A wavelet-based multivariate multiscale approach for forecasting

Working Papers 2016

António Rua

BANCO DE PORTUGAL
EUROSYSTEM

12

12

A wavelet-based multivariate multiscale approach for forecasting

Working Papers 2016

António Rua

Please address correspondence to
Banco de Portugal, Economics and Research Department
Av. Almirante Reis 71, 1150-012 Lisboa, Portugal
T +351 213 130 000 | estudos@bportugal.pt

BANCO DE PORTUGAL
EUROSYSTEM

Lisbon, 2016 • www.bportugal.pt

A wavelet-based multivariate multiscale approach for forecasting

António Rua

Banco de Portugal

NOVA School of Business and
Economics

July 2016

Abstract

In an increasingly data rich environment, factor models have become the workhorse approach for modelling and forecasting purposes. However, factors are non-observable and have to be estimated. In particular, the space spanned by the unknown factors is typically estimated via principal components. Herein, it is proposed a novel procedure to estimate the factor space resorting to a wavelet based multiscale principal component analysis. Through a Monte Carlo simulation study, it is shown that such an approach allows to improve both factor model estimation and forecasting performance. In the empirical application, one illustrates its usefulness for forecasting GDP growth and inflation in the United States.

---

1. Introduction

In an increasingly data-rich environment, factor models have become one of the most popular forecasting tools in the literature and among practitioners in central banks and international institutions

In fact, there is by now a huge strand of literature using factor models to forecast macroeconomic variables, namely GDP growth and inflation. See Stock and Watson (1999, 2002a), Banerjee and Marcellino (2006), Giannone, Reichlin and Small (2008) for the United States, Marcellino, Stock and Watson (2003), Camba-Méndez and Kapetanios (2005), Angelini *et al.* (2011) for the euro area, Artis, Banerjee and Marcellino (2005) for the United Kingdom, Barhoumi, Darné and Ferrara (2010, 2013) for France, Schumacher (2007, 2010) and Schumacher and Breitung (2008) for Germany, Runstler *et al.* (2009) for several european countries, among many others.

Typically, the factors in such forecasting models are estimated via principal components analysis. Such an approach draws on the work by Stock and Watson (1998, 2002b) and Bai and Ng (2002) who have shown that the principal components are consistent estimators of the true latent factors when the cross-section dimension and the number of observations tend to infinity. Moreover, feasible forecasts, constructed using the estimated factors and estimated parameters, are shown to be asymptotically efficient.

However, the finite sample performance of the principal components estimator worsens when the relative explanatory power of the factors decreases vis-à-vis the idiosyncratic components. Intuitively, when the relative importance of the idiosyncratic term increases, it becomes more difficult to distinguish the common from the idiosyncratic component. Based on Monte Carlo evidence, Boivin and Ng (2006) and Bai and Ng (2008) show that the estimated factors and forecasts are negatively affected by decreasing the relative importance of the common component. Such findings have been reinforced by the theoretical results concerning the finite sample properties of the principal components estimator, within a weakly influential factor asymptotics framework, as provided by Johnstone and Lu (2009) and Onatski (2012).

To cope with the above mentioned issue, we suggest a wavelet-based approach. Although wavelet analysis has been developed in other fields, such an approach has already proved to be useful in economics and finance. See, for instance, the pioneer work of Ramsey and Zhang (1996, 1997) and Ramsey and Lampart (1998a,b)). Recent applications of wavelets in the literature can

be found in Aguiar-Conraria and Soares (2011a, 2011b, 2014), Gallegati (2008, 2012, 2014), Gallegati *et al.* (2011, 2014), Rua (2010, 2011, 2012), Rua and Nunes (2009, 2012), among others (see Crowley (2007) for a survey).

In particular, we resort to a wavelet-based multiscale principal components analysis to improve factor model estimation and forecasting performance. Multiscale principal components analysis has been initially proposed by Bakshi (1998) for multivariate statistical process control. The multiscale principal components analysis is a generalisation of principal components analysis and involves decomposing each variable on a selected family of wavelets. A principal components analysis is conducted independently at each scale and combined in an efficient scale-recursive way to yield the multiscale model. Hence, multiscale principal components analysis harvest the benefits of both principal components and wavelet analysis. On the one hand, the relationship between the variables is decorrelated by principal components while, on the other hand, each variable is decorrelated by the wavelet decomposition.

