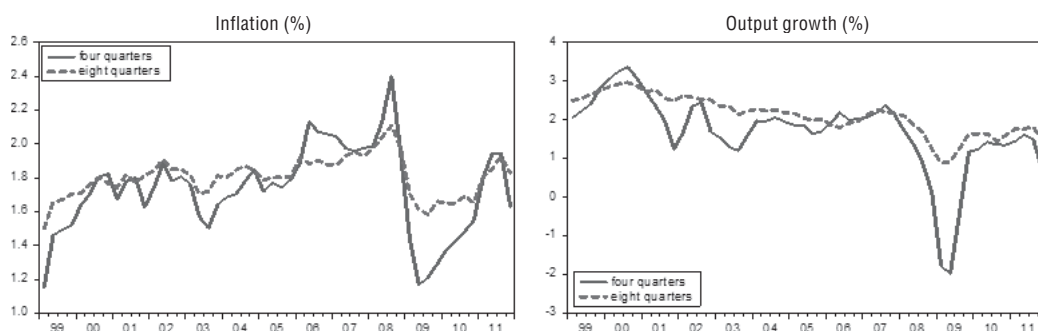
Next, we consider the basic features of the data, then in Section 3 we analyse the relationships between different survey variables, in Section 4 we scrutinise the uncertainty proxies and in Section 5 we present some concluding remarks.

2. Alternative data sets

Of the numerous survey sources of micro-level survey data we use the quarterly Survey of Professional Forecasters of the European Central Bank (ECB SPF) as our main data source.³ Since 1999, the ECB has asked a panel of some 75 forecasters about their short- and long-term views of HICP inflation, real GDP growth and unemployment in the euro area. Respondents represent the financial sector, non-financial research institutes and employer or employee organisations in the European Union (EU). Both *fixed event* forecasts and *fixed horizon* forecasts for different time horizons are surveyed. In addition to the point estimates, probabilities surrounding point estimates (density forecasts) are published for all variables and all horizons.⁴

Mean expectations of inflation and real GDP growth one and two years ahead in the ECB SPF survey are displayed in Figure 1. Until mid-2008, the forecasts were quite stable, but after that the price and output forecasts decreased sharply and the forecasting errors expanded substantially. After 2009, the expectations and expectations errors returned to more or less “normal levels”. As shown in Figure 1, long-term forecasts are typically more stable than short term forecasts.

Figure 1. **Forecasts of inflation and output growth for one and two years ahead (ECB SPF)**



For the sake of comparison, two surveys for the US are also analysed: Consensus Forecasts survey provided by Consensus Economics (CF) and the Survey of Professional Forecasts provided by the Federal Reserve Bank of Philadelphia (US SPF). Two surveys are used because they are both somewhat different from the ECB SPF survey (see Annex A). The CF survey, which has been conducted since 1989, publishes forecasts for the US and many other industrialised countries every month. Respondents in this survey are public and private economic institutions in all major economies. The US SPF is a quarterly survey, which began already in 1968Q4. It includes over 30 variables for different forecast horizons. Both surveys provide expectations on a micro level. Consensus Forecast provides only *fixed event* forecasts of several macroeconomic variables while the US SPF survey publishes both *fixed event* and *fixed horizon* forecasts.

We follow Gerlach (2007) and Dovern et al. (2012) and approximate *fixed horizon* forecasts by weighted averages of *fixed event* forecasts. Denote by $F[y_0, m, y_1(x)]$ the *fixed event* forecast of variable x for year y_1 made in month m of the previous year, y_0 , and by $F[y_0, m, 12(x)]$ the *fixed horizon*, twelve-month-ahead forecast made at the same time. We can then approximate the *fixed horizon* forecast for the next twelve months as the average of forecasts for the current and next calendar years weighted by their shares in the forecasting horizon:

$$F[y_0, m, 12(x)] = [(12 - m)/12] * F[y_0, m, y_0(x)] + (m/12) * F[y_0, m, y_1(x)] \quad [1]$$

For example, the July 2010 twelve-month-ahead forecast of inflation rate $F[2010, 7, 12(\Delta p)]$ is approximated by the sum of $F[2010, 7, 2010(\Delta p)]$ and $F[2010, 7, 2011(\Delta p)]$ weighted by 5/12 and 7/12 respectively. Using Equation [1], we construct expected inflation and real GDP growth series for both the US (using the CF survey) and the euro area (using the ECB SPF survey).

Next, we assess the empirical relevance of Equation [1] using the ECB SPF survey. Figure 2 illustrates the relationship between fixed horizon data and constructed series, and Table 1 provides a test of the equality of the series. It is a bit disturbing to see that the original *fixed horizon* data seem to differ from the constructed series. The deviation is quite

Figure 2. **Relationship between survey forecasts four quarters ahead and corresponding approximations based on Equation [1] (ECB SPF)**

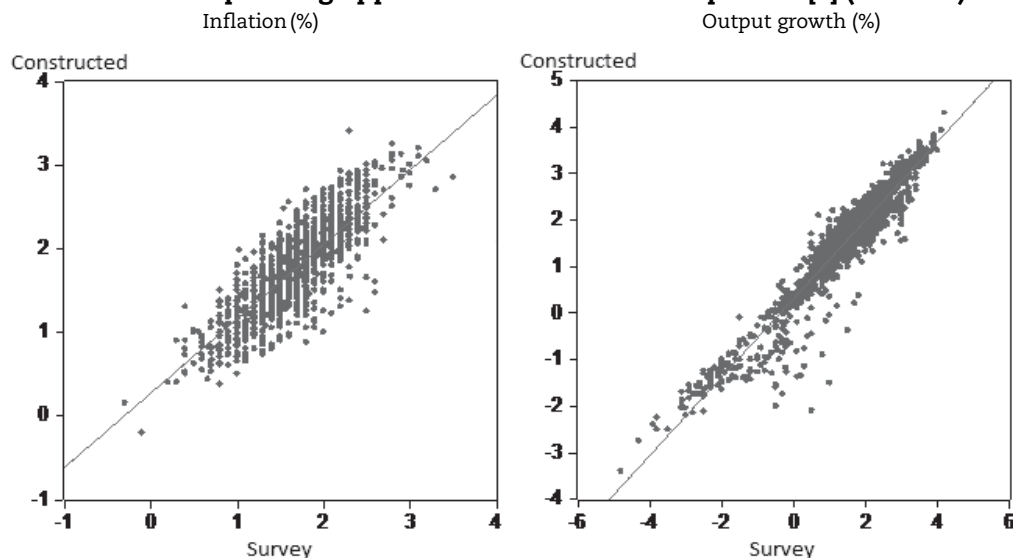


Table 1. **Comparison of actual and constructed forecast data in the ECB SPF survey**

Test of equality of the ECB SPF time series

	t1	t2	R2	SEE	DW	F
$\Delta p_{it,4}^e = .335 + .770\Delta p_{it,4g}^e$	14.27	61.61	0.684	0.219	1.23	
$\Delta p_{it,4}^e = -.418 + 1.185\Delta p_{it,4g}^e$	11.13	57.61	0.800	0.174	1.55	29.38
$\Delta y_{it,4}^e = -.207 + 1.085\Delta y_{it,4g}^e$	7.35	84.43	0.915	0.320	1.49	
$\Delta y_{it,4}^e = -.308 + 1.144\Delta y_{it,4g}^e$	4.78	31.22	0.940	0.271	1.58	20.78

Notes: $\Delta p_{it,4}^e$ ($\Delta y_{it,4}^e$) denotes the original inflation (output growth) forecast four quarters ahead in the survey and $\Delta p_{it,4g}^e$ ($\Delta y_{it,4g}^e$) the corresponding four-quarter forecast from the fixed event (current and next calendar year) forecasts computed via Equation [1]. t1 and t2 denote the (corrected) t-ratios of the coefficients and the F test statistic for fixed period time effects. All estimates are OLS estimates which include cross section fixed effects. Equations [2] and [4] also include period fixed effects.

noticeable for inflation, whereas for output growth the difference is more moderate. But in both cases equality of the two series is clearly rejected (Table 1). The results for fixed time effects (Equations [2] and [4] in Table 1) suggest that to some extent the difference is related to specific periods. The results can be interpreted in many ways, but the most obvious interpretation is that, instead of making mechanical calculations, forecasters do indeed have in mind some nontrivial path of future inflation and output growth when they form expectations. Therefore, in empirical studies, expectations based on weighted averages should be interpreted with caution.

3. Empirical evidence

In this section, we explore the internal consistency of survey expectations on a micro level. More precisely, we investigate individual forecasters' trade-off between inflation and real activity. In practice, this means estimating the Phillips curve. Phillips curves have been estimated extensively, even with micro data. Micro-level analysis has not been done for the euro area, whereas Fendel et al. (2011) have used individual Consensus Forecast data to explore the Phillips curve relationship in G7 countries. For the US, individual FOMC forecasts have been used to study the relevance of the Phillips curve (Tillmann, 2010) and to examine the Taylor rule in monetary policy (Fendel and Rülke, 2012). Earlier in Kortelainen et al. (2011b), we used a VAR model to study the relationships between different forecast variables (expectations). The results of this analysis appeared to be consistent with the idea of a New-Keynesian macro model, in particular with a New-Keynesian Phillips curve.

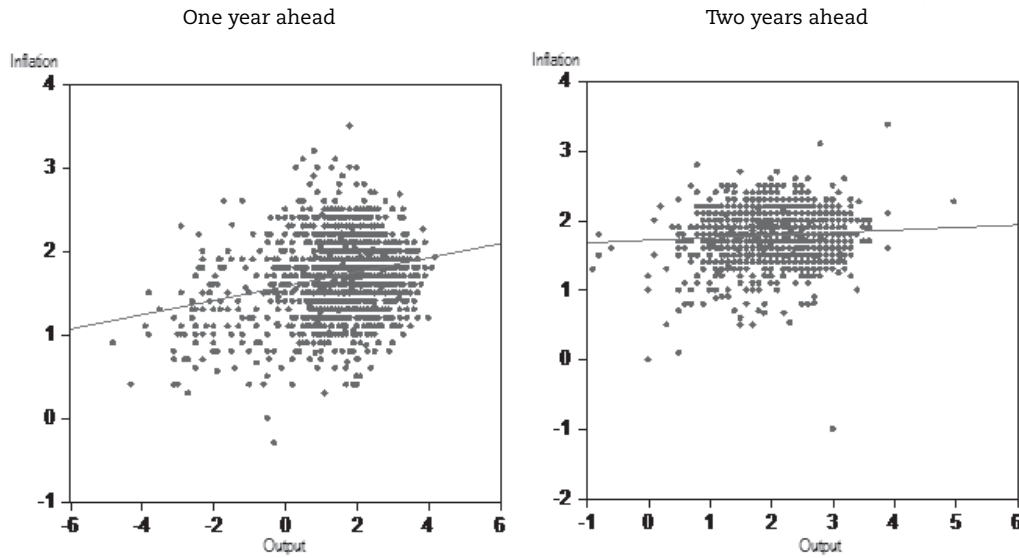
First, we scrutinise cross-plots of individual inflation and real GDP growth forecasts in the ECB SPF survey. One year- and two years-ahead relationships are displayed in Figure 3. Both panels indicate that individual forecasters have quite heterogeneous views about the future trade-off between inflation and output. Views regarding the Phillips curve seem to be even more divergent for the shorter forecast horizon. Even so, the two variables are positively related at both forecast horizons.

Next, we examine in more detail the heterogeneity of individual forecasts in the ECB SPF survey. For that purpose, we test the importance of cross-section fixed effects and period fixed effects in the panel regression:

$$\Delta p_{it,T}^e = c + c_i + c_t + u_{it} \quad [2]$$

where $\Delta p_{it,T}^e$ denotes inflation expectations of forecaster i for period T , assessed in period t . The term c is a constant and c_i refers to cross section fixed effects, c_t to period

Figure 3. Relationship between inflation and output growth forecasts (ECB SPF)



fixed effects. The error term, which is assumed to be uncorrelated, is denoted by u_{it} . As dependent variables, we consider forecasts for fixed calendar years $T = t + 0$, $T = t + 1$, $T = t + 2$ or $T = t + L$ (where L refers to long term) as well as fixed horizon forecasts for four quarters ahead ($T = t + 4$), eight quarters ahead ($T = t + 8$) and 20 quarters ahead ($T = t + 20$). In all cases, we test the hypothesis according to which both the cross section fixed effects c_i and period fixed effects c_t are the same across forecasters (see F test statistics in Table 2).

Table 2. Testing the importance of fixed effects in the ECB SPF survey

Dependent variable	Degrees of freedom	F	χ^2
Fixed event forecasts			
Δp_{t+0}^e	99 2880	2.71	152
Δp_{t+1}^e	98 2806	6.09	264
Δp_{t+2}^e	93 1034	4.35	258
Δp_{t+L}^e	97 1993	10.93	812
Fixed horizon forecasts			
Δp_{t+4}^e	96 2551	5.95	270
Δp_{t+8}^e	96 2771	6.44	422
Δp_{t+20}^e	65 62	4.50	214

Notes: Tests are based on OLS regression (2). The SUR estimator could not be used for the whole data set, because of missing observations (unbalanced panel). A balanced panel with 21 forecasters was used, however. It turned out that with these data, the test statistics became much larger (e.g., with Δp_{t+4}^e the F statistic increased from 4.25 to 22.47).

We also test the restriction that the relationship between inflation and output growth forecast is the same for all forecasters (see Chi-squared test statistics in Table 2). All test statistics indicate rejection of the null hypothesis of equal coefficients. Only for current period values (first row) do the significance levels come close to 0.001, which indicate that the difference between individual coefficients is not of several light-years' magnitude. Otherwise, the differences between individual forecasts are large and persistent.

After dealing with these measurement issues, we explore how the survey respondents' views of future price and output developments are related to each other in the ECB SPF survey (see Table 3). First, using alternative fixed event and fixed horizon forecasts, we

Table 3. Relationship between inflation and output growth forecasts in the ECB SPF survey

Dep. var.	Const.	Δy^e_{t+4}	$\frac{\Delta y_{t+4} - \Delta y^e_{t+4}}{\Delta y^e_{t+4}}$	Δy^e_{t+8}	Δy^e_{t+0}	Δy^e_{t+1}	y^e_{t+2}	Δy^e_{t+L}	R ² /SEE
Fixed horizon forecasts									
Δp^e_{t+4}	1.596 (96.27)	0.082 (10.20)							0.169 0.358
$\Delta p_{t+4} - \Delta p^e_{t+4}$	0.453 (25.56)		0.258 (22.06)						0.287 0.663
Δp^e_{t+8}	1.697 (57.68)			0.049* (3.75)					0.196 0.270
Fixed event forecasts									
Δp^e_{t+0}	1.628 (127.81)				0.219* (33.29)				0.280 0.561
Δp^e_{t+1}	1.607 (62.35)					0.090* (7.56)			0.137 0.343
Δp^e_{t+2}	1.721 (39.03)						0.050* (2.54)		0.230 0.248
Δp^e_{t+L}	2.030 (36.66)							-0.057* (2.21)	0.342 0.196
Phillips curve estimates									
		Δy^e_t	Δp^e_{t+1}	$s1 * \Delta p^e_{t+1}$	$s2 * \Delta p^e_{t+1}$	$s3 * \Delta p^e_{t+1}$	$s4 * \Delta p^e_{t+1}$	$\Delta^e p_{t-1}$	R ² /SEE
Δp^e_t		0.108 (16.84)	1.003 (127.11)						0.508 0.457
Δp^e_t		0.111 (17.32)		0.364 (17.42)	0.511 (25.72)	0.551 (26.17)	0.485 (22.91)	0.485 (22.54)	0.694 0.348
Δp^e_t		0.130 (17.32)		0.268 (11.07)	0.411 (18.13)	0.450 (18.63)	0.401 (15.91)	0.524 (21.73)	0.682 0.352
Δp^e_t		0.151 (11.32)		0.858 (55.17)					0.507 0.329
Δp^e_t		0.088 (5.98)			1.010 (60.00)				0.482 0.429
Δp^e_t		0.103 (8.69)				1.071 (77.85)			0.618 0.456
Δp^e_t		0.119 (9.68)					1.045 (67.64)		0.465 0.528

Notes: Numbers inside parentheses are corrected t-ratios. Δp_{t+4} denotes expected inflation for the subsequent four quarters and Δp_{t+1} the corresponding measure for the next calendar year. Δp_{t+L} denotes the long-run inflation expectations. The growth rate of output, Δy^e_t , is defined accordingly. The last seven rows represent a (hybrid) Phillips curve. All equations were estimated by OLS, the simple equations on rows 1-6 also include cross-section fixed effects. s1-s4 are seasonal dummies. Equations in the last four rows are estimated with the first, second, third and fourth quarter data only. Δp and Δy denote actual inflation and output growth, respectively.

explore whether inflation and output growth expectations are positively correlated (for example, in the case of forecasts four quarters ahead, we examine how Δp^e_{t+4} and Δy^e_{t+4} are related). The relationship between corresponding expectation errors, i.e. between $(\Delta p_{t+4} - \Delta p^e_{t+4})$ and $(\Delta y_{t+4} - \Delta y^e_{t+4})$, is also examined. Then, we estimate alternative specifications of the Phillips curve relationship. For comparison, we estimate Phillips curves also for the US using the CF survey (see Table 4). In this case expectations variables are constructed via Equation [1]. Our analytical framework is quite straightforward mainly because we have just two variables (for a more sophisticated approach in a multivariate setting, see Sinclair et al., 2012; and Komunjer and Owyang, 2012).

The first six rows in Table 3 indicate that inflation and output growth expectations and the corresponding expectation errors are indeed positively correlated in the euro area. Only if we consider long run (five-year) expectations, does the relationship seem to vanish. Thus,

forecasters seem to believe that in the future rising prices will be related to increasing real activity. The finding is consistent with the basic features of the data (see Figure 3).⁵

Estimation results for the US with the Consensus Forecast data (and a much longer sample period) in Table 4 confirm that inflation and output forecasts are indeed positively correlated in the US, although allowing for both cross-section and period time effects produces rather low t-ratios for the respective coefficients. When the cross-section fixed effects are eliminated, the t-values increase substantially, to 3.42 for Equation [1] (first row) and 7.41 for Equation [2] in Table 4. This finding is consistent with evidence on Dutch households (see Christensen et al., 2006), although the relationship seems to be much stronger in the Dutch data. Recent cross-country evidence with Consensus Economics micro data on professional economists (Fendel et al., 2011) also point in the same direction. The interesting point for this study is the strong support found for a nonlinear Phillips curve.

Next, we use relationship [3] to estimate the (New Keynesian) Phillips curve. We use the survey values to examine whether the expected values reflect this basic relationship in the same way as the actual data (see e.g. Kortelainen et al., 2011a):

$$\Delta p_{it,T}^e = \alpha \Delta p_{it-1,T}^e + \beta \Delta p_{it,T+1}^e + \gamma \Delta y_{it,T}^e + \Sigma \text{Seas}_t + u_{it} \quad [3]$$

where $\Delta y_{it,T}^e$ denotes the expected growth rate of output for the current calendar year expected in period t by forecaster i . T denotes the period to which the forecast applies. “Seas _{i} ” with $i = 1, \dots, 4$ is a seasonal dummy for quarter i . The equation was also estimated using fixed 4 quarters ahead expectations.⁶ Equation [3] was estimated using both the ECB SPF and Consensus Forecast data (see the last seven rows in Table 3, and Table 4). The dependent variable is the forecast of current-period inflation. Hence it is in a sense the equivalent of real-time estimate of the current rate of inflation.

Table 4. Relationship between US inflation and output growth forecasts in the CF survey

Dep. var.	Const.	$\Delta y_{t,T}^e$	$\Delta y_{t,T+1}^e$	$\Delta p_{t,T+1}^e$	$\Delta p_{t-1,T}^e$	FE	Seas	R ²	SEE	X ² test
$\Delta p_{t,T}^e$	2.617 (57.12)	0.032 (1.74)				x		0.963	0.222	784.4 (0.00)
$\Delta p_{t,T+1}^e$	2.601 (65.16)		0.038 (2.25)			x		0.828	0.366	2 495 (0.00)
$\Delta p_{t,T}^e$		0.149 (12.71)		0.855 (51.08)			x	0.464	0.835	599.17 ^a (0.00)
$\Delta p_{t,T}^e$		0.053 (7.63)		0.138 (8.86)	0.851 (63.51)		x	0.828	0.389	130.08 ^a (0.001)
$\Delta p_{t,T}^e$		0.043 (6.96)		0.283 (21.77)	0.684 (102.0)		x	0.824	0.478	161.46 ^a (0.00)
$\Delta p_{t,T}^e$		0.062 (11.02)		0.264 (27.43)	0.684 (NA)			0.815	0.492	320.60 ^a (0.00)
$\Delta y_{t,T+1}^e$	1.597 (16.09)	0.430 (10.91)				x		0.792	0.400	1 200 (0.00)

Notes: Notation is the same as in Table 2. Superscript “a” denotes the case where the alternative is a model with forecaster-specific coefficients of the output growth variable. For equations in the fifth and sixth rows, actual past inflation is used instead of lagged value of expected current period inflation. In the latter equation, the coefficient of lagged inflation is allowed to be forecaster-specific. No fixed-effects are used with the Phillips curves. “FE” indicates that fixed cross-section effects and “seas” that seasonal dummies are included in the equation. X² test denotes the Chi-square test of equality of individual coefficients, or the test for fixed effects.

As for the Phillips curve, we find that a conventional hybrid specification performs strikingly well: the coefficients are correctly signed and of reasonable magnitude. This is not usually the case when a rational expectations model is estimated by GMM using lagged

values of inflation and output as instruments.⁷ The sum of coefficients of future and past inflation comes quite close to one in hybrid specification. The coefficients represent a not-so-new 50-50 split in weights between past and future inflation.⁸

If inflation expectations only partially reflect output growth expectations, we should consider other sources of differences in inflation expectations. The most obvious option is a difference in the Phillips curve parameters. Thus, we estimate Equation [2] with equal slopes for all forecasters and, alternatively, allowing for different (forecaster-specific) slopes. The corresponding test results for the hypothesis of equal coefficients for all forecasters are reported in the last column of Table 4. We see that the null hypothesis of equal coefficients is rejected, so that the slopes are indeed different (although only “marginally” in the hybrid specification), suggesting that the “forecasting model” may indeed also produce differences in inflation forecasts. Thus, differences in inflation forecasts may not simply reflect differences in the optimism-pessimism balance.

Finally, note that the estimates of the Phillips curve (rows 3-6 in Table 4) generally make sense; to some extent the results make more sense than those for the actual data as regards the imposition REH orthogonality conditions via the GMM estimator (see e.g. Adam and Padula, 2011). Thus, again the coefficient of output is positive and the coefficients for both the forward- and backward-looking inflation terms are of reasonable magnitude and the coefficients can be estimated quite precisely. This time, the coefficient of the lagged inflation term is much larger than that of the forward-looking term, probably reflecting the way in which the forecast value is constructed (i.e. using Equation [1]).

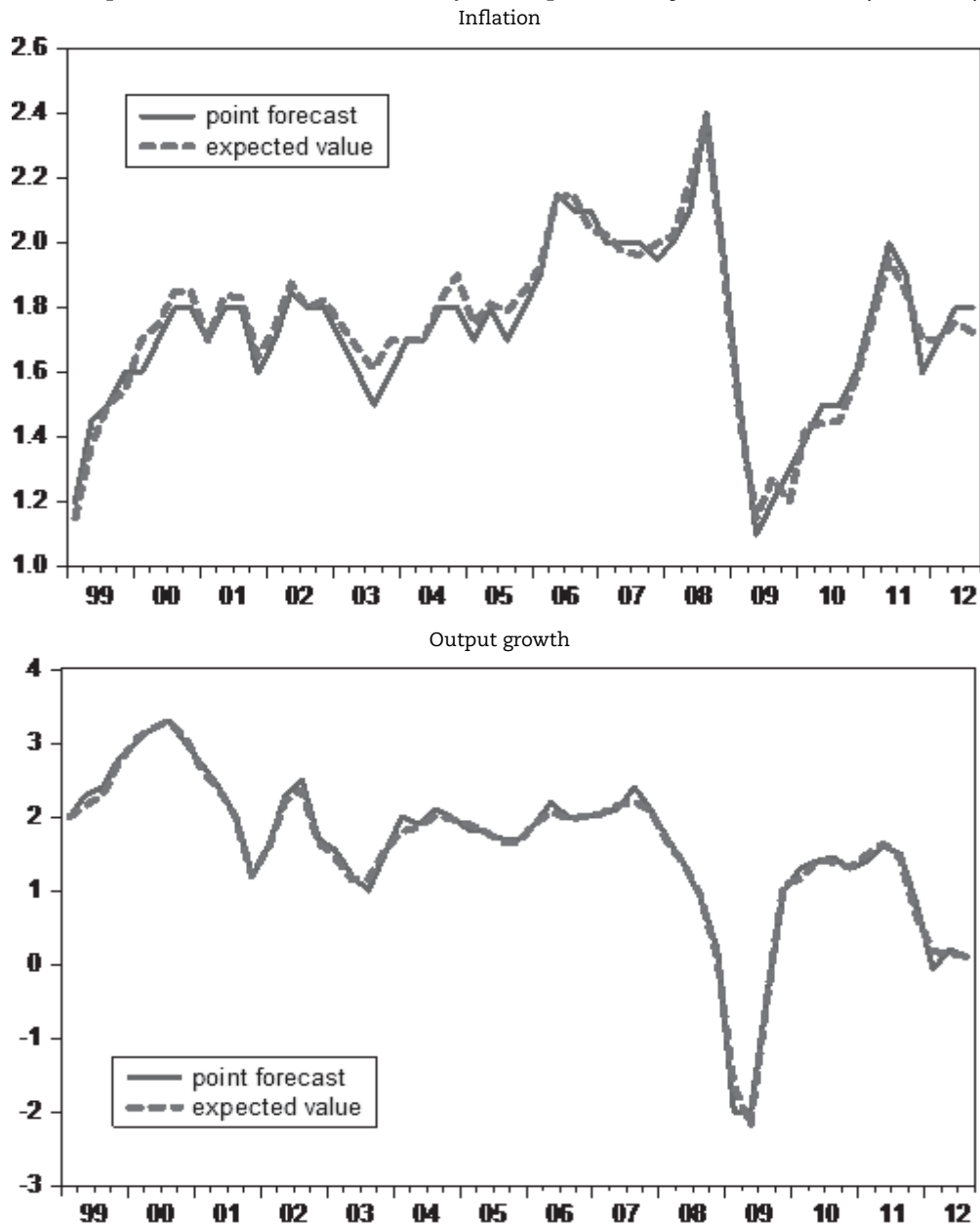
4. Analysis of forecast uncertainty

Finally, we turn to the analysis of forecast uncertainty and disagreement. The recent financial crisis clearly highlighted the fact that mean values of survey forecasts do not necessarily reveal all relevant information about forecasters' expectations. An analysis of forecast uncertainty may also provide useful information on market participants' behaviour. Forecast uncertainty and disagreement in the ECB SPF survey has been analysed in some recent studies, but not in terms of view of internal consistency of different variables (see e.g. Bowles et al., 2010; and Conflitti, 2011). Uncertainty and disagreement in the Bank of England Survey of External Forecasters has been widely analysed in Boero et al. (2008).

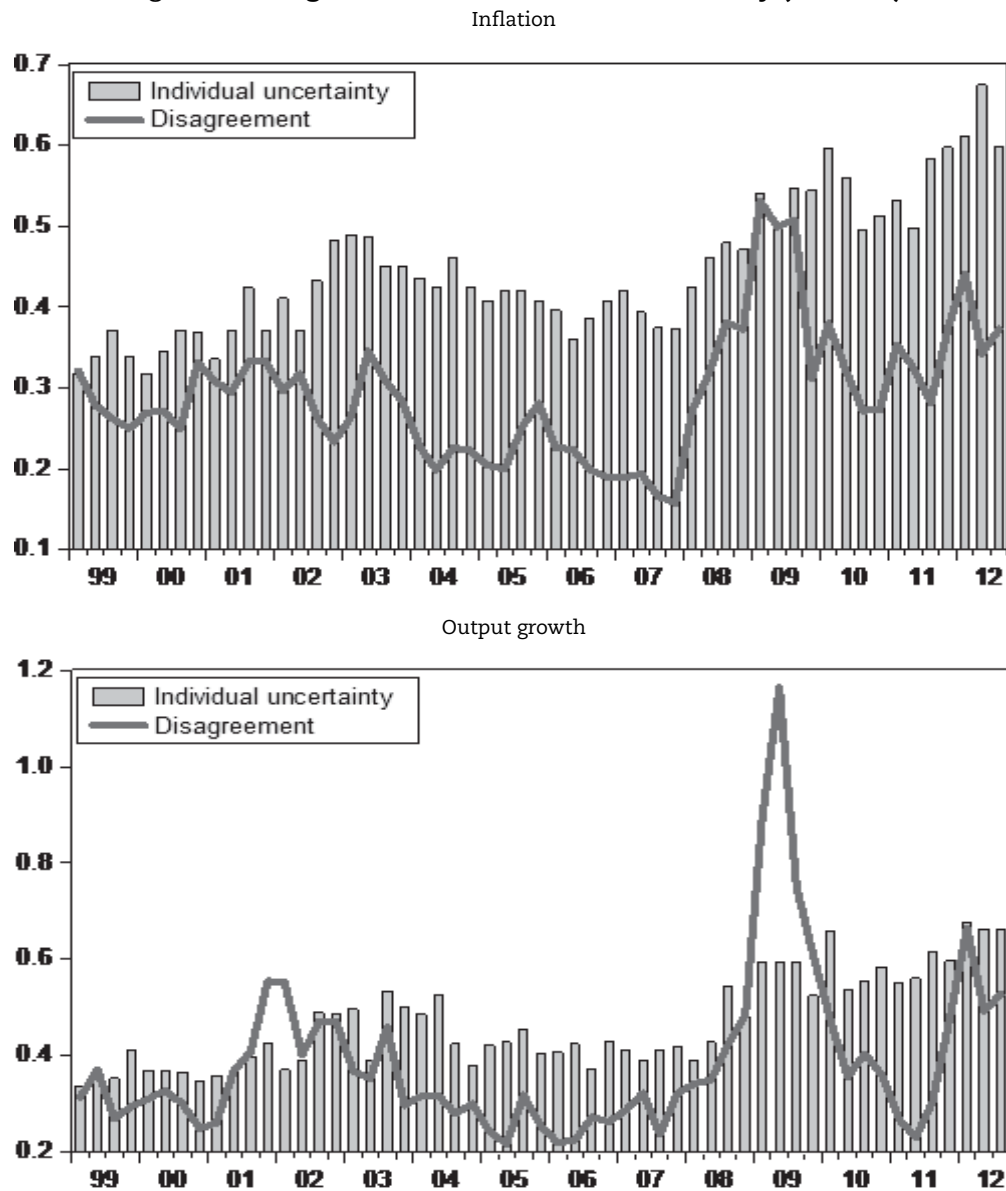
The ECB SPF survey includes both individual point estimates and individual probability distributions (essentially for all forecast horizons). In Figure 4, we compare averages of individual point forecasts with corresponding expectations based on subjective probability distributions. Both inflation and real GDP growth forecasts for four quarters ahead are displayed. Figure 4 shows that the alternative forecast series are so closely related as to be almost identical, which suggests that forecasters use the point estimates as benchmarks for forming their subjective probability distributions.

Individual forecast uncertainty is investigated using two alternative measures: 1) the mean value of individual forecasters' subjective uncertainty, measured by the second moment of the distribution of forecast values; and 2) the conventional standard deviation of point forecasts, which measures disagreement between individual forecasters. Figure 5 represents these two measures for both inflation and output growth expectations four quarters ahead in the ECB SPF survey. The same information in a slightly different form and corresponding series for the US SPF survey are shown in Figure 7.

Figure 4. **Comparison of mean values of individual point forecasts and expected values based on subjective probability distributions (ECB SPF)**



Inflation disagreement and output growth disagreement are closely related (the correlation coefficient between these variables is 0.77, see Table 5). Moreover, they seem to be relatively closely related to forecast errors (absolute deviations between actual and expected values). By contrast, disagreement measures are only weakly (albeit statistically significantly) correlated with subjective uncertainty measures. Figure 5 provides at least a partial explanation. Forecast disagreement seems to have increased notably in response to the financial crisis.⁹ The impact of the crisis on disagreement as to inflation expectations seems to be permanent while the impact on output growth disagreement has been more temporary. This is also the case for the “individual uncertainty” measure: the higher-than-normal levels of individual inflation uncertainty and output growth uncertainty have

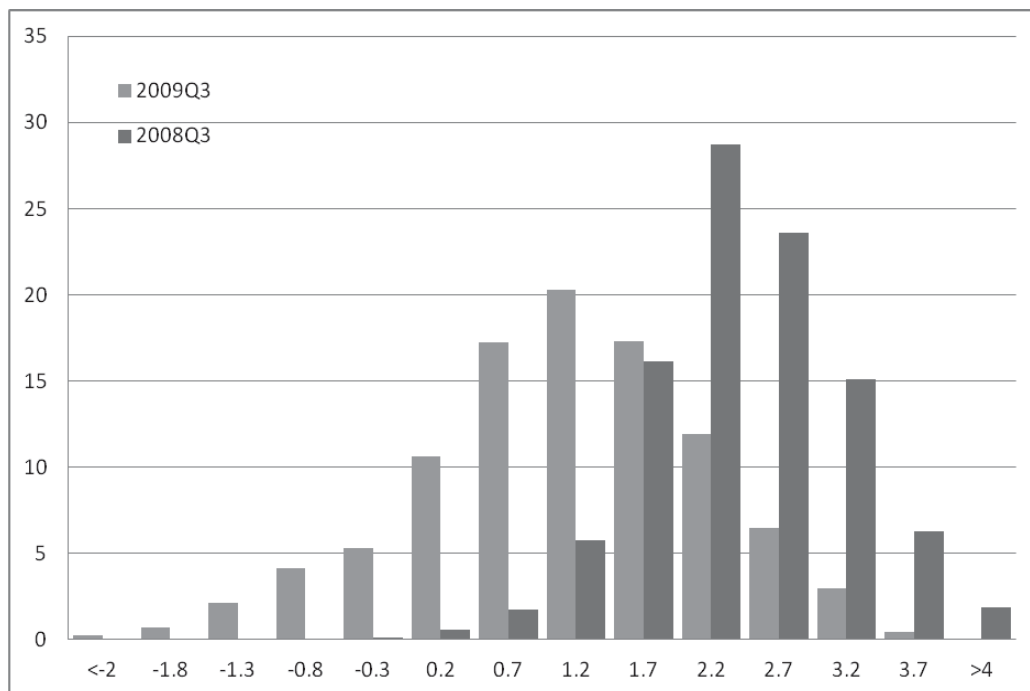
Figure 5. **Disagreement vs. individual uncertainty (ECB SPF)**

persisted since the peak of the financial crisis. It is worth noting that, compared to the pre-crisis period, individual point forecasts were highly divergent after mid-2008: in the middle of the crisis, both inflation and deflation were anticipated and similarly both positive and negative output growth values were forecasted. Therefore, increasing disagreement across forecasters may simply reflect increased polarisation of individual views, without any substantial change in individual uncertainties. Individual forecasters may still be very confident in their views even though they disagree with each other more than before. This “camp view” approach is introduced and tested e.g. by Geiger et al. (2009) and Andrade et al. (2012). Their results suggest that this possibility must be taken seriously in interpreting the uncertainty results.

Table 5. **Correlation coefficients between uncertainty measures in the ECB SPF survey**

	$\delta_{g,D}$	$\delta_{p,D}$	$\delta_{g,S}$	$\delta_{p,S}$	$ \Delta y - \Delta y^e $	$ \Delta p - \Delta p^e $
$\delta_{g,D}$	1.00	0.77	0.19	0.18	0.70	0.42
$\delta_{p,D}$	0.77	1.00	0.19	0.21	0.57	0.50
$\delta_{g,S}$	0.19	0.19	1.00	0.81	0.10	0.14
$\delta_{p,S}$	0.18	0.21	0.81	1.00	0.10	0.16
$ \Delta y - \Delta y^e $	0.70	0.57	0.10	0.10	1.00	0.26
$ \Delta p - \Delta p^e $	0.42	0.50	0.14	0.16	0.26	1.00

Notes: $\delta_{g,D}$ denotes the standard deviation of point estimates of output growth forecasts and $\delta_{g,S}$ the standard deviation of individual forecasters' forecast outcomes. $|\Delta y - \Delta y^e|$ refers to absolute forecast errors in terms of output growth. Subscript p refers to inflation forecasts. 5% significance level for correlations is 0.058.

Figure 6. **Average values of perceived probabilities of inflation expectations (ECB SPF)**

The “camp view” might also explain – at least partially – the puzzling features in Figure 5, in particular that the two uncertainty measures are only weakly related (and that the subjective uncertainty measures are only weakly related to the forecast errors; see Table 5).¹⁰ Thus, one is well advised to interpret uncertainty developments in the aftermath of the crisis with caution and perhaps not to jump to the conclusion that the survey data tell us that inflation expectations have been firmly anchored in recent years.¹¹

To get more insight into the nature of the change in subjective uncertainty we use the ECB SPF data to compare average probability distributions of individual forecasters for inflation expectations (computed four quarters ahead). We explore just two periods during the crisis: 2008Q3 and 2009Q3 (see Figure 6). The comparison indicates that the distributions are indeed quite different for the two periods. However, they are different largely because of the difference in means. A level shift in inflation expectations

from 2008Q3 to 2009Q3 can be easily discerned, but the change in the dispersion is not equally obvious. Between the two periods the mean of variances computed from individual PDFs decreased from 12.2 to 9.6, but that mainly reflects the disappearance of the highest values for inflation expectations. Thus, it is questionable to claim that a genuine change in uncertainty (measured in this way) had taken place.¹²

Finally, we consider the GDP price index –based measures of inflation uncertainty in the US SPF survey since 1992Q1, shown in Figure 7.¹³ The first panel illustrates mean expected inflation and corresponding minimum and maximum values for each period. Disagreement, i.e. standard deviation of point forecasts for four quarters ahead, is reported in the second panel. The graph of the series is quite reasonable during the crisis although the peak in the series is very short-lived and differs greatly from its European counterpart. Because the data seem to include several outliers, we focus not only on the standard deviation but also on the interquartile range (the difference between the 75th and 25th percentiles of the projections for Q/Q growth of the GDP price index, expressed in annualised percentage points). The corresponding series seem reasonable. However, we find relatively little systematic time-variation, so that this uncertainty measure is not very informative either.

The subjective uncertainty measure in Figure 7 is based on individual probabilities that in turn are related to the expected annual-average over annual-average percent change in the GDP price index.¹⁴ The individual uncertainty has been surprisingly stable in the whole sample period. This can be partly explained by the chosen probability ranges in the survey (0 to 0.9, 1 to 1.9, 2 to 2.9, etc.). Since these ranges are relatively wide compared to the actual inflation history, there are typically only a few non-zero ranges in the survey responses. The time-variation in the response rate (last panel) may also explain why the subjective uncertainty measure for the US SPF survey is not very informative. The average number of non-zero entries per respondent is only 3.5, although the crisis increased the ratio somewhat (Figure 7). In the ECB SPF survey, things are somewhat better. With inflation, there has been 4.7 and with output growth 5.1 non-zero entries per respondent on average. However, this may simply reflect the clearly narrower probability ranges in the ECB questionnaire.