Herein, we show how one can take advantage of the multiscale principal components analysis to enhance the estimation of the space spanned by the true latent factors and to improve factor model forecasting behaviour. We provide a Monte Carlo simulation study to assess the suggested approach and we find that it delivers noteworthy gains vis-à-vis the principal components estimator. Furthermore, to illustrate the empirical usefulness of the suggested wavelet-based multivariate multiscale principal components approach, we assess its performance for forecasting GDP growth and inflation in the United States. Drawing on the large dataset compiled by Stock and Watson (2012), we find that the proposed approach outperforms significantly the factor model advocated therein which relies on principal components.

The paper is organized as follows. In section 2, the wavelet multiscale decomposition is overviewed and the multiscale principal components analysis is discussed. In section 3, the design of the Monte Carlo exercise is described and the corresponding simulation results are reported. In section 4, the empirical application is conducted by forecasting GDP growth and inflation in the United States and the results of the out-of-sample forecasting exercise are presented. Finally, section 5 concludes.

2. A wavelet-based multiscale approach

2.1. From wavelets to multiscale decomposition

The term wavelet denotes a small wave. The wave refers to the condition that this function $\psi(.)$ is oscillatory. The smallness refers to the condition that it is of finite length, that is, compactly supported. Thus, a wavelet should satisfy two basic properties namely the integral of $\psi(.)$ is zero,

$$\int \psi(t)dt = 0 \qquad (1)$$

that is, the average value of the wavelet in the time domain must be zero; and the square of $\psi(.)$ integrates to unity,

$$\int \psi^2(t)dt = 1 \qquad (2)$$

which means that $\psi(.)$ is limited to an interval of time. Although it departs from zero for a limited interval of time, the excursions above zero must cancel out with the excursions below zero. Thus, a wavelet is any function that integrates to zero and is square integrable.[1]

The wavelet transform decomposes a time series in terms of some basis functions, the wavelets, analogous to the use of sines and cosines in Fourier analysis. Wavelets are a family of basis functions $\psi_{\tau,s}(t)$ that are localized in both time and frequency and are obtained by translation and dilation of the mother wavelet $\psi(t)$ as

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \qquad (3)$$

where τ determines the time position (translation parameter), s is the scale (dilation parameter) and $\frac{1}{\sqrt{s}}$ is for energy normalization across the different scales ($\|\psi_{\tau,s}\|^2 = \|\psi\|^2$).

---

1.   In addition, it should also satisfy the so-called admissibility condition,

$$0 < \int_0^{+\infty} \frac{\left|\widehat{\psi}(\omega)\right|^2}{\omega}d\omega < +\infty$$

where $\widehat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$, that is, $\widehat{\psi}(\omega) = \int_{-\infty}^{+\infty}\psi(t)e^{-i\omega\tau}dt$, so as to allow for the reconstruction of the time series without loss of information.

The Discrete Wavelet Transform (DWT) involves the discretization of the translation and dilation parameters (see, for example, Percival and Walden (2000)). In particular, these parameters are discretized dyadically as $s = 2^{-j}$ and $\tau = 2^{-j}k$. Any time series $x(t)$ can be decomposed as a weighted sum of dyadically discretized orthonormal basis functions as

$$x(t) = \sum_k d_{1,k}\psi_{1,k}(t) + \cdots + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k a_{J,k}\varphi_{J,k}(t) \qquad (4)$$

where J is the number scales, $d_{j,k}$ is the detail coefficient at scale j and location k whereas $a_{J,k}$ is the scaling function coefficient at the coarsest scale J and location k. The terms $\sum_k d_{j,k}\psi_{j,k}(t)$ for $j = 1, 2, ..., J$ and $\sum_k a_{J,k}\varphi_{J,k}(t)$ represent the detail and smooth components, respectively. The wavelet transform coefficients are given by[2]

$$d_{j,k} = \int x(t)\psi_{j,k}(t)dt \qquad j = 1, 2, ..., J \qquad (5)$$

$$a_{J,k} = \int x(t)\varphi_{J,k}(t)dt \qquad (6)$$

where

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right) \qquad j = 1, 2, ..., J \qquad (7)$$

$$\varphi_{J,k}(t) = 2^{-J/2}\varphi\left(2^{-J}t - k\right) \qquad (8)$$

The scaling function or father wavelet $\varphi(.)$ captures the low frequency content of the series that is not captured by wavelets at the corresponding or finer scales.