Examination of survey response patterns and survey response rates suggests that the higher-order moments of the data may not be very informative in assessing forecast uncertainty, even though there seem to be some indication of non-normal distributions. Take for instance the behaviour of maximum and minimum values *vis-à-vis* the expected inflation in the upper panel of Figure 7. The observed asymmetry may not reflect true changes in forecasting behaviour. The number of respondents is so small and the number of outliers so large that even the standard deviations may be on relatively shaky ground. That is a pity, because we would like to use the information on uncertainty proxies in evaluating the general public's perceptions on monetary policy.

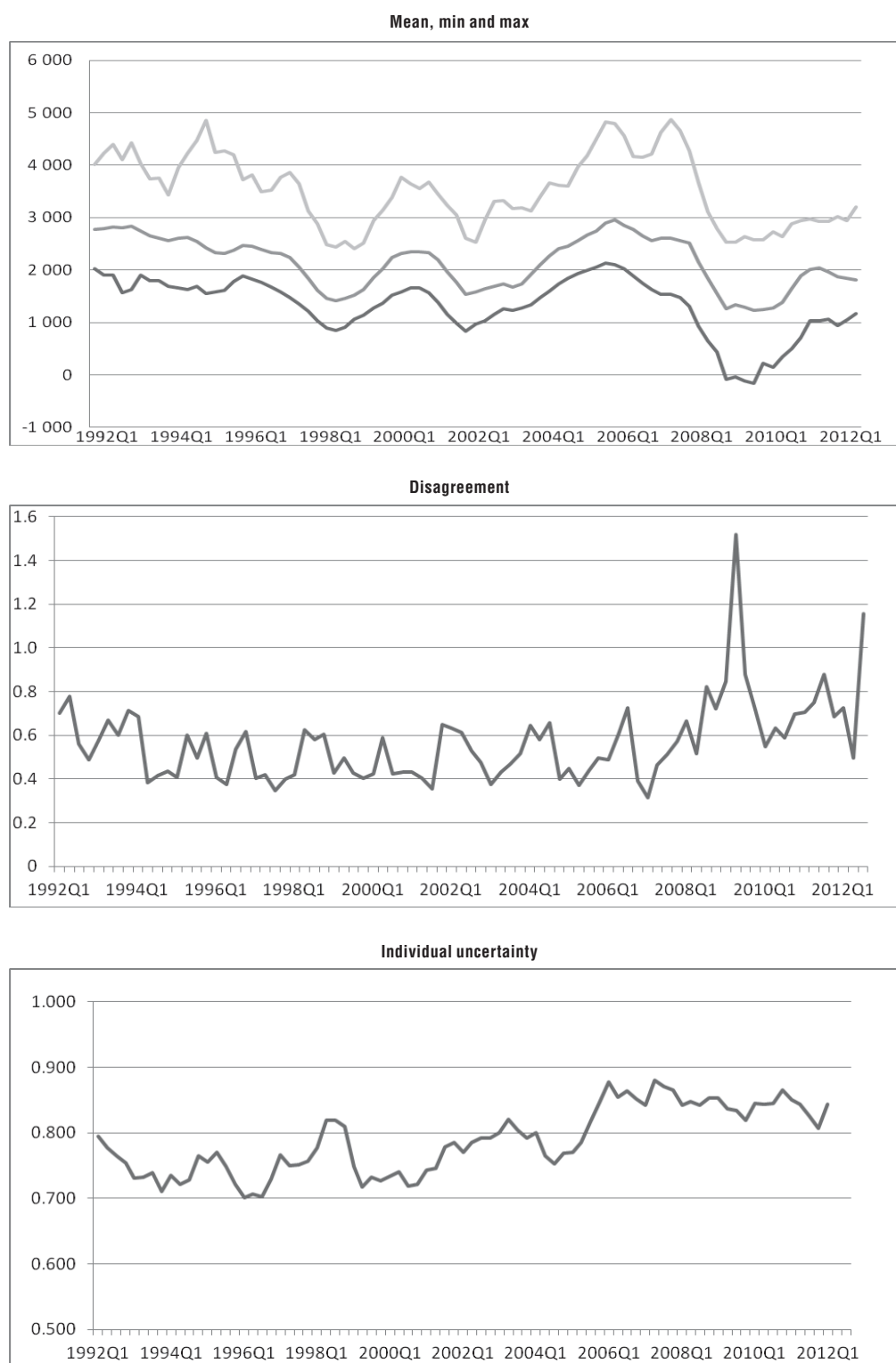
Figure 7. **Different measures of inflation uncertainty in the US SPF survey**

Figure 7. **Different measures of inflation uncertainty in the US SPF survey (cont.)**

Notes: The series “entries/respondent” are derived by dividing the total number of entries (categories in the distribution with non-zero values) by the number of forecasters. 75%-25 % is the Philadelphia FED measure of dispersion: the difference between the 75th percentile and 25th percentile (interquartile range) of the forecasts. It corresponds to our measure of disagreement, the cross-section standard deviation of point forecasts.

5. Concluding remarks

This paper has examined individual inflation and real GDP growth expectations using three alternative panel data sets. The analyses produced several interesting results. First of all, the results show that individual forecasters seem to deviate systematically from each other. Differences between individual forecasts are large and persistent, and individual views are more divergent for shorter forecast horizons. The results also indicate that basing forecasts on weighted averages of fixed event forecasts is not necessarily the optimal way to approximate fixed horizon forecasts.

The forecasters produce values that are largely consistent with basic principles of economics: inflation and output growth expectations are positively and significantly correlated and, moreover, consistent with the hybrid specification of the New Keynesian Phillips curve. This result is common for all data sets and data transformations. The slope

parameters in the Phillips curve relationship seem not to be identical across individual forecasters, however, indicating that there is considerable amount of persistent heterogeneity across all forecasts. Finally, we find that individual inflation uncertainty based on subjective probability distributions is quite closely related to corresponding individual output uncertainty. However, the relationship between these uncertainty measures with corresponding standard deviations of individual point forecasts is relatively weak. To some extent, this can be explained by changes in the number of respondents and changes in probability ranges in the questionnaires. Also outliers play some role in this respect. Anyway, the respective uncertainties are positively (albeit more weakly) related.

Looking ahead, it would be very useful to explore micro data sets more extensively. One would then require more extensive data sets – more respondents and hopefully more detailed information in the questionnaire. Better information would facilitate more detailed examination of distributions and thus provide a more accurate measure of underlying forecasting uncertainty. Thus, for instance the so-called “twin peaks” hypothesis for different sets of forecasters could be tested. Additional data would also allow testing hypotheses that predict changes in inflation expectations on the basis of higher-order moments of the probability distribution of forecast values.

Notes

1. Kenny et al. (2012) provide some evidence of the usefulness of these data by showing that the distributional information helps to predict future inflation and output developments.
2. Schnader and Stekler (1991) and Kolb and Stekler (1996) in fact show that it is not clear whether we can simply use central moments, as we usually do, to characterise “consensus forecast”.
3. Alternative data sets are summarised in Annex A.
4. The ECB SPF survey is described in detail in Bowles et al. (2007).
5. With the long time-horizon data we have a degrees-of-freedom problem (see Table 2) in that the results are not fully comparable with the shorter-time horizon data.
6. We have no monthly data on actual GDP growth, which would be needed to estimate the Phillips curve. Hence, we have to use some sort of “real time” proxy for current output. That is done by computing a 12-month moving average of expectations for current-year output growth (results reported in rows 3-6 of Table 4).
7. With the ECB SPF data we have a sort of seasonality problem due to the fact that the current values of inflation and output growth are expressed only in terms current calendar year.
8. Interpreting the results is problematic, because the forecasts do not appear to be unbiased. That is true both for inflation and output growth forecasts and for unbalanced and balanced panels. With output growth, the bias ($\Delta y_t^e - \Delta y_t$) is relatively small (0.19), but with inflation ($\Delta p_t^e - \Delta p_t$) it is more troublesome (-0.37).
9. This finding is consistent with Döpke and Fritsche (2006), who analysed forecast dispersion of German professional forecasts for 1970-2004. They find that forecast dispersion varies over time and is particularly high before and during recessions.
10. Unfortunately, we face a measurement problem here, because the ECB questionnaire was changed in 2008 to allow more negative values for inflation and output growth, which obviously affects the computed standard deviation. Careful scrutiny of the data however reveals that the new extreme values were used very often, so that the overall importance of the change is not devastating.
11. Typically, long-term forecast uncertainty is measured by standard deviations. Unfortunately, the sample size for long-term forecasts is so small as render problematic any proper statistical analysis.
12. The puzzling features of Figure 6 probably reflect the fact that the probability distributions of the questionnaire changed during the crisis. As a result of deteriorating growth and sharply decreasing

inflation, the scales in the questionnaire were changed substantially (more “extreme” alternatives were added).

13. Survey information is not fully comparable before and after 1992Q1 due to significant changes in the questionnaire. Here we focus only on inflation, but we would mention that in terms of uncertainty relationships inflation and output growth data follow the same pattern as in Europe. Thus, for instance, the correlation with the respective interquartile ranges was clearly positive, having a value of 0.38.
14. Due to changing forecast horizons in annual-average over annual-average series, we report here the moving averages.

References

- Adam, K., and M. Padula (2011), “Inflation Dynamics and Subjective Expectations in the United States”, *Economic Inquiry*, 49 (1), 13-25.
- Andrade, P., E. Ghysels and J. Idier (2012), “Tails of Inflation Forecasts and Tales of Monetary Policy”, unpublished mimeo, Bank of France.
- Boero, G., J. Smith and K.F. Wallis (2008), “Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters”, *The Economic Journal*, 118 (July), 1107-1127.
- Bowles, C., R. Friz, V. Genre, G. Kenny, A. Meyler and T. Rautanen (2007), “The ECB Survey of Professional Forecasters: A Review after Eight Years’ Experience”, *ECB Occasional Paper*, No. 58.
- Bowles, C., R. Friz, V. Genre, G. Kenny, A. Meyler and T. Rautanen (2010), “An Evaluation of the Growth and Unemployment Forecasts in the ECB Survey of Professional Forecasters”, *OECD Journal: Journal of Business Cycle Measurement and Analysis*, Vol. 2.
- Christensen, C., P. van Els and M. van Rooij (2006), “Dutch households’ perceptions of economic growth and inflation”, *De Economist*, Springer, Vol. 154(2), pp. 277-294, June.
- Conflitti, C. (2011), “Measuring Uncertainty and Disagreement in the European Survey of Professional Forecasters”, *OECD Journal: Journal of Business Cycle Measurement and Analysis*, Vol. 2.
- Dovern, J., U. Fritsch and J. Slacalek (2012), “Disagreement Among Forecasters in G7 Countries”, *The Review of Economics and Statistics* 94(4), 1081-1096.
- Döpke, J. and U. Fritsche (2006), “When Do Forecasters Disagree? An Assessment of German Growth and Inflation Forecast Dispersion”, *International Journal of Forecasting*, 22, 125-135.
- Fendel, R., E. Lis and J.C. Rülke (2011), “Do Professional Forecasters Believe in the Phillips Curve? Evidence from G7 Countries”, *Journal of Forecasting*, 30, 268-287.
- Fendel, R. and J.C. Rülke (2012), “Are Heterogeneous FOMC Forecasts Consistent with the Fed’s Monetary Policy?”, *Economics Letters*, 116, 5-7.
- Geiger, F. O. Sauter and K. Schmid (2009), “The Camp View of Inflation Forecasts”, *Hochheimer Diskussionsbeiträge*, Universität Hohenheim, 320.
- Gerlach, S. (2007), “Interest Rate Setting by the ECB, 1999-2006: Words and Deeds”, *International Journal of Central Banking*, September, 1-42, 389-439.
- Kenny, G., T. Kostka and F. Masera (2012), “How Informative Are the Subjective Density Forecasts of Macroeconomists?”, *ECB Working Paper*, 1446.
- Kolb, R. and H. Stekler (1996), “Is there a Consensus Among Financial Forecasters?”, *International Journal of Forecasting*, No. 12, 455-464.
- Komunjer, I. and M. Owyang (2012), “Multivariate Forecast Evaluation and Rationality Testing”, *The Review of Economics and Statistics*, No. 94, 1066-1080.
- Kortelainen, M., M. Paloviita and M. Viren (2011a), “Observed Inflation Forecasts and the New Keynesian Macro Model”, *Economics Letters*, 112 (1), 88-90.
- Kortelainen, M., M. Paloviita and M. Viren (2011b), “Expectations and Inflation: Evidence from Euro Area, Japan and USA”, *International Economics*, 64 (4), 465-487.
- Pesaran, M.H. and M. Weale (2006), “Survey Expectations”, in G. Elliot, C.W.J. Granger and A. Timmermann (eds.), *Handbook in Economic Forecasting*, Elsevier.
- Sinclair, P. (2010), “Inflation Expectations”, Routledge International Studies in Money and Bankin, London.

- Sinclair, T, H. Stekler and W. Carnow (2012), "A New Approach for Evaluating Economic Forecasts", *Economics Bulletin*, No. 32, 2332-2342.
- Schnader, H. and H. Stekler, (1991), "Do Consensus Forecasts Exist?", *International Journal of Forecasting*, No. 7, 165-170.
- Tillmann, P. (2010), "The Fed's Perceived Phillips Curve: Evidence from Individual FOMC Forecasts", *Journal of Macroeconomics*, No. 32, 1008-1013.

ANNEX A

*Dataset details***ECB Survey of Professional Forecasters**

Data source	European Central Bank, www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.htm .
Frequency	Quarterly.
Period	1999Q1-2012Q1.
Actual data	HICP and GDP growth rates over past 4 quarters from OECD.
Forecast variables for the euro area	HICP and real GDP forecasts for fixed calendar years ($T = t + 0$, $T = t + 1$, $T = t + 2$ or $T = t + L$, where L refers to longterm) and for four, eight and 20 quarters ahead.

US SPF Survey of Professional Forecasts

Data source	Federal Reserve Bank of Philadelphia, www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/
Frequency	Quarterly.
Period	1992Q1-2012Q1.
Actual data	Growth rates of GDP price deflator over past 4 quarters from OECD.
Forecast variable for the US	Moving average of annual-average over annual-average of the GDP price deflator forecasts using Equation [1].

Consensus Forecasts

Data source	Consensus Economics, www.consensuseconomics.com/con1/index.htm
Frequency	Monthly.
Period	1988M1-2011M4.
Actual data	Growth rates of CPI and GDP over past 4 quarters from OECD.
Forecast variables for the US	CPI and real GDP forecasts based on Equation [1].

On the robustness of balance statistics with respect to nonresponse

by
Christian Seiler*

A general problem for survey conductors is the fact that the response decision can be connected to the intended answer of the non-respondents. This nonresponse bias might have a substantial effect on the aggregated results. In this paper, a participation framework for the widely used business cycle balance statistics indicators is examined. An extensive simulation study is performed to analyse their effects. The analyses show that these indicators are extremely stable towards nonresponse biases.

Keywords: balance statistics, business tendency survey, nonresponse bias, simulation study

JEL classification: C810, C830

* Ifo Institute, Poschingerstr. 5, 81679 München.

1. Introduction

Monitoring and forecasting of economic activity is nowadays high on the agenda of both public and private institutions. As official data are commonly released with a delay, timely business cycle indicators are needed. Balance statistics indicators based on survey data play a central role in this context. Their major advantages are timely availability and the fact that they are subject to almost no revisions. Indeed, aggregated business survey data has proved to be one of the most competitive indicators for analysing macroeconomic variables, for example, the Consumer Confidence Index in the United States (Ang, Bekeart and Wei, 2007), the Economic Sentiment Indicator for the European Union (Gayer, 2005) or the Ifo Business Climate Index for Germany (Drechsel and Scheufele, 2010; Kholodilin and Siliverstovs, 2006; Robinsonov and Wohlrabe, 2010).

Most papers using survey based business cycle indicators (usually not explicitly) assume that the indicator is measured without any nonresponse bias. While a large body of literature exists concerning the effects of nonresponse in household or individual surveys, less is known about the processes and reasons for participation and responding behaviour in business surveys (Janik and Kohaut, 2012; Thompson, 2009; Willimack and Nichols, 2010), particularly in the case of surveys aimed at evaluating the business cycle (Wang, 2004). Therefore, the indicators cannot generally be said to be unbiased. Seiler (2010) gives evidence for a dependence of the responding behaviour in the *Ifo Business Survey* from the business cycle which confirmed the results of Harris-Kojetin and Tucker (1999) for population surveys. Seiler and Heumann (2013) estimated the missing observations in the *Ifo Business Survey* with several imputation methods and found substantial differences given the (estimated) answers of the company. So, the question arises how a potential nonresponse bias affects the results of these type of business surveys. How do indicators look when they are biased by systematic nonresponse patterns? How strongly is the performance reduced? In this paper, we can show that these macro level indicators are extremely stable even for very strong bias structures in the data.

The paper is organised as follows: in Section 2, we define the methodologic framework for the calculation of the balance statistics and show how nonresponse biases affect the indicators. We subsequently introduce measurements to explore the magnitude of the biases. In Section 3, we perform an extensive simulation study for a wide variety of nonresponse patterns and different types of business cycles. Finally, Section 4 sums up the results.

2. Methodological framework

In market economies, one can typically observe that general long-term growth is accompanied by temporary fluctuations. In general, the classical business cycle is divided into four phases: upswing, boom, downswing and recession. In its ideal form, the cycle would have a sinusoidal wave. The business cycle itself is a theoretical construct which can not be measured directly, but can be visualised by observing economic indicators, such as

GDP growth rates or production in the manufacturing sector. We can think of a business cycle $g(t)$ as a stationary function in time, for example, $g(t) = \sin(t)$. Due to the fact that official data are published with a delay (and often revised after being first released too), business cycle tendency surveys can monitor the recent economic situation considerably more quickly. In line with the Joint Harmonised EU Programme of Business and Consumer Surveys (European Union, 2006), the indicators base on two variables (business situation and business expectations) which are measured on a 3-level scale representing a good, equal or bad state. Due to the construction of the questions in the questionnaire, the indicators obtained actually measure the business cycle without trend, see OECD (2003). For more theoretical aspects of the balance statistics see Anderson (1951 and 1952) and Theil (1952).

For the current study, we abstract from any formal definition of the business cycle. We define a macroeconomic time series for which a qualitative survey with a 3-level scale is constructed or referenced to. Usually, a variable on a 3-level scale is influenced by an unobserved process (for example, the business cycle in our case). However, continuous variables may also be surveyed on a 3-level scale, realisations of which are only available on a very delayed basis, for example, the *World Economic Survey* of the Ifo Institute asks for inflation in the appropriate country (Stangl, 2007). In this section, we leave the exact shape of $g(t)$ undefined, but show the effects for different types of $g(t)$ in Section 3.

2.1. Construction of the balance statistics

All surveys asking on a 3-level scale mentioned above calculate a so-called *balance statistics* after data collection. The balance statistics is defined as the fraction of positive answers subtracted by the fraction of negative answers for a certain variable. We assume that respondent i is affected in his opinion formation by the cycle function $g(t)$, for which the survey is conducted, and an individual error term ε_i :

$$s_{i,t}^* = g(t) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),$$

$$i = 1, \dots, n, t = 1, \dots, T \text{ with}$$

$$E_t(s_{i,t}^*) = g(t), \quad \forall i,$$

where $E_t(\cdot)$ denotes the expected value at time t . As each survey participant i is restricted to give answers on a 3-level scale, we observe

$$s_{i,t} = \begin{cases} + & \text{if } s_{i,t}^* > \tau^+ \\ = & \text{if } \tau^+ \geq s_{i,t}^* > \tau^- \\ - & \text{if } \tau^- \geq s_{i,t}^* \end{cases}$$

where τ^s define time-independent thresholds.¹ This is the standard model for latent variable models (Bartholomew and Knott, 1999). The assumption of time- and individual-independent thresholds as well as time- and individual-independent normality of ε_i is required for the calculation of the balance statistics. In addition, τ^s are assumed to be symmetric around zero, that is, $-\tau^- = \tau^+$ (Henzel and Wollmershäuser, 2005). Therefore, the balance statistics is a special case according to the Carlson-Parkin framework (Carlson and Parkin, 1975) which allows for differences for thresholds and error terms in time and across individuals. These very strong assumptions have been criticised by several papers, see Nardo (2003) for an overview. As the standard calculation of the balance statistics requires the conditions mentioned above, we proceed assuming τ^s and $\varepsilon_i = \varepsilon$ as independent from time and individuals. In Annex C, we discuss the impact of these assumptions on our analysis.

To enable the calculation of means, we define $' + ' \equiv 1$, $' = ' \equiv 0$ and $' - ' \equiv -1$. Then, the expected mean of $s_{i,t}$ for given t is defined as $E_t(s_{i,t}) = \sum_{s=-1}^1 s \cdot P(s_{i,t} = s)$. With $E_t(s) = E_t(s_{i,t})$, we follow:

$$E_t(s) = E_t(s_{i,t}) = \phi^+(t) - \Phi^-(t) \quad [1]$$

with $\phi(\cdot)$ as the cumulative distribution function of the standard normal distribution.² $\Phi^+(t) := 1 - \Phi[(\tau^+ - g(t))/\sigma]$ denotes the probability for a positive reply whereas $\Phi^-(t) := \Phi[(\tau^- - g(t))/\sigma]$ the probability for a negative reply. $E_t(s) = E_t(s_{i,t})$ in Equation [1] is given since $\left[1 - \Phi\left(\frac{\tau^+ - g(t)}{\sigma}\right)\right] - \Phi\left(\frac{\tau^- - g(t)}{\sigma}\right)$ includes no individual information i . In the following, we call $E_t(s)$ the unbiased mean function or the unbiased indicator. In Annex A, we show that Equation [1] is the same as the balance statistics for a 3-level scale.

2.2. Inclusion of response decision

The mean function $E_t(s)$ in Equation [1] is the expected function in time for $n \rightarrow \infty$ when no nonresponse bias is present. To include the decision of the respondent to participate in the survey, we define a binary variable m which indicates this decision, so that $m_{i,t} = 1$ if the observation i is missing at time t and $m_{i,t} = 0$ if not. According to Little and Rubin (2002), not missing at random (NMAR), that is, a nonresponse bias, occurs if the probability for missing $P(m_{i,t} = 1)$ depends on the outcome of the variable of interest s , that is, $P(m_{i,t} = 1 | s_{i,t} = s, \xi)$, with ξ as additional parameters which are unrelated to the variable of interest (including t in our case). Therefore, the missing process would depend on the state s . To evaluate the effects of the bias patterns on the balance statistics, we assume that each respondent has the same probability not to respond to the survey for given s (and t), that is, $P(m_{i,t} = 1 | s, \xi) = P(m = 1 | s)$, $\forall i$. Then, $\pi^s(t) = P(m_{i,t} = 0 | s_{i,t} = s)$ is the (not necessarily time-dependent) probability to be observed at time being in state s . Therefore, the mean function of the observed units $s_{i,t}^{obs}$ is given by:

$$E_t(s^{obs}) = E_t(s_{i,t}^{obs}) = \sum_{s=-1}^1 s \cdot P(s_{i,t} = s | m_{i,t} = 0) \quad [2]$$

$$\begin{aligned} &= \sum_{s=-1}^1 s \cdot \frac{P(m_{i,t} = 0 | s_{i,t} = s) \cdot P(s_{i,t} = s)}{P(m_{i,t} = 0)} \\ &= \sum_{s=-1}^1 s \cdot \frac{P(m_{i,t} = 0 | s_{i,t} = s) \cdot P(s_{i,t} = s)}{\sum_{s=-1}^1 P(m_{i,t} = 0 | s_{i,t} = s) \cdot P(s_{i,t} = s)} \\ &= \frac{\pi^+(t) \cdot [1 - \Phi(\frac{\tau^+ - g(t)}{\sigma})] - \pi^-(t) \cdot \Phi(\frac{\tau^- - g(t)}{\sigma})}{\sum_s \pi^s(t) \cdot \Phi^s(t)} \\ &= \frac{\pi^+(t)}{\bar{\pi}(t)} \cdot \Phi^+(t) - \frac{\pi^-(t)}{\bar{\pi}(t)} \cdot \Phi^-(t) \end{aligned}$$

with $\bar{\pi}(t) := \sum_s \pi^s(t) \cdot \Phi^s(t)$, $\Phi^-(t) := 1 - [\Phi^+(t) + \Phi^-(t)]$ and $0 \leq \pi^s(t) \leq 1$. Note that $\pi^s(t)$ is the probability to answer ("acceptance rate") being in state s at time t and $\Phi^+(t)$ and $\Phi^-(t)$ directly depend on $g(t)$ and beyond only on time-invariant variables. A comparison of Equation [1] and [2] clearly shows that $E_t(s^{obs})$ is a *weighted balance statistics* with weights $\frac{\pi^s(t)}{\bar{\pi}(t)}$, that is, the relation between the acceptance rate $\pi^s(t)$ of state s and the sum $\bar{\pi}(t)$

of all (weighted) acceptance rates at time t . Therefore, NMAR occurs in cases of $\exists t \in \{1, \dots, T\} \wedge \exists (r, s): \pi^r(t) \neq \pi^s(t)$ with $r, s \in S$. Note that the probability $\pi^r(t)$ for the center category enters the mean function $E_t(s^{obs})$ in $\bar{\pi}(t) = \sum_s \pi^s(t) \cdot \Phi^s(t)$. Thus, $E_t(s^{obs})$ is the mean function in t of our observed (and potentially biased) indicator for $n \rightarrow \infty$. Therefore, $E_t(s^{obs})$ is designated as the observed or biased mean function or the observed or biased indicator. In addition, $E_t(s)$ denotes the unbiased case of $E_t(s^{obs})$.

2.3. Correlation with the cycle function

When inspecting nonresponse biases and their effects on the observed indicator $E_t(s^{obs})$, we have to define which type of bias we want to analyse. The first “natural” comparison would be to evaluate the difference between the observed indicator $E_t(s^{obs})$ and the unbiased mean function $E_t(s)$, for example, by (absolute) differences. However, we have two notes on this issue: Firstly, survey indicators that are constructed by balance statistics are, at least, artificial. Unlike official data, for example, their level does not reflect a certain quantity. This particularly holds for latent variables such as the “business situation”. At best, a positive/negative value indicates an increase/decrease. Secondly, these indicators are constructed to display the cyclical development $g(t)$ for different economic parameters over time. A shift of the whole time series $E_t(s)$ or a stretch/compression by a constant factor would leave the relationship between $E_t(s)$ and the cycle function $g(t) = E_t(s^*)$ unaffected. Therefore, the main subject of investigation is (in contrast to most bias analyses) not the level but the correlation over t .

Before we start to evaluate the effects on the correlation with the cycle function $g(t)$, we have to think about the correlation that can be obtained in the unbiased case for $n, T \rightarrow \infty$. We define

$$\lim_{n, T \rightarrow \infty} \rho[E_t(s), g(t)] = \rho_C^{unbiased},$$

where $\rho_C^{unbiased}$ denotes the maximum correlation in cases of an unbiased mean function $E_t(s)$ with the cycle function $g(t)$. As Pearson's correlation coefficient measures the linear relationship between two variables, it is only invariant towards linear transformations $h(\cdot)$. Since $E_t(s)$ only depends on the cycle function $g(t)$ and $\Phi(\cdot)$ is continuous, but not linear in Equation [1], $\rho_C^{unbiased} < 1$. This result is not surprising since a 3-level trait always includes less information. For this reason, we have to rescale $\rho_C^{obs} = \rho[E_t(s^{obs}), g(t)]$ when inspecting possible bias effects to

$$\tilde{\rho}_C := \tilde{\rho}_C^{obs} = \frac{\rho_C^{obs}}{\rho_C^{unbiased}} \quad [3]$$

We notice that $\tilde{\rho}_C$ is not a real correlation, but the adjustment in Equation [3] is necessary to display the differences in correlation. However, in nearly all of the cases $\rho_C^{unbiased}$ is close to 1. The non-linearity arises from the cumulative density function of ε . Linearity can only be received in cases of a rectangular distribution $\varepsilon \sim U(c_l, c_u)$. However, this might be unrealistic in most settings. Stangl (2009) evaluated both variables of the Ifo index on a visual analog scale and found evidence that the distribution of s^* is similar to a bell curve. A consequence of $\rho_C^{unbiased} < 1$ is that an observed biased indicator $E_t(s^{obs})$ may have a higher correlation with the cycle function $g(t)$ than the unbiased indicator $E_t(s)$. We can not analytically derive in which cases this effect would appear, but we show in Annex B that a perfect correlation can never be obtained.

3. Simulation study

3.1. Definition of cycle functions

Before we run our simulation study, we have to specify our cycle functions $g(t)$. As noted in Section 2, the ideal case of a business cycle is a sinus function in time, that is, $\sin(t)$. In addition, historical business cycle theory identifies four types of overlapping cycles: Kitchin (3-5 years), Juglar (7-11), Kuznets (15-25) and Kondratiev (45-60).³ We first want to inspect the effects of our acceptance functions $\pi^s(t)$ on a simple sinus-function for the cycle $g(t)$, so $g(t) \approx \sin(t)$. We refer t to represent months and define our simple sinus-function as a Juglar cycle with 10 years, that is, 120 months. So, we have to re-scale $\sin(t)$ by a constant k to $\sin(t/k)$. The first full cycle is reached at 2π , so $\sin(120/k) \approx \sin(2\pi) \Rightarrow k = k_{Jug} \approx 20$.⁴ Then, our first cycle $g(t)$ to inspect is defined as

$$g_1(t) := \sin(t/k_{Jug}) = \sin(t/20).$$

Secondly, we define a cycle function $g(t)$ including all four cycles mentioned above. One full Kitchin 4-years cycle covers 48 months, one Kuznets 20-years cycle 240 months and one Kondratiev 50-years cycle 600 months. As above, the scaling parameters k turn to: $k_{Kit} \approx 8$, $k_{Kuz} \approx 38$ and $k_{Kon} \approx 95$. Therefore, the cycle function is defined as

$$\begin{aligned} g_2(t) &:= \sin(t/k_{Kit}) + \sin(t/k_{Jug}) + \sin(t/k_{Kuz}) + \sin(t/k_{Kon}) \\ &= \sin(t/8) + \sin(t/20) + \sin(t/38) + \sin(t/95) \end{aligned}$$

Of course, $g_1(t)$ and $g_2(t)$ display a very ideal case of a business cycle, which is usually not observed in reality. To come closer to a more realistic process, we define a third cycle function $g_3(t)$ by a stationary AR(1)-process

$$g_3(t) := 0.9 \cdot g_3(t-1) + u_t$$

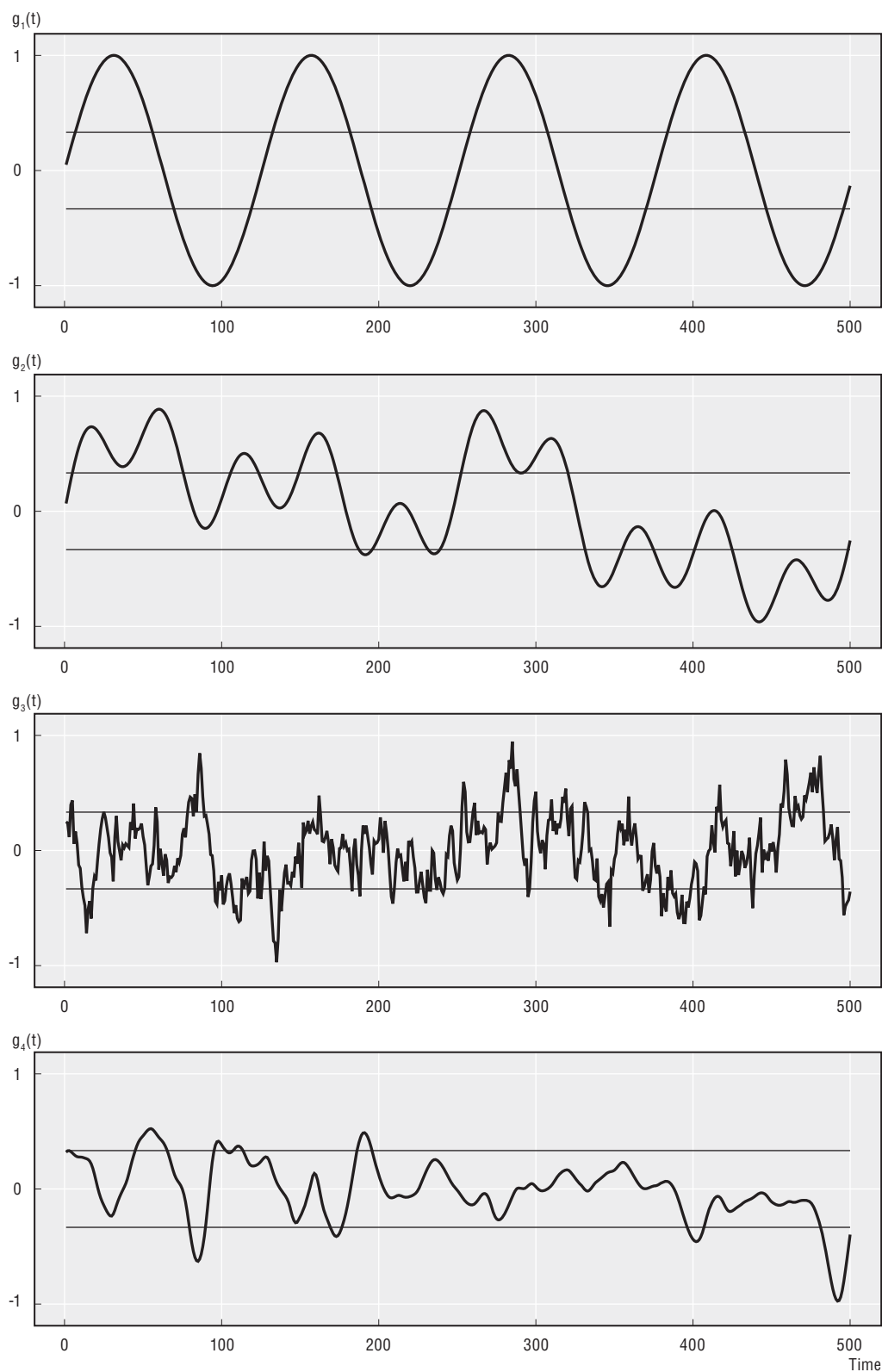
with $g_3(0) = 0$ and $u_t \sim N(0,1)$. The series $g_3(t)$ can be regarded as a monthly growth rate. One example for such a latent process is the inflation question of the Ifo World Economic Survey, which is also asked on a 3-level scale. Our fourth cycle function is based on a real data example, the US industrial production. We define $g_4(t)$ by extracting the output gap with the Hodrick-Prescott filter for time series (Hodrick and Prescott, 1997).

All cycle functions $g_1(t)$, $g_2(t)$, $g_3(t)$ and $g_4(t)$ are displayed in Figure 1 for $T = 500$. For an easier comparison, we standardise all cycle functions to $g_j(t) \in [-1,1]$. The thresholds τ^s are defined as $\tau^+ = 1/3$ and $\tau^- = -1/3$ so that the range of every function is divided into three parts of equal size which fit the assumptions for the calculation of the balance statistics in Section 2.1.

3.2. Definition of response functions

After defining cycle functions $g_j(t)$ we introduce different nonresponse patterns determined by the acceptance functions $\pi^s(t)$. Of course, the number of different functions $\pi^s(t)$ is infinite (and thus the number of combinations), as well as the number of cycle functions $g_j(t)$. Nevertheless, we examine the effects of some main types of $\pi^s(t)$ on the correlations and perform a simulation study on this.

There exists only a small fraction of analyses about nonresponse biases in business tendency surveys. A dependence of the overall response rate from the business cycle was found in Seiler (2010) for the Ifo Business Survey and Harris-Kojetin and Tucker (1999) for the US Current Population Survey. Seiler and Heumann (2013) estimated the missing

Figure 1. **Different cycle functions with the appropriate thresholds τ^S** 

observations in the *Ifo Business Survey* and found notable cycle-dependent differences between the three states for the business situation and business expectations variable. Wang (2004) analysed the *Norwegian Business Tendency Survey* and states that “there is variation in the distribution of non-response”.

We therefore focus on different types of nonresponse patterns which seem to be problematic in the case of business tendency surveys: fixed (time-independent), cycle-dependent, cycle-shifted, monotone and random. All time-independent structures describe general differences in the responding behaviour when being in the appropriate state. For example, one can imagine that the respond rate is higher when the firms’ situation is bad and the company wants to complain about that. The other effect can occur, when the firms’ state is good and they want to let the others know it. Trends, nonlinear as well as linear, reflect general changes of response behaviour over time. For example, the probability to respond may have decreased for those firms which are in a good situation. A dependence from the cycle, as well as cycle shifts, may occur as the decision to respond can be affected by the latent variable, for example, the company has a higher probability for nonresponse when the firm remarks that their situation is bad in relation to the market. For robustness checks, we also include random probabilities for $\pi^s(t)$. Regardless of the type of $\pi^s(t)$, we have to define $0 \leq \pi^s(t) \leq 1$, $\forall t$, to obtain a correct specified probability function.

3.3. Simulation study

As we are able to evaluate the mean function for the observed units $E_t(s^{obs})$ directly with Equation [2], we focus our simulation study on analysing different situations of nonresponse patterns. In Section 3.2, we defined five general types of acceptance rates: random, fixed, cycle-dependent, cycle-shifted and monotone. For every cycle function $g_j(t)$ and each of these bias patterns, we draw $Z = 1\,000$ different situations for the triple $\pi^+(t)$, $\pi^-(t)$ and $\pi^s(t)$. For example, for a time-independent bias pattern we draw a probability for each $\pi^s(t)$ and fix these over t . In contrast, a random bias pattern would lead to a draw of each $\pi^s(t)$ for every t . For cycle-dependent and cycle-shifted bias patterns, $\pi^s(t)$ are functions of $g_j(t)$ whereas monotone and linear bias patterns are functions in t . In every iteration z , different scaling parameters (which are defined in the subsequent paragraphs) are drawn. Therefore, we are able to reflect very different situations of nonresponse patterns. For each of these patterns, we calculate the correlations defined in Section 2.3.

3.3.1. Random and fixed

We start our analysis with random, unstructured response functions $\pi^s(t)$. For every iteration z , response function $\pi^s(t)$ and t , we draw from the Uniform distribution $U(0, 1)$ to obtain randomness, that is, $\pi_z^s(t) \sim U(0, 1)$, $\forall s, t, z$. For the time-independent case, we draw only once for every iteration z and state s and fix this value over t , that is, these functions may be different across s but do not fluctuate over t , which leads to $\pi_z^s(t) = \pi_z^s \sim U(0, 1)$.

3.3.2. Cycle-dependent

In contrast to random and fixed bias patterns, all cycle-related as well as monotone patterns are functions in $g_j(t)$ or t . As we have to ensure that $0 \leq \pi^s(t) \leq 1$, $\forall t$, we can not use $g_j(t)$ directly and have to standardise it. In Section 3.1 we defined $g_j(t) \in [-1, 1]$. So, our cycle-dependent acceptance rate function has the form

$$\pi_{j,k_c}^C(t) := \frac{g_j(t)+1}{2k_c} \quad [4]$$

for $j = 1, \dots, 4$. For our simulation study, we draw scaling parameters $k_c \in \{1, \dots, 10\}$. As Equation [4] defines a positive relationship between $\pi^s(t)$ and $g_j(t)$, we also allow negative relationships $1 - \pi_{j,k_c}^c(t)$. This is done by drawing a 0/1-dummy d_l if to use $\pi_{j,k_c}^c(t)$ or $1 - \pi_{j,k_c}^c(t)$. d_l will also be used for the cycle-shifted and monotone response functions.

3.3.3. Cycle-shifted

As survey business cycle indicators including expectation questions are commonly used for now- and forecasting, that is, should contain leading information, we now assume that the selection process also includes leading or lagging information. It can be assumed that the respondents might anticipate coming development⁵ or might react on the most recent development. To reflect this, we define the acceptance rate functions $\pi^s(t)$

$$\pi_{j,k_c}^s(t + k_T) = \pi_{j,k_c}^c(t + k_T) \quad [5]$$

for $j = 1, \dots, 4$. Equation [5] has the same form as the cycle-dependent acceptance rates in Equation [4], but is shifted in time by k_T units. Therefore, we first draw $k_T \in \{-12, 11, \dots, 12\}$, that is, we allow a maximum of 12 months for leads or lags. In addition, we draw k_c and d_l as in the cycle-dependent case.

3.3.4. Monotone

Cycle-dependent and cycle-shifted response rates $\pi^s(t)$ are direct transformations of $g(t)$, that is, the underlying cycle function. To allow $\pi^s(t)$ to be dependent from t but not from $g(t)$, we specify functions $\pi^s(t)$ which display general trends of monotone or linear form. Therefore, we define two different types of acceptance rates $\pi^s(t)$:

$$\pi_{k_L}^L(t) = \frac{t + 1000}{k_L}$$

and

$$\pi_{k_M}^M(t) = \Phi\left(\frac{t - 250}{k_M}\right)$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. $\pi_{k_L}^L(t)$ was chosen to have values between 0 and 1 for $t \in \{1, \dots, T\}$. k_L and k_M are scaling parameters which control for the steepness of $\pi_{k_M}^M(t)$ and $\pi_{k_L}^L(t)$. In a first step, we draw d_L which decides to use $\pi^L(t)$ or $\pi^M(t)$. Conditional on d_L we draw $k_L \in [2\,000, 4\,000]$ or $k_M \in [100, 500]$. Finally, we draw d_l as in the cycle-dependent case to allow for decreasing trends, for example, by $1 - \pi^L(t)$.

3.3.5. Mixed

Of course, the five main bias patterns described above can also appear simultaneously. To reflect this situation, we mix these patterns by introducing a variable d_p which specifies the bias pattern to use. After this classification, the simulations are done as described above.

3.4. Dispersion measure

From Section 2.2, we know that MCAR occurs in cases of $\pi^r(t) = \pi^s(t)$, $\forall r, s \in S$, and $\forall t \in \{1, \dots, T\}$. We want to analyse if a higher average variation in the acceptance rates $\pi^s(t)$ between the different states for given t leads to a lower correlations as high differences of $\pi^s(t)$

lead to high differences for $\pi^s(t)/\bar{\pi}(t)$ and thus the weighted balance statistics in Equation [4]. To evaluate this variation for the whole time series, we introduce a variation measure

$$v = \frac{1}{T} \sum_{t=1}^T \text{Var}_s[\pi^s(t)|t] \quad [6]$$

The inner part $\text{Var}_s[\pi^s(t)|t]$ is the variance across $\pi^s(t)$ for given t which is then averaged over t . As $\pi^s(t) < 1, \forall t$, $\text{Var}_s[\pi^s(t)|t] < 1$ and therefore also $v < 1$.

3.5. Results

Figure 2 shows the scatterplots for the pairs $(\tilde{\rho}_c, v)$ according to the five different main bias patterns plus the mixed case as defined in Section 3.3. The vertical axis are scaled to $[-1, 1]$ whereas the horizontal axis are not scaled due to the fact that the range of v differs strongly between the various patterns. In addition, boxplots for the correlations and the dispersion parameters are drawn and the underlying cycle functions $g_j(t)$ are displayed with different shapes. The boxplots on the y-axis clearly show that the majority of biased indicators still receives high correlations close to 1. This holds in particular for the fixed, cycle-dependent, cycle-shifted and monotone types $\pi^s(t)$. On average, the lowest correlations appear in the random and mixed case. Furthermore, clusters according to the different cycle functions can be seen in the random case. For this bias pattern type, the ideal cycle functions $g_1(t)$ and $g_2(t)$ seem to be less affected by the nonresponse bias. As the random bias pattern includes more unstructured uncertainty in the data, the correlations become smaller on average. With exception of the monotone bias types, there seems to be no clear connection between the dispersion v and the correlations. Correlations remain high, even if the dispersion increases.

To get an idea of how the acceptance rates $\pi^s(t)$ transform the biased indicator $E_t(s^{obs})$, we show some extreme cases for every main bias type in Figures 2 and 3. We select the cycle function $g_2(t)$ to display the effects. On the left side of Figure 3, the cycle function $g_2(t)$ (... line), the unbiased mean function $E_t(s)$ (--- line) and the biased mean function $E_t(s^{obs})$ (— line) are drawn. The right side of Figure 3 shows the appropriate acceptance rates $\pi^+(t)$ (- - - line), $\pi^-(t)$ (- line) and $\pi^{\sim}(t)$ (... line). In Figure 2, a random, fixed and cycle-dependent case, including correlations $\tilde{\rho}_c$ as well as the dispersion parameter v , is displayed. The random bias pattern clearly shows that the correlations decrease, but the general underlying structure from the cycle function is still present. Smoothing approaches, such as the Hodrick-Prescott filter, could still be used in this case to reduce this effect. However, we have to notice here that sampling size effects will include even more uncertainty. The second row displays the effect of time-independent acceptance rates $\pi^s(t)$. With $\pi^-(t) = 0.1$ we chose a low acceptance for the negative replies. This led to a shift of $E_t(s^{obs})$ upwards.⁶ However, the correlation still remained high, as we already saw in the scatterplots in Figure 3. In the last row, the cycle-dependent case is shown. $\pi^s(t)$ were chosen to display a very extreme case with high negative correlations. This effect appears when the acceptance rates are anti-cyclical, that is, negative replies are seldom in recessions and positive ones in boom times. In our case, we have $\pi^+(t) = 1 - \pi^{c,1}(t)$ and $\pi^-(t) = \pi^{c,1}(t)$.⁷ Figure 4 shows the other three bias types: cycle-shifted, monotone and mixed cases. Although the correlations remained high, it can clearly be seen for the cycle-shifted case that the lag/lead structure of the observed indicator has changed. The bias pattern led to a shift into the future which caused a decrease of leading information of the indicator. The second row shows the effect of monotone and linear acceptance rates $\pi^s(t)$ on the indicator. This type introduced a clear trend movement of the indicator which is diametrical to the underlying cycle. This was caused by an increase of acceptance of the positive replies and a simultaneous decrease of the negative replies. In this case, the correlations are also negative. The last type of bias pattern is the mixed case where we combined a cycle-shifted [for $\pi^+(t)$], a monotone [for $\pi^-(t)$] and a fixed [for $\pi^{\sim}(t)$] bias pattern. Although the acceptance rates are of very different kind, the correlations still remain high.

Figure 2. Scatterplots for correlations $\tilde{\rho}_c$ and dispersions v for six different types of acceptance rates $p^s(t)$ as defined in Section 3.3

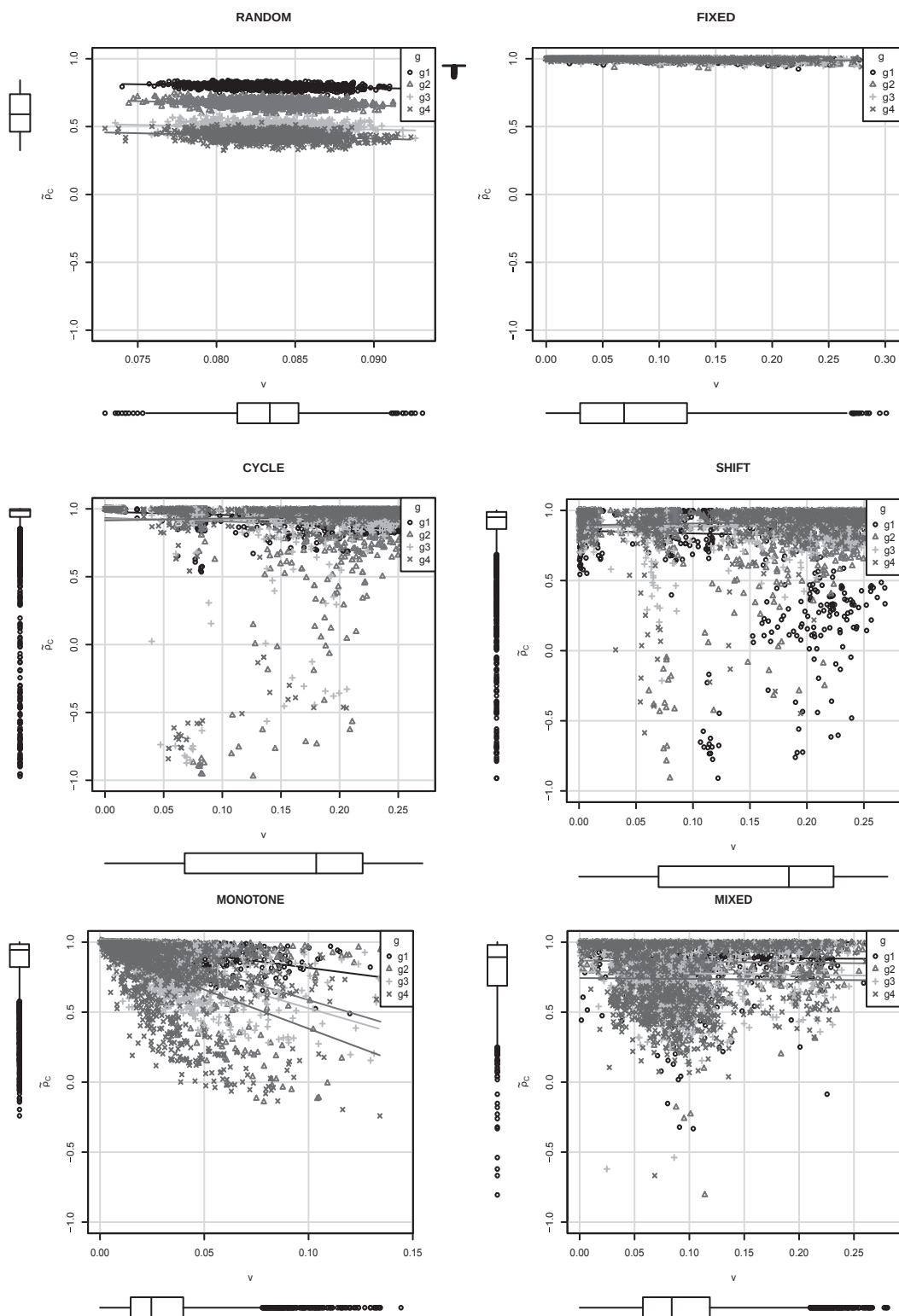
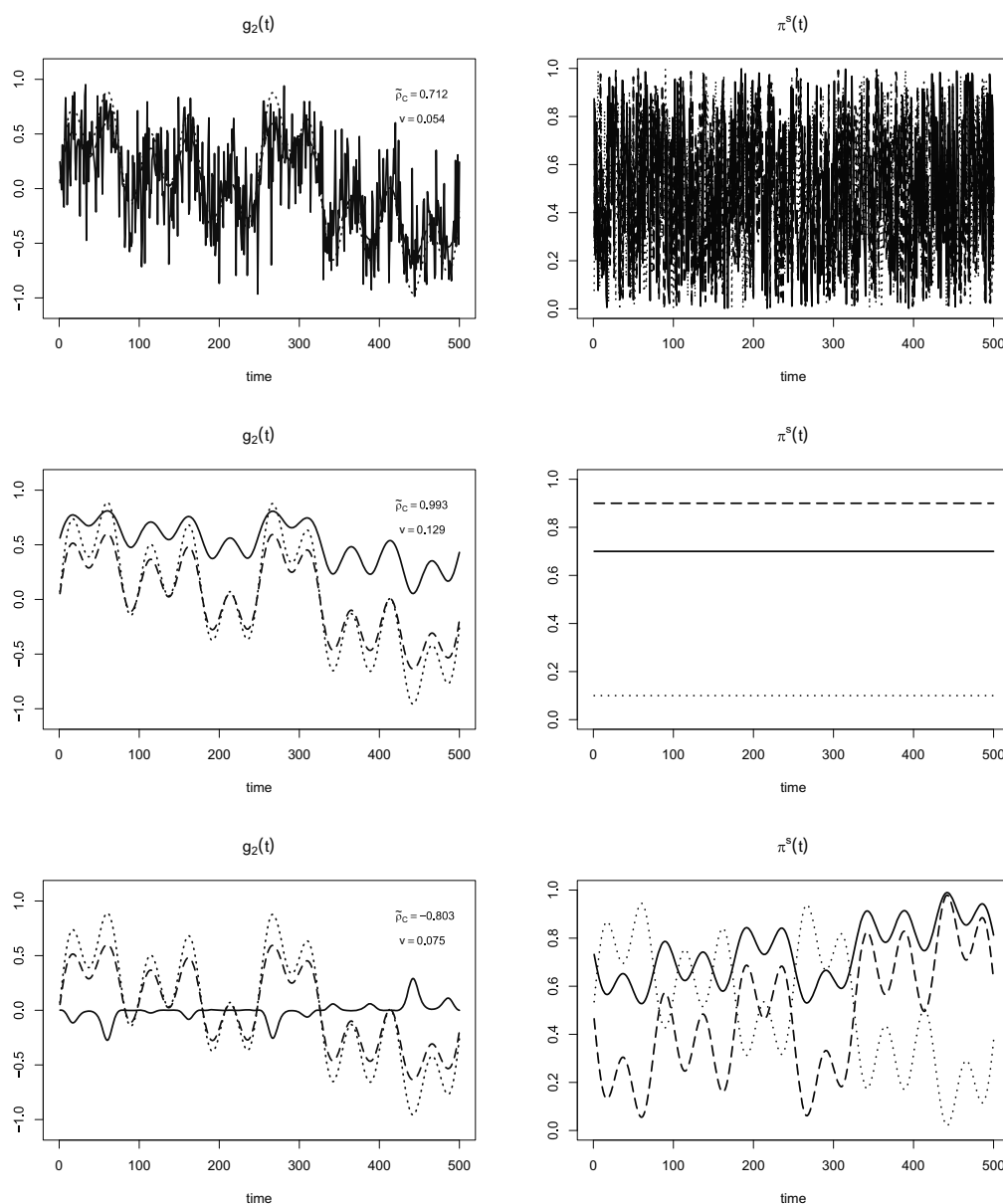
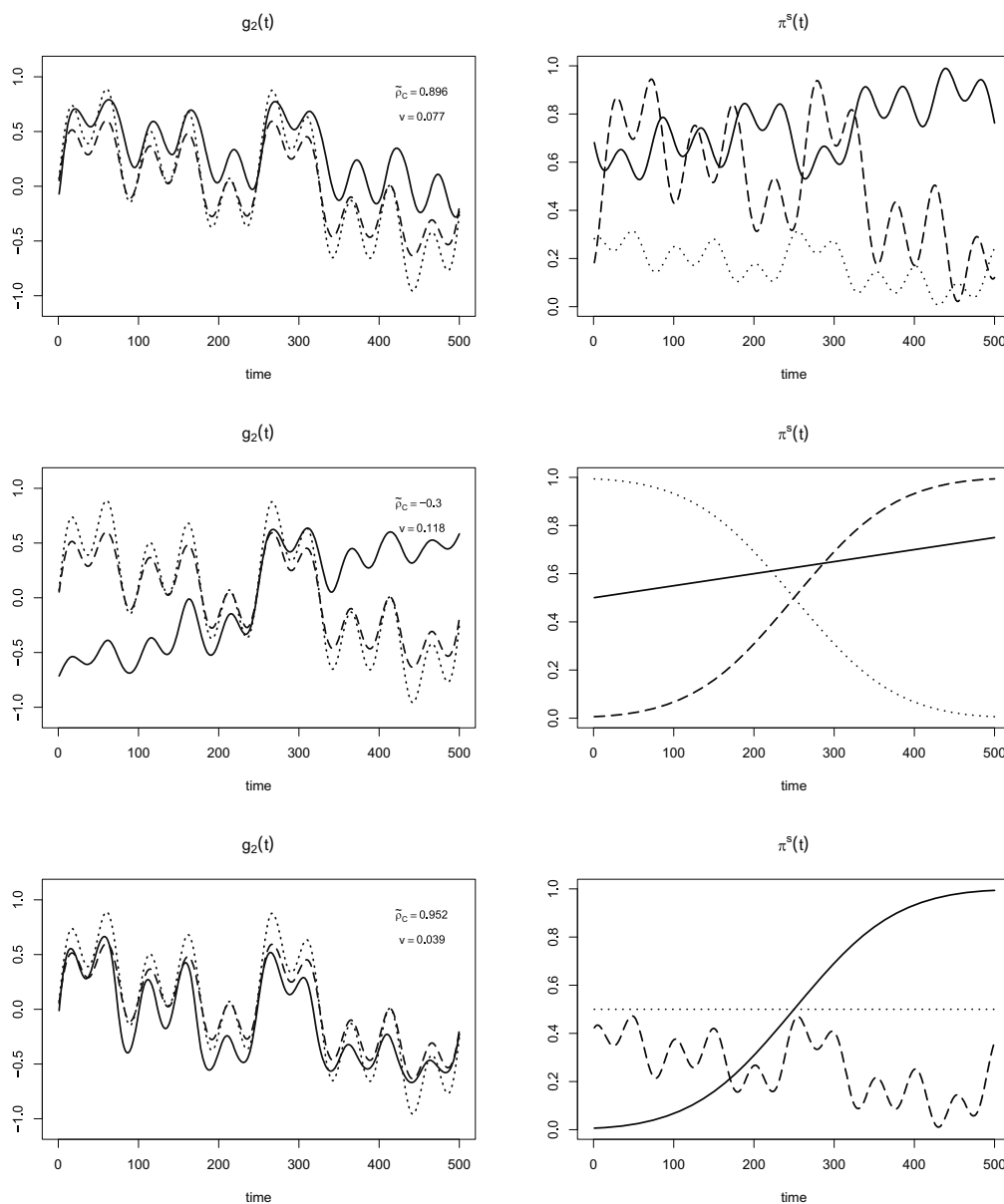


Figure 3. **Effects of bias patterns on cycle function $g_2(t)$: Part I**

Notes: Left: cycle function $g_2(t)$ (... line), unbiased mean function $E_t(s)$ (--- line) and observed mean function $E_t(s^{obs})$ (— line).

Right: acceptance rates $\pi^+(t)$ (- - line), $\pi^-(t)$ (dark grey line) and $\pi^*(t)$ (... line).

From top to bottom: Random, fixed [with $\pi^+(t) = 0.9$, $\pi^-(t) = 0.7$ and $\pi^*(t) = 0.1$] and cycle-dependent [with $\pi^+(t) = 1 - \pi_{2,1}^c(t)$, $\pi^-(t) = 1 - \pi_{2,2}^c(t)$ and $\pi^*(t) = \pi_{2,1}^c(t)$] patterns.

Figure 4. Effects of bias patterns cycle function $g_2(t)$: Part II

Notes: Left: cycle function $g_2(t)$ (... line), unbiased mean function $E_t(s)$ (--- line) and observed mean function $E_t(s^{obs})$ (— line).

Right: acceptance rates $\pi^+(t)$ (- - - line), $\pi^-(t)$ (— line) and $\pi^s(t)$ (... line).

From top to bottom: Cycle-shifted [with $\pi^+(t) = \pi_{2,1}^s(t-12)$, $\pi^-(t) = 1 - \pi_{2,2}^s(t+3)$ and $\pi^s(t) = \pi_{2,3}^s(t+12)$], monotone [with $\pi^+(t) = \pi_{100}^M(t)$, $\pi^-(t) = \pi_{2000}^L(t)$ and $\pi^s(t) = 1 - \pi_{100}^M(t)$] and mixed [with $\pi^+(t) = \pi_{2,2}^s(t+12)$, $\pi^-(t) = \pi_{100}^M(t)$ and $\pi^s(t) = 0.5$] patterns.

4. Conclusion

In this paper, a methodological framework for the widely used balance statistics indicators for economic time series was built. As these indicators are based on surveys, we included nonresponse into this framework to evaluate their effects. Due to the fact that the indicators are artificial, we focused our analyses on the effects of nonresponse on correlation. The analyses showed that the correlations between the observed indicators and the underlying cycle still remain high in nearly all of the cases. Of course, some bias patterns exist where the shape of the indicator is transformed strongly, but we notice that these patterns are seldom and unrealistic to appear in real-life situations. However, some patterns might cause certain problems such as a complete switch of the indicator or a shift which might affect the leading performance. The analyses also showed that there exists only a small connection between the dispersion in the acceptance rates and the correlations. Stronger biases seem to be caused by a simultaneous presence of unfavorable combinations of nonresponse patterns. If the acceptance rates differ strongly for random or monotone transformations of t , the correlation with the underlying cycle decreases on average. The reasons for this robustness are probably due to the fact that all states contain the oscillation of the underlying cycle. All of the three states are affected by $g(t)$ and therefore include its information. When the probability for one state to be observed decreases, the other states still include enough information. Only for very extreme, in particular time- or cycle-dependent cases, significant decreases in correlation might appear.

Notes

1. τ^- left undefined.
2. We abstract from different types of weighting that may occur in real data sets.
3. For a recent study to the 2008-09 economic crisis with respect to the different cycles, see Korotayev and Tsirel (2010).
4. We notice that this procedure is inaccurate but the cycles, as defined in the literature, do also not have an exact length. The cycles in this section are only selected for representation of the nonresponse bias effects.
5. The participants of the Ifo Business Survey are asked every month to give an assessment on the future business situation. This might also affect the decision to respond.
6. A stretch of the indicator can appear in cases where $\pi^-(t) \ll \pi^+(t)$, $\pi^-(t)$ whereas a compression of the indicators appears in cases of $\pi^-(t) \gg \pi^+(t)$, $\pi^-(t)$.
7. As the negative replies have higher probability in recession times, $\pi_{2,1}^C(t)$ is anti-cyclical for the “-” category.

References

- Anderson, O. (1951), “Konjunkturtest und Statistik”, *Allgemeines Statistisches Archiv*, 35, 209-220.
- Anderson, O. (1952), “The Business Test of the IFO-Institute for Economic Research”, *Revue de l'Institut International de Statistique*, 20, 1-17.
- Ang, A., G. Bekaert and M. Wei (2007), “Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?”, *Journal of Monetary Economics*, 54, 1163-1212.
- Bartholomew, D.J. and M. Knott (1999), *Latent Variable Models and Factor Analysis: A Unified Approach*, Charles Grin.
- Breitung, J. and M. Schmeling (2013), “Quantifying Survey Expectations: What’s Wrong with the Probability Approach?”, *International Journal of Forecasting*, 29, 142-154.
- Carlson, J.A. and M. Parkin (1975), “Inflation Expectations”, *Economica*, 42, 123-138.

- Drechsel, K. and R. Scheufele (2012), "The Performance of Short-Term Forecasts of the German Economy Before and During the 2008/2009 Recession", *International Journal of Forecasting*, 28, 428-445.
- European Union (2006), "Joint Harmonised EU Programme of Business and Consumer Surveys", *Official Journal of the European Union*, 49, 5-8.
- Gayer, C. (2005), "Forecast Evaluation of European Commission Survey Indicators", *Journal of Business Cycle Measurement and Analysis*, 7, 157-183, Doi: <http://dx.doi.org/10.1787/jbcm-a-v2005-2-en>.
- Harris-Kojetin, B. and C. Tucker (1999), "Exploring the Relation of Economical and Political Conditions with Refusal Rates to a Government Survey", *Journal of Official Statistics*, 15, 167-184.
- Henzel, S. and T. Wollmershäuser (2005), "Quantifying Inflation Expectations with the Carlson-Parkin Method: A Survey-Based Determination of the Just Noticeable Difference", *Journal of Business Cycle Measurement and Analysis*, 2, 321-352, Doi: <http://dx.doi.org/10.1787/jbcm-a-v2005-3-en>.
- Hodrick, R. and E. Prescott (1997), "Postwar US Business Cycles: An Empirical Investigation", *Journal of Money, Credit, and Banking*, 29, 10-16.
- Janik, F. and S. Kohaut (2012), "Why Don't They Answer? Unit Non-response in the IAB Establishment Panel", *Quality & Quantity*, 46, 917-934.
- Kholodilin, K.A. and S. Siliverstovs (2006), "On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence", *Journal of Economics and Statistics*, 226, 234-259.
- Korotayev, A.V. and S.V. Tsirel (2010), "A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and Kitchin Cycles in Global Economic Development, and the 2008-2009 Economic Crisis", *Structure and Dynamics*, 4, 3-57.
- Little, R.J.A. and D. Rubin (2002), *Statistical Analysis with Missing Data*, Wiley.
- Nardo, M. (2003), "The Quantification of Qualitative Survey Data: A Critical Assessment", *Journal of Economic Surveys*, 17, 645-668.
- OECD (2003), *Business Tendency Surveys: A Handbook*, OECD Publishing, Paris, Doi: <http://www.dx.doi.org/10.1787/9789264177444-en>.
- Robinsonov, N. and K. Wohlrabe (2010), "Freedom of Choice in Macroeconomic Forecasting", *CESifo Economic Studies*, 56, 192-220.
- Seiler, C. (2010), "Dynamic Modelling of Nonresponse in Business Surveys", *Ifo Working Paper*, 93, Ifo Institute for Economic Research.
- Seiler, C. and C. Heumann (2013), "Microdata Imputation and Macrodata Implications: Evidence from the Ifo Business Survey", *Economic Modelling*, 35, 722-733.
- Stangl, A. (2007), "European Data Watch: Ifo World Economic Survey Micro Data", *Journal of Applied Social Science Studies*, 127, 487-496.
- Stangl, A. (2009), "Essays on the Measurement of Economic Expectations", Ph.D. thesis, Ludwig-Maximilians-Universität, München.
- Theil, H. (1952), "On the Shape of Economic Microvariables and the Munich Business Test", *Revue de l'Institut International de Statistique*, 20, 105-120.
- Thompson, K.J. (2009), "Conducting Nonresponse Bias Analysis for Two Business Surveys at the US Census Bureau: Methods and (Some) Results", in *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1288-1301.
- Wang, J.H. (2004), "Non-Response in the Norwegian Business Tendency Surveys", Document, No. 2004/3, Statistics Norway, Department of Economic Statistics.
- Willimack, D.K. and E. Nichols (2010), "A Hybrid Response Process Model for Business Surveys", *Journal of Official Statistics*, 26, 3-24.

ANNEX A

Calculation of balance statistics

As balance statistics indicators are based on balances and not on the calculation of the mean, we have to ensure that we obtain the same results. We define, as in Section 2.1, $+$ $\equiv 1$, $=$ $\equiv 0$ and $-$ $\equiv -1$. Then, let b_t the balance statistics at time t , that is,

$$b_t = \frac{\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)}{n_t}$$

Therefore, the mean is defined as

$$\begin{aligned} E(b_t) &= E_t(b) = E \left[\frac{\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1)}{n_t} \right] \\ &= \frac{1}{n_t} E \left[\sum_{i=1}^{n_t} I(s_{i,t} = 1) - \sum_{i=1}^{n_t} I(s_{i,t} = -1) \right] \\ &= P(s_{i,t} = 1) - P(s_{i,t} = -1) \\ &= \left\{ 1 - \Phi \left[\frac{\tau^+ - g(t)}{\sigma} \right] \right\} - \Phi \left[\frac{\tau^- - g(t)}{\sigma} \right] \\ &= E_t(s) \end{aligned}$$

ANNEX B

Proof of $\rho_c^{obs} \neq 1$

In Section 2.3, we stated that $\rho_c^{obs} \neq 1$. The correlation $\rho_c^{obs} = \rho[E_t(s^{obs}), g(t)] = 1$ if

$$E_t(s^{obs}) = a \cdot g(t) + b$$

$\frac{1}{\bar{\pi}(t)} \cdot [\pi^+(t) \cdot \Phi^+(t) - \pi^-(t) \cdot \Phi^-(t)] = a \cdot g(t) + b$
Assuming $a \neq 0$, Equation [4] can be written as

$$\left[\frac{\pi^+(t)}{\bar{\pi}(t)} \cdot \Phi^+(t) = a^+ \cdot g(t) + b^+ \right] \wedge \left[\frac{\pi^-(t)}{\bar{\pi}(t)} \cdot \Phi^-(t) = a^- \cdot g(t) + b^- \right]$$

with $a^+, a^- \neq 0$. It is important to notice that a, a^+, a^-, b, b^+ and b^- have to be constant and process: As $\Phi^+(\cdot)$ and $\Phi^-(\cdot)$ are nonlinear if $\varepsilon \sim U(c_l, c_u)$, linearity in $\Phi^s(\cdot)$ can only be obtained directly via the inverse of $\Phi^s(\cdot)$. As the $\pi^s(t)$ have a *multiplicative effect* on $\Phi^s(\cdot)$ in Equation [4], this way for correction is not possible. Another approach to obtain a linear relationship would be to constrain $\pi^+(t) = a \cdot g(t) + b$ or $\pi^-(t) = a \cdot g(t) + b$ and to oblige that $\Phi^+(t)$ and $\Phi^-(t)$ are reduced to time-invariant constants via division by $\bar{\pi}(t)$. But because we have stated above that $\pi^+(t)$ or $\pi^-(t)$ have to be a linear transformation of $g(t)$ and they are also included in $\bar{\pi}(t)$, $\Phi^+(t)$ and $\Phi^-(t)$ can not be changed simultaneously into time-invariant effects. Therefore, it is not possible to obtain a perfect correlation regardless of the type of $\pi^s(t)$. However, this fact does not exclude the possibility of receiving a higher correlation for a biased indicator $E_t(s^{obs})$. In these cases, the functions $\frac{\pi^+(t)}{\bar{\pi}(t)}$ and $\frac{\pi^-(t)}{\bar{\pi}(t)}$ partially eliminate the bias of $\Phi(\cdot)$.

ANNEX C

Discussion on thresholds

In Section 2.1, we made the assumption of time- and individual-independent thresholds as well as time- and individual-independent normality of ε_i . For the ZEW Indicator of Economic Sentiment, Breitung and Schmeling (2013) found that these thresholds are generally not time-independent. In our study, however, a change of the thresholds and/or the form of the density function of ε_i would affect both the unbiased and the biased indicator since these variables define the probabilities $\Phi^s(t)$ which are included in $\bar{\pi}(t) = \sum_s \pi^s(t) \cdot \Phi^s(t)$. Therefore, a change of τ^s and/or ε_i would lead to a different *unbiased* indicator $E_t(s)$ and maximum correlation $\rho_c^{unbiased}$.

Cohesion within the euro area and the US: A wavelet-based view

by

António Rua* and Artur Silva Lopes**

The analysis of synchronisation of macroeconomic fluctuations across countries or regions has been crucial, for example, for the debate on economic integration. In this paper, we propose a multivariate measure of synchronisation to assess cohesion across countries or regions by resorting to wavelet analysis. This wavelet-based measure of cohesion allows one to study how synchronisation has evolved over time and across frequencies simultaneously. In particular, we investigate the cohesion among euro area countries and within the US, both at the regional and state levels, over the last decades. In addition, an analysis at the sectoral level is also conducted. The results obtained unveil a noteworthy heterogeneity and highlight the usefulness of a wavelet-based measure of cohesion.

Keywords: cohesion, wavelets, time-frequency, output growth

JEL classification: C40, E32

* Corresponding author: Banco de Portugal and ISEG-Technical University of Lisbon. Address: Economic Research Department, Banco de Portugal, Av. Almirante Reis, No. 71, 1150-012 Lisboa, Portugal. antonio.rua@bportugal.pt.

** CEMAPRE and ISEG-Technical University of Lisbon.

1. Introduction

The assessment of comovement among different countries or regions has long been a topic of interest in economics. In fact, the degree of synchronisation of macroeconomic fluctuations across countries or regions plays a key role in the discussion about the attractiveness of economic integration. In this respect, the debate about the European monetary union has dominated the literature over the past decade [see, for example, de Haan et al. (2008) for a literature survey regarding business cycle synchronisation in the euro area]. In particular, building on the work of Mundell (1961) on optimum currency areas, it has been argued that the cost of joining a monetary union will be low if countries have highly synchronised business cycles. However, it has also been pointed out that economic integration itself can affect the synchronisation of macroeconomic fluctuations. For instance, Frankel and Rose (1998) claim that the removal of trade barriers induces a higher symmetry of output fluctuations, while Rose (2000) provides evidence that a common currency results in more trade. In contrast, Krugman (1993) argues that a higher level of trade can lead to a higher economic specialisation and less synchronised business cycles. See, for example, Kalemli-Ozcan et al. (2001) for an overview of the effects of economic integration on output fluctuations symmetry.

Several measures have been used in the literature to assess synchronisation but the Pearson correlation coefficient remains the most popular because it summarises the degree of comovement through time in a single value. Alternatively, one can resort to spectral analysis, to obtain insights about the relationship at the frequency level. Croux et al. (2001) have suggested a spectral-based measure, the dynamic correlation, which is conceptually similar to the contemporaneous correlation but allows to measure comovement at the frequency level [which has been used by Tripier (2002), Rua and Nunes (2005) and Camacho et al. (2006), among others]. However, while the Pearson correlation coefficient completely disregards the relationship at the frequency level, with the dynamic correlation proposed by Croux et al. (2001) all the information about the time dependence of comovement is lost.

To overcome such caveats, Rua (2010) has proposed a wavelet-based measure of comovement to assess simultaneously how two variables are contemporaneously related at different frequencies and how such relationship has evolved over time. Wavelet analysis constitutes a very promising tool as it represents a refinement in terms of analysis, in the sense that both time and frequency domains are taken into account. Recent applications of wavelets in economics can be found, for instance, in Gallegati and Gallegati (2007), Gallegati et al. (2008), Yogo (2008), Crowley and Mayes (2008), Rua and Nunes (2009, 2012), and Rua (2011, 2012) [see Crowley (2007) for a survey].

In order to take on board more than two series when assessing comovement, Croux et al. (2001) have extended the dynamic correlation to the multivariate case and named this generalised measure as cohesion. As cohesion draws on the dynamic correlation measure, the phase shifts are disregarded and the focus is on the contemporaneous

comovement among the series. As stressed by de Haan et al. (2008), this measure provides a useful summary statistic on the degree of comovement across countries or regions while avoiding the problem of choosing a base country or region. Cohesion has been applied by Croux et al. (2001) to assess the synchronisation of output fluctuations between European countries and across the US states and regions, by Carlino and DeFina (2004) to study the comovement in employment across the US states and sectors, by Crone (2005) to evaluate the business cycle cohesion within US regions, and *inter alia*, by Eickmeier and Breitung (2006) for assessing output growth and inflation cohesion between European countries.

Following Croux et al. (2001), we extend the bivariate measure proposed in Rua (2010) to the more general case in order to obtain a measure of cohesion in the wavelet domain. The resulting measure allows one to assess how cohesion has evolved over time and across frequencies simultaneously. Note that it can provide information about the strength of the relationship at a given frequency and time but it cannot uncover the lead-lag relation between the variables since, by definition, the phase differences are discarded. Although the lead-lag relation may be important by itself, one can interpret synchronisation precisely as the degree of comovement when one disregards phase shifts, that is, one is interested in the contemporaneous relationship between the series.

Focusing on output growth, we investigate the cohesion among euro area countries and within the US, at both the regional and state levels, over the last decades. We find that cohesion within the euro area has been higher at low frequencies, which is in line with Crowley and Mayes (2008) who also found a higher synchronisation of output growth for longer cycles than for shorter ones for the main euro area countries. We also find that cohesion in the euro area has increased for most frequencies, being very high since the end of the 1990s. In this respect, Eickmeier and Breitung (2006) also found that cohesion increases significantly across frequencies when the time span includes the period after the introduction of the euro. Moreover, all these findings are broadly in line with the results of Rua (2010), who considers only the major euro area countries and assesses all possible country pairs individually. Likewise in the euro area case, we find that output growth cohesion within US is positive at both the regional and state levels (Carlino and DeFina, 2004, found similar evidence with US employment data). It is also found that US cohesion is higher at the regional level than at the state level, both across frequencies and over time. This is in line with the findings of Croux et al. (2001) and Carlino and DeFina (2004) who argue that aggregation results in a higher comovement. In the US case, we find that cohesion is higher at the typical business cycle frequency range but it declined during the 1980s. However, this has been reverted with cohesion attaining a high level during the most recent economic and financial crisis episode. In addition, besides taking into account the spatial perspective when assessing cohesion, we also conduct an analysis at the sectoral level. Resorting to disaggregated data by eleven sectors for the euro area countries and US regions and states, we find a noteworthy heterogeneity in the results at the sectoral level.

The paper is organised as follows. In Section 2, the main building blocks are discussed and the wavelet-based measure of cohesion is presented. In Section 3, a data description is provided, while in Section 4 the empirical application is carried out for both the euro area and the US. Finally, Section 5 concludes.

2. Measuring cohesion in the wavelet domain

While the well-known Fourier transform decomposes the time series into infinite length sines and cosines (see, for example, Priestley, 1981), discarding all time-localisation information, the wavelet transform uses local base functions that can be stretched and translated with a flexible resolution in both frequency and time. In particular, the wavelet transform decomposes a time series in terms of some elementary functions, the daughter wavelets or simply wavelets $\psi_{\tau,s}(t)$. These wavelets result from a mother wavelet $\psi(t)$, that can be expressed as function of the time position τ (translation parameter) and the scale s (dilation parameter), which is related with the frequency,¹ that is,

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad [1]$$

To be a mother wavelet $\psi(t)$ must fulfil several conditions [see, for example, Percival and Walden (2000) for further discussion]. The continuous wavelet transform of a time series $x(t)$ with respect to $\psi(t)$ is given by the following convolution:

$$W_x(\tau, s) = \int_{-\infty}^{+\infty} x(t) \psi_{\tau,s}^*(t) dt = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \psi^*\left(\frac{t-\tau}{s}\right) dt \quad [2]$$

where $*$ denotes the complex conjugate.

The most commonly used mother wavelet for the continuous wavelet transform is the Morlet wavelet [see, for example, Aguiar-Conraria and Soares (2013) and references therein], which can be simply defined as:

$$\psi(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-\frac{t^2}{2}} \quad [3]$$

One can observe that the Morlet wavelet is a complex sine wave within a Gaussian envelope whereas ω_0 is the wavenumber. By increasing (decreasing) the wavenumber one achieves better (poorer) frequency localisation but poorer (better) time localisation. In practice, ω_0 is set to 6 as it provides a good balance between time and frequency localisation [see, for example, Adisson (2002) for further details]. Since the wavelength for the Morlet wavelet is given by $\frac{4\pi s}{\omega_0 + \sqrt{2 + \omega_0^2}}$ (see Torrence and Compo, 1998), then for $\omega_0 = 6$, the wavelet scale s is almost equal to the Fourier period, which eases the interpretation of wavelet analysis.

Given two time series $x_i(t)$ and $x_j(t)$, with wavelet transforms $W_{x_i}(\tau, s)$ and $W_{x_j}(\tau, s)$, one can define the cross-wavelet spectrum as $W_{x_i x_j}(\tau, s) = W_{x_i}(\tau, s) W_{x_j}^*(\tau, s)$. As the mother wavelet is in general complex, the cross-wavelet spectrum is also complex valued and it can be decomposed into real and imaginary parts. The measure proposed in Rua (2010) is given by:

$$\rho_{x_i x_j}(\tau, s) = \frac{\Re [W_{x_i x_j}(\tau, s)]}{\sqrt{|W_{x_i}(\tau, s)|^2 |W_{x_j}(\tau, s)|^2}} \quad [4]$$

where $\Re$ denotes the real part of the cross-wavelet spectrum. This wavelet-based measure $\rho_{x_i x_j}(\tau, s)$ allows one to evaluate the degree of comovement in the time-frequency space and assess over which periods of time and over which frequencies is the comovement higher. Basically, it plays a role as a contemporaneous correlation coefficient around each moment in time and for each frequency. Since it provides information about the

comovement not only at the frequency level but also over time, it can be seen as a generalisation of the dynamic correlation measure suggested by Croux et al. (2001).

In a similar fashion to Croux et al. (2001), who extended the dynamic correlation measure to the multivariate case, providing a measure of cohesion in the frequency domain, we extend the bivariate measure proposed by Rua (2010) to the more general case in order to obtain a measure of cohesion in the wavelet domain. In particular, cohesion is defined as the weighted average of the wavelet-based measure $\rho_{x_i x_j}(\tau, s)$ between all possible pairs of series:

$$\text{coh}(\tau, s) = \frac{\sum_{i \neq j} \varpi_{ij} \rho_{x_i x_j}(\tau, s)}{\sum_{i \neq j} \varpi_{ij}} \quad [5]$$

where ϖ_{ij} is the weight attached to the pair of series (i, j) s. As the $\rho_{x_i x_j}(\tau, s)$ range between -1 and 1 , the wavelet-based cohesion also varies between -1 and 1 . This measure allows one to quantify the extent of cohesion among several series at different frequencies and investigate if such global relationship has changed over time. Hence, it enables a richer analysis than the one possible with the cohesion measure suggested in Croux et al. (2001), which focus on the frequency level only. This is of particular importance as there is by now evidence that comovement can vary across frequencies as well as over time [see Crowley and Mayes (2008), Rua and Nunes (2009), Rua (2010) among others]. Moreover, the suggested wavelet-based cohesion allows to capture both features within an unified framework.

3. Data

Regarding the United States, we considered data at the regional and state levels provided by the Bureau of Economic Analysis (BEA) of the US Department of Commerce.² Annual real GDP by region and state is available from 1977 up to 2011. However, as nominal data is available from 1963, in order to cover a longer time span we resorted to the US GDP deflator prior to 1977. The disaggregation of regional and state output by sectors is also provided by the BEA for the same sample periods, and again we used the US corresponding deflator, so as to obtain volume series since 1963 and up to 2011. In order to ease the comparison between the US and the euro area, we considered eleven sectors, namely: 1) Agriculture, hunting, forestry, and fishing; 2) Mining; 3) Manufacturing; 4) Utilities; 5) Construction; 6) Wholesale and retail trade; 7) Transport and storage; 8) Information; 9) Finance, insurance, real estate and rental and leasing; 10) Services including professional and business services, educational services, health care, social assistance, arts, entertainment, recreation, accommodation and food services and other services; and 11) Government.

Concerning the euro area, we considered all the member countries as of 1 January 2001, namely, Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain. Data regarding annual real GDP has been collected from the *European Commission AMECO Database*.³ Concerning sectoral data for the euro area countries, we used the *EU KLEMS Database* provided by the Groningen Growth and Development Centre (GGDC, which is financially supported by the European Commission).⁴ As this data ranges from 1970 up to 2007, we updated it up to 2011 resorting to the *AMECO Database*.

The weights used in Equation [5] correspond to the share of output of the pair (i, j) in the year 2005.⁵ As usual, all series are taken in logs and first differenced. By considering growth rates, it is possible to assess how output synchronisation differs across frequencies, allowing

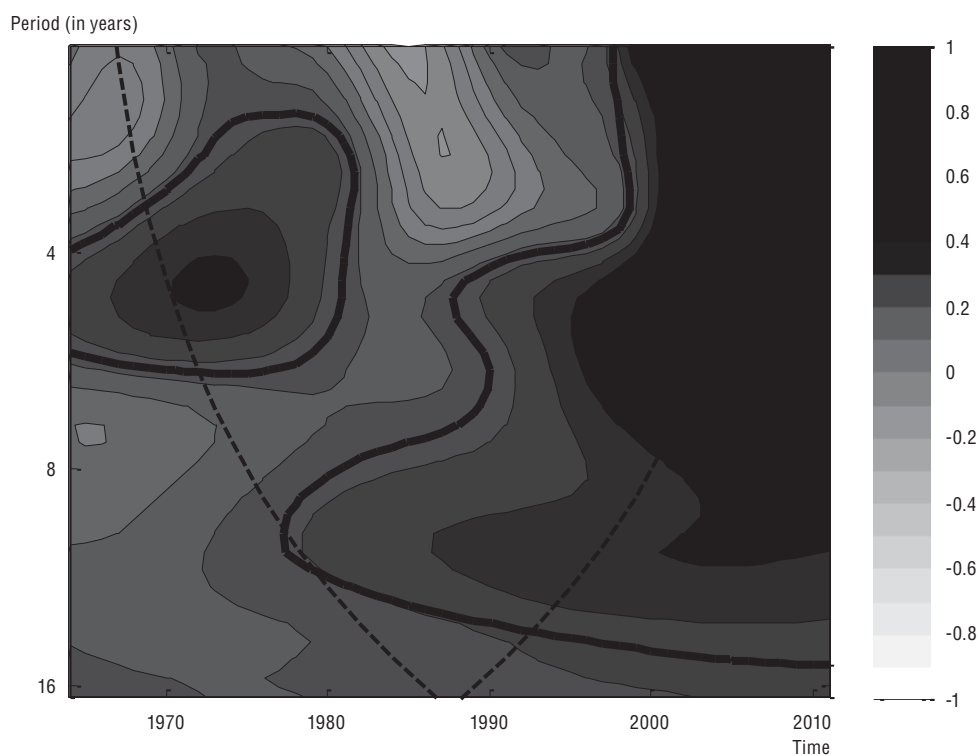
to assess both at the business cycle frequency range as well as at other frequencies. If one had used, for example, a band-pass filter the latter information would be lost.

4. Cohesion within euro area and US

4.1. Spatial cohesion

In this section, we proceed to compute the suggested measure so as to assess the cohesion among euro area countries and the cohesion within the US at both the regional and state levels. As three dimensions are involved, the wavelet-based cohesion is presented through a contour plot. The horizontal axis refers to time while the vertical axis refers to frequency. To ease interpretation, the frequency is converted to time units (years). The gray scale is for the wavelet-based cohesion, where increasing darkness corresponds to an increasing value and mimics the height in a surface plot. Hence, through the inspection of the graph one can identify both frequency bands (in the vertical axis) and time intervals (in the horizontal axis) where cohesion is larger and whether it has changed over time. As the continuous wavelet transform at a given point in time uses information of neighbouring data points, the values of the wavelet transform are generally less accurate as the wavelet approaches the edges of the time-series (this region is known as the cone of influence (see Torrence and Compo, 1998). With finite length time-series, edge effects occur at the beginning and end of the sample period, as the Fourier transform assumes data to be periodic. The region affected increases with the temporal support (or width) of the wavelet and to limit the wraparound effects one usually resorts to zero padding (see, for example, Aguiar-Conraria and Soares, 2013). Hence, the results should be read carefully close to the beginning or the end of the time series. The dashed line delimits the cone of influence. Note that we restricted the figures to periods up to around sixteen years which corresponds to the lowest frequency where there is at least some part outside the cone of influence. Concerning the statistical significance of cohesion, as the distribution is not known, the critical values for a 5% significance level were determined through a Monte Carlo simulation exercise.⁶ The bold line delimits the area where cohesion is statistically different from zero with a 95% confidence level.

In Figure 1, we present the wavelet-based measure of cohesion among euro area countries in terms of GDP growth. Several findings emerge. Firstly, one can observe that cohesion has been typically higher at low frequencies, i.e., long-run dynamics, than at the remaining frequencies. In this respect, Crowley and Mayes (2008) also found a higher synchronisation of output growth for longer cycles than for shorter ones when considering the three major euro area countries (Germany, France and Italy). Moreover, at low frequencies, cohesion has increased throughout time becoming statistically significant since the beginning of the 1980s. At the standard business cycle frequency range, that is fluctuations between two and eight years, cohesion was particularly high at the beginning of the 1970s, during the oil crisis, while decreasing during the 1980s. This lends support to Inklaar and de Haan's (2001) finding that correlations of euro area countries business cycle are higher in the period 1971-79 than 1979-87. However, cohesion within euro area has increased again since the mid-1990s which is in line with the results of Massmann and Mitchell (2004) obtained by computing pairwise correlations with rolling windows. Similarly, Altavilla (2004) presents evidence that synchronisation of several euro countries has increased after 1991, while Darvas and Szapáry (2008) also find evidence in support of more business cycle comovement in the euro area since 1993, the start of the run-up to the

Figure 1. **Cohesion within euro area**

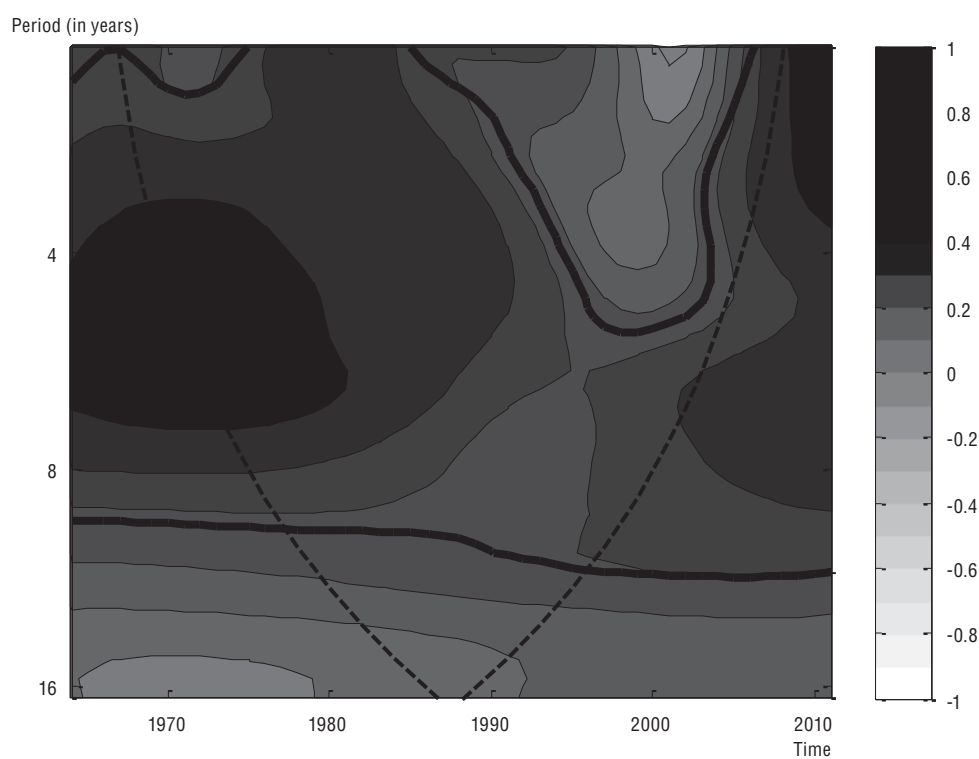
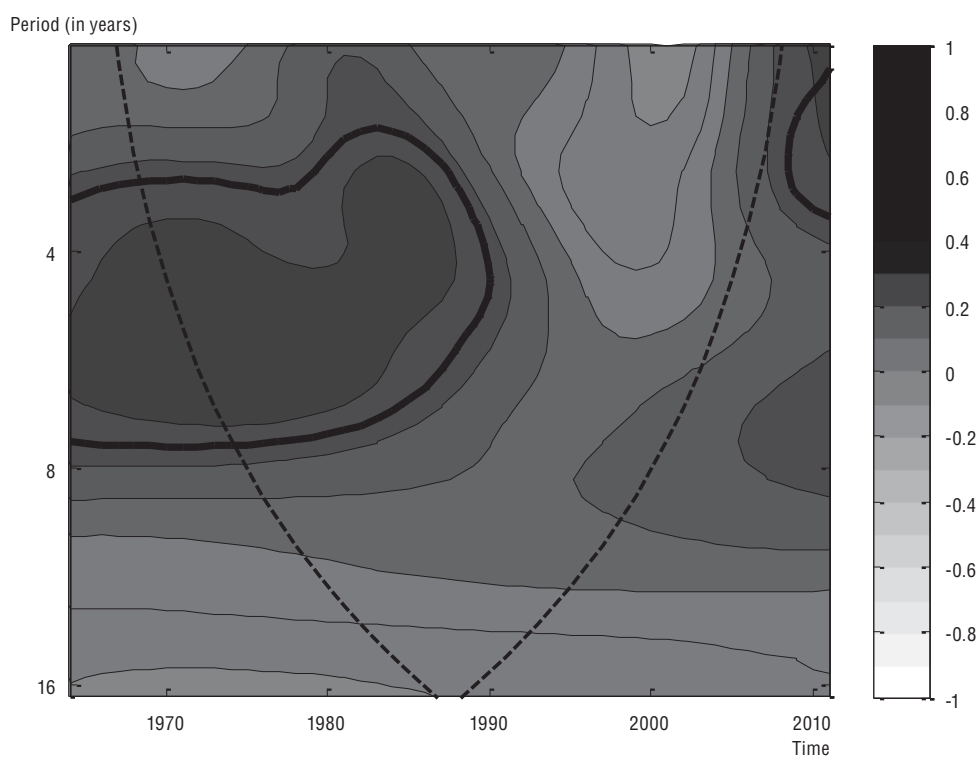
European Monetary Union. Finally, one should note that cohesion has been very high for most frequencies since the beginning of the euro area (see also Rua, 2010).

Regarding the US, we compute the cohesion at both the regional and state levels (see Figures 2 and 3, respectively). Likewise in the euro area case, cohesion is positive within the US at both the regional and state levels [Carlino and DeFina (2004) report similar evidence with US employment data]. Comparing Figures 2 and 3, one can observe that the US cohesion at the regional level is higher than at the state level, whatever the frequency and the time period. This is in line with the results of Croux et al. (2001), who argue that this is due to the fact that by aggregating the states the idiosyncratic sources of variations are diminished (see also Tootell, 1990, and Carlino and DeFina, 2004). The frequency range where cohesion is higher and statistically significant is at the business cycle frequency band. A noteworthy and distinct finding is that, while in the euro area cohesion seems to have increased, reflecting most probably the deepening of the process of European economic integration, in the US there seems to be evidence of a decrease in cohesion since the 80s. In this respect, Wyne and Koo (2000) found that the synchronisation between the 12 US Federal Reserve districts was higher in the run-up to the first oil shock and decreased during the 1980s. However, one should note that cohesion has increased again and attained a high level during the most recent economic and financial crisis episode.

To shed some more light on the above mentioned results, in Section 4.2 we conduct a cohesion analysis at the sectoral level.

4.2. Cohesion at the sectoral level

Up to now, the focus has been on assessing the spatial cohesion within the euro area and the US. To complement and provide further insights about the previous results, we

Figure 2. **Cohesion within US (at the regional level)**Figure 3. **Cohesion within US (at the state level)**

also investigate cohesion at the sectoral level. This analysis allows us to identify whether the above findings are broadly based or if they are being driven by any particular sector.

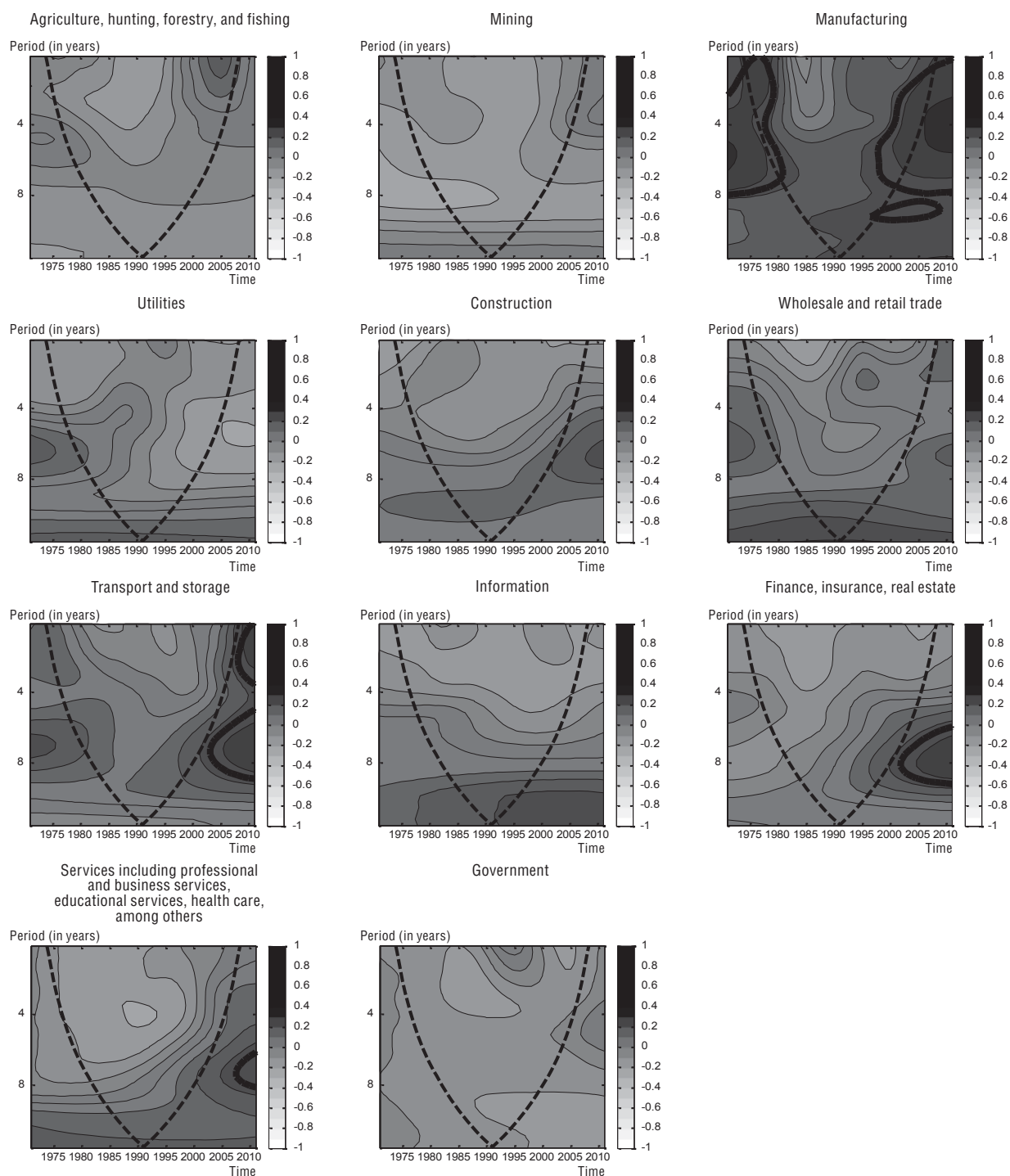
The results at the sectoral level for the euro area are reported in Figure 4. Each plot presents the cohesion among the euro area countries for a given sector. In general, sectors like Agriculture, Mining, Government and to a lesser extent, Utilities and Construction present relatively weak cohesion across frequencies and over time. In contrast, Manufacturing is the sector that presents the highest cohesion. Darvas and Szapáry (2008) also found higher synchronisation for industry than for services in the euro area. Sectors such as Manufacturing, Wholesale and retail trade, and Information denote a relatively high cohesion at low frequencies, which has increased throughout time, and seem to be responsible for the above mentioned time-varying behaviour of cohesion within the euro area at low frequencies. At the business cycle frequency range, one should note that in the case of Manufacturing, after presenting a high cohesion up to the 1980s, it recorded a decrease in the subsequent period. However, it increased again since the 1990s. Such increase was also observed in the sector related to Finance, insurance and real estate reflecting the financial integration that has been taking place in the euro area. Other sectors that have also recorded an increase in the more recent period are Transport and storage and Services including professional and business services and others.

We now turn to the US results (see Figures 5 and 6). As pointed out in Section 4.1, cohesion at the regional level is higher than at the state level and this finding seems to hold for all the sectors across all the frequencies and over time. Since the results at the regional and state levels are qualitatively similar, henceforth we focus on the results obtained at the state level. The results suggest a noteworthy heterogeneity in terms of cohesion across sectors. In this respect, Carlino and DeFina (2004) also found considerable variation in the degree of cohesion across sectors using employment data at the state level. Overall, the sectors that present the lowest cohesion are Mining, Finance, insurance and real estate, and Government whereas the highest cohesion is recorded by Manufacturing, Wholesale and retail trade, and Transport and storage. This is also broadly in line with the results of Carlino and DeFina (2004). In the sectors where cohesion has been higher, it is more marked at the business cycle frequency range. The exception is the Information sector where cohesion has been higher at low frequencies, namely in the first half of the sample period. At the latter part of the sample, cohesion in the Information sector has been statistically significant only at higher frequencies. In sectors like Construction and Agriculture, cohesion was high only during the 1970s and 1980s, whereas in the Utilities sector it was more pronounced during the last decade. In the cases of Manufacturing, Wholesale and retail trade, Transport and storage and Services including professional and business services and others cohesion was high up to the 1980s and then decreased progressively. These sectors appear to be behind the overall decrease of cohesion within the US at the business cycle frequency range which has been mentioned earlier. However, all these sectors present a noteworthy increase in the latter part of the sample period driving up again the cohesion within the US.

5. Conclusions

The assessment of output synchronisation across countries or regions has been of key importance in several strands of the literature as, for example, in the discussion of economic integration. Though comovement is traditionally measured in the time domain resorting to the well-known Pearson correlation coefficient, there has been an increasing focus on frequency domain analysis. However, while the former approach ignores the relationship at the frequency level, the latter disregards the fact that comovement can change over time. To overcome such shortcomings one can resort to wavelet analysis. In particular, we propose a wavelet-based

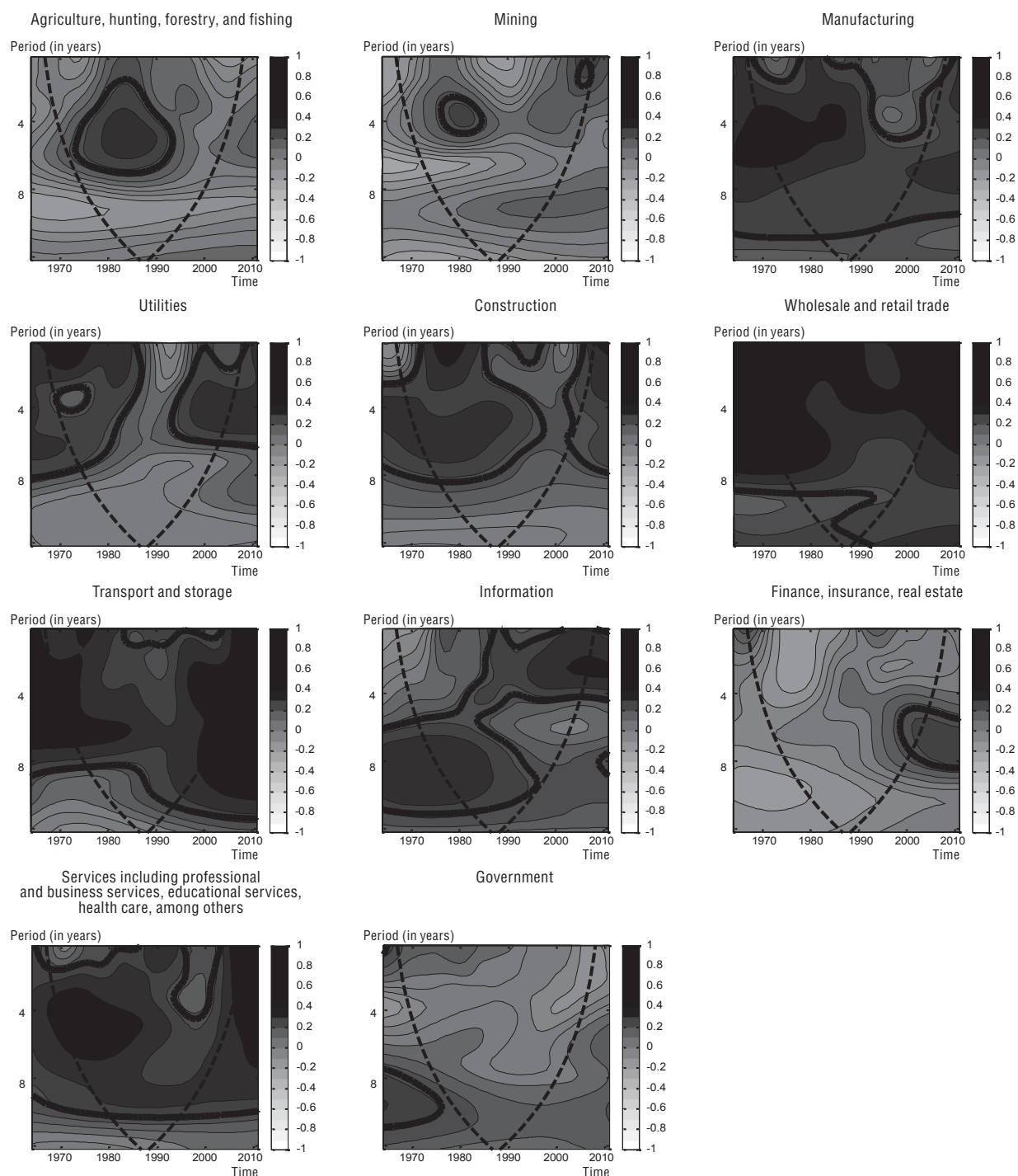
Figure 4. Cohesion across euro area countries by sector



measure of cohesion which allows one to assess how synchronisation has evolved over time and across frequencies simultaneously, within a set of countries or regions.

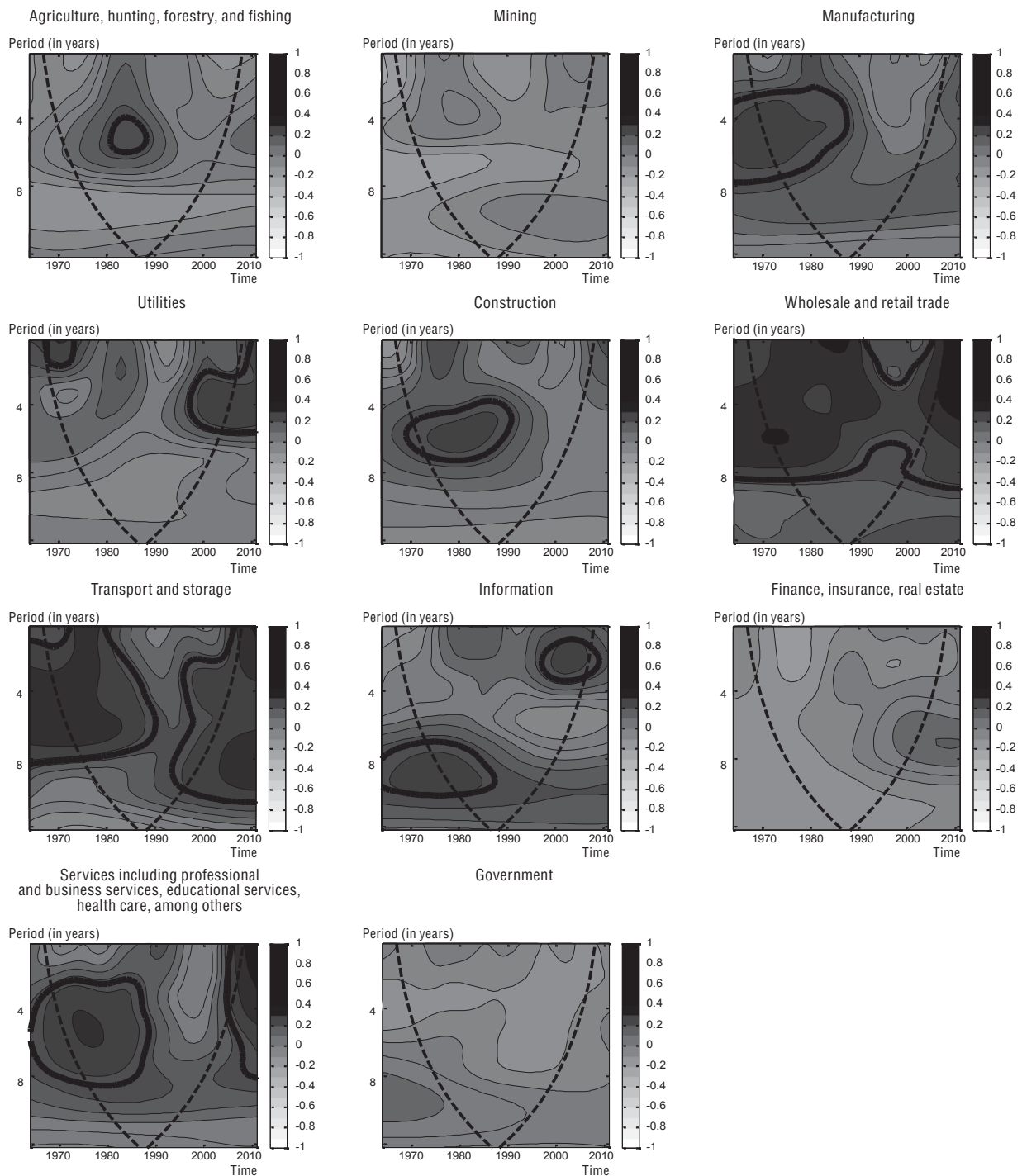
To illustrate its empirical application we focus on the output growth synchronisation within the euro area and the US. We study cohesion among the euro area countries and within the US, both at the regional and state levels over the last decades. The results obtained

Figure 5. Cohesion across US regions by sector



highlight the usefulness of a wavelet-based measure of cohesion so as to uncover both frequency and time-varying features. We find that cohesion within the euro area has been higher at low frequencies and that it has increased for most frequencies, being very high since the introduction of the euro. In the case of the US, cohesion has been higher at the business cycle frequency range. Despite the decrease observed during the 1980s, there was a

Figure 6. Cohesion across US states by sector



noteworthy increase during the most recent economic and financial crisis episode. These findings for the US hold both at the regional and state levels. Furthermore, we find that the US cohesion at the regional level is higher than at the state level across all frequencies and over the whole sample period. Additionally, we find a noteworthy heterogeneity in the results at the sectoral level. The sectors that seem to lie behind the overall results are identified.

Notes

1. For a general discussion about the relationship between scale and frequency, see, for example, Aguiar-Conraria and Soares (2013).
2. The eight BEA regions are: New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, Far West. The 51 BEA states are: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming. See www.bea.gov/regional.
3. See http://ec.europa.eu/economy_finance/db_indicators/ameco.
4. See www.euklems.net.
5. One should note that the results are not sensitive to the chosen year.
6. In particular, the critical values are obtained by considering 10 000 sets of autoregressive processes based on the series under analysis (see, for example, Torrence and Compo, 1998). In fact, we found that for most series the AR(1) model is the preferred one. All computations are done using Matlab and the codes are available from the authors upon request.

References

- Adisson, P. (2002), *The Illustrated Wavelet Transform Handbook*, The Institute of Physics, London.
- Aguiar-Conraria, L. and M.J. Soares (2013), "The Continuous Wavelet Transform: Moving Beyond Uni- and Bi-variate Analysis", *Journal of Economic Surveys*, Vol. 28, Iss. 2, 344-375, Doi: <http://dx.doi.org/10.1111/joes.12012>.
- Altavilla, C. (2004), "Do EMU Members Share the Same Business Cycle?", *Journal of Common Market Studies*, Vol. 42, No. 5, 869-896.
- Camacho, M., G. Perez-Quiros and L. Saiz (2006), "Are European Business Cycles Close Enough to Be Just One?", *Journal of Economic Dynamics and Control*, 30, 1687-1706.
- Carlino, G. and R. DeFina (2004), "How Strong Is Co-Movement in Employment Over the Business Cycle? Evidence from State/Sector Data", *Journal of Urban Economics*, 55, 298-315.
- Crone, T. (2005), "An Alternative Definition of Economic Regions in the United States Based on Similarities in State Business Cycles", *Review of Economics and Statistics*, 87, 617-626.
- Croux, C., M. Forni and L. Reichlin (2001), "A Measure of Comovement for Economic Variables: Theory and Empirics", *Review of Economics and Statistics*, 83, 232-241.
- Crowley, P. (2007), "A Guide to Wavelets for Economists", *Journal of Economic Surveys*, 21, 207-264.
- Crowley, P. and D. Mayes (2008), "How Fused Is the Euro Area Core? An Evaluation of Growth Cycle Co-Movement and Synchronization Using Wavelet Analysis", *Journal of Business Cycle Measurement and Analysis*, Vol. 2008, No. 1, 63-95, Doi: <http://dx.doi.org/10.1787/jbcm-a-v2008-1-en>.
- Darvas, Z. and G. Szapáry (2008), "Business Cycle Synchronization in the Enlarged EU", *Open Economies Review*, Vol. 19, No. 1, 1-19.
- Eickmeier, S. and J. Breitung (2006), "How Synchronized Are New EU Member States with the Euro Area? Evidence from a Structural Factor Model", *Journal of Comparative Economics*, 34, 538-563.
- Frankel, J. and A. Rose (1998), "The Endogeneity of the Optimum Currency Area Criterion", *Economic Journal*, 108, 1009-1025.
- Gallegati, M. and M. Gallegati (2007), "Wavelet Variance Analysis of Output in G7 Countries", *Studies in Nonlinear Dynamics and Econometrics*, Vol. 11, No. 3, Art. 6.
- Gallegati, M., A. Palestini and M. Petrin (2008), "Cyclical Behavior of Prices in the G7 Countries Through Wavelet Analysis", *Advances in Complex Systems*, Vol. 11, 1, 119-130.
- Haan, J. (de), R. Inklaar and R. Jong-A-Pin (2008), "Will Business Cycles in the Euro Area Converge: A Critical Survey of Empirical Research", *Journal of Economic Surveys*, Vol. 22, No. 2, 234-273.

- Inklaar, R. and J. de Haan (2001), "Is There Really a European Business Cycle? A Comment", *Oxford Economic Papers*, 53, 215-220.
- Kalemli-Ozcan, S., B. Sorensen and O. Yosha (2001), "Economic Integration, Industrial Specialization and the Asymmetry of Macroeconomic Fluctuations", *Journal of International Economics*, 55, 107-137.
- Krugman, P. (1993), "Lesson of Massachusetts for EMU", in F. Giavazzi and F. Torres (eds), *The Transition to Economic and Monetary Union in Europe*, Cambridge University Press, 241-261.
- Massmann, M. and J. Mitchell (2004), "Reconsidering the Evidence: Are Eurozone Business Cycles Converging?", *Journal of Business Cycle Measurement and Analysis*, Vol. 1, No. 3, 275-308, Doi: <http://dx.doi.org/10.1787/jbcm-a-v2004-3-en>.
- Mundell, R. (1961), "A Theory of Optimum Currency Areas", *American Economic Review*, 51, 657-665.
- Percival, D. and A. Walden (2000), *Wavelet Methods for Time Series Analysis*, Cambridge University Press.
- Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Vol. I and II, Academic Press, London.
- Rose, A. (2000), "One Money, One Market: Estimating the Effect of Common Currencies on Trade", *Economic Policy*, 30, 9-45.
- Rua, A. (2010), "Measuring Comovement in the Time-frequency Space", *Journal of Macroeconomics*, 32, 685-691.
- Rua, A. (2011), "A Wavelet Approach for Factor-Augmented Forecasting", *Journal of Forecasting*, 30, 666-678.
- Rua, A. (2012), "Money Growth and Inflation in the Euro Area: A Time-Frequency View", *Oxford Bulletin of Economics and Statistics*, 74, 875-885.
- Rua, A. and L.C. Nunes (2005), "Coincident and Leading Indicators for the Euro Area: A Frequency Band Approach", *International Journal of Forecasting*, 21, 503-523.
- Rua, A. and L.C. Nunes (2009), "International Comovement of Stock Market Returns: A Wavelet Analysis", *Journal of Empirical Finance*, 16, 632-639.
- Rua, A. and L.C. Nunes (2012), "A Wavelet-Based Assessment of Market Risk: The Emerging Markets Case", *Quarterly Review of Economics and Finance*, Vol. 52, 84-92.
- Tootell, G. (1990), "Central Bank Flexibility and the Drawbacks to Currency Unification", *New England Economic Review*, May/June, 3-17.
- Torrence, C. and G. Compo (1998), "A Practical Guide to Wavelet Analysis", *Bulletin of the American Meteorological Society*, Vol. 79, No. 1, 61-78.
- Tripier, F. (2002), "The Dynamic Correlation Between Growth and Unemployment", *Economics Bulletin*, Vol. 5, No. 4, 1-9.
- Wyne, M.A. and J. Koo (2000), "Business Cycles Under Monetary Union: A Comparison of the EU and US", *Economica*, 67, 347-374.
- Yogo, M. (2008), "Measuring Business Cycles: A Wavelet Analysis of Economic Time Series", *Economics Letters*, 100, 208-212.

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 2014. Without their devoted and honorary effort, this publication would not be feasible.

Abberger, Klaus
Abrahamsen, Yngve
Achuthan, Lakshman
Addo, Peter Martey
Altug, Sumru
André, Christophe
Antonakakis, Nikolaos
Askitas, Nikos
Bessec, Marie
Billio, Monica
Bracke, Philippe
Conflitti, Cristina
Fenga, Livio
Fonzo, Tommaso Di
Froni, Claudia
Gallegati, Marco

Garda, Paula
Golinelli, Roberto
Granda-Carvajal, Catalina
Hori, Masahiro
Jaba, Elisabeta
Kadilli, Anjeza
Kliesen, Kevin L.
Lahiri, Kajal
Li, Yushu
Ma, Xiaohan
Mallic, Debdulal
Mikosch, Heiner
Molnar, Krisztina
Pedersen, Michael
Ponomareva, Natalia
Proietti, Tommaso

Rua, António
Sandqvist, Pauliina
Sarferaz, Samad
Schneider, Friedrich
Seiler, Christian
Silva Lopes, Artur
Silva, José I.
Theobald, Thomas
Tongbin, Zhang
Uebele, Martin
Viren, Matti
Von der Lippe, Michael Peter
Wang, Peijie
Yilmazkuday, Hakan

Aims and scope

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.

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.

CCIRET is a global forum for economists and institutions that 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 (editor-in-chief@ciret.org). All papers undergo a referee process.

Notes for authors regarding manuscripts are available at www.ciret.org/jbcma or www.oecd.org/std/jbcma.

Correspondence about editorial and other matters related to the journal may also be addressed to the CIRET office (jbcma@ciret.org) or the OECD Statistics Directorate (stat.contact@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.

Contents

Tracking economic activity in the euro area: Multivariate direct filter approach

Ginters Buss

What's behind the survey values? An analysis of individual forecaster's behaviour

Maritta Paloviita and Matti Viren

On the robustness of the balance statistics with respect to non-response

Christian Seiler

Cohesion within the euro area and the US: A wavelet-based view

António Rua and Artur Silva Lopes

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/jbcma-v2014-2-en>.

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 2014/2

OECD publishing

www.oecd.org/publishing

ISSN 1995-2880
VOLUME 2014(2 ISSUES)
33 2014 02 1 P



9 771995 288001