For a given familiy of wavelets, efficient methods to compute the wavelet decomposition are based on the convolution of the time series with the corresponding wavelet filter $\mathbf{H}$ and scaling filter $\mathbf{G}$ (see, for example, Percival and Walden (2000)). Thus the coefficients can be obtained by

$$\mathbf{d}_j = \mathbf{H}_j\mathbf{x} \qquad (9)$$

$$\mathbf{a}_j = \mathbf{G}_j\mathbf{x} \qquad (10)$$

---

2.   When the number of observations, T, is divisible by 2^J there are $T/2^j$ $d_{j,k}$ coefficients at scale $j = 1, ..., J - 1$, while at scale J there are $T/2^J$ $d_{J,k}$ coefficients and $T/2^J$ $a_{J,k}$ coefficients. In total, there are T wavelet coefficients, that is, $T = T/2^1 + T/2^2 + ... + T/2^{J-1} + T/2^J + T/2^J$.

where $\mathbf{x}$ denotes the data vector, $\mathbf{d}_j$ is the vector of detail coefficients at scale j, $\mathbf{H}_j$ is obtained by applying the $\mathbf{G}$ filter $j-1$ times and the filter $\mathbf{H}$ once, $\mathbf{a}_j$ is the vector of scaling coefficients at scale j and $\mathbf{G}_j$ is obtained by applying the $\mathbf{G}$ filter j times. Hence, the DWT is implemented via the Mallat pyramid algorithm where in the first stage it consists in transforming $\mathbf{x}$ into first level wavelet coefficients $\mathbf{d}_1$ and first level scaling coefficients $\mathbf{a}_1$. In the second stage, it transforms the vector $\mathbf{a}_1$ into the second level details $\mathbf{d}_2$ and second level scaling coefficients $\mathbf{a}_2$ and so on. This means that at the jth stage (for $j = 2, ..., J$) the vector $\mathbf{a}_{j-1}$ is transformed, likewise $\mathbf{x}$ in the first stage, by applying the filters $\mathbf{H}$ and $\mathbf{G}$ to obtain $\mathbf{d}_j$ and $\mathbf{a}_j$. At the end of the Jth stage one can form the DWT coefficient vector as

$$\mathbf{W} = \begin{bmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_J \\ \mathbf{a}_J \end{bmatrix} \tag{11}$$

One can also write

$$\mathbf{W} = \mathcal{W}\mathbf{x} \tag{12}$$

where $\mathcal{W}$ is $T \times T$ real valued orthonormal matrix defining the DWT and satisfying $\mathcal{W}'\mathcal{W} = I$ ($T \times T$ identity matrix). Hence, from (9)-(12),

$$\mathcal{W} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_J \\ \mathbf{G}_J \end{bmatrix} \tag{13}$$

Based on (12) one can write

$$\mathbf{x} = \mathcal{W}'\mathbf{W} \tag{14}$$

and drawing on the orthonormality of $\mathcal{W}$ one can show that

$$
\begin{aligned}
\|\mathbf{x}\|^2 &= \mathbf{x}'\mathbf{x} \\
&= \left(\mathcal{W}'\mathbf{W}\right)'\left(\mathcal{W}'\mathbf{W}\right) \\
&= \left(\mathbf{H}_1'\mathbf{d}_1 + ... + \mathbf{H}_J'\mathbf{d}_J + \mathbf{G}_J'\mathbf{a}_J\right)'\left(\mathbf{H}_1'\mathbf{d}_1 + ... + \mathbf{H}_J'\mathbf{d}_J + \mathbf{G}_J'\mathbf{a}_J\right) \\
&= \left(\mathbf{d}_1'\mathbf{H}_1 + ... + \mathbf{d}_J'\mathbf{H}_J + \mathbf{a}_J'\mathbf{G}_J\right)\left(\mathbf{H}_1'\mathbf{d}_1 + ... + \mathbf{H}_J'\mathbf{d}_J + \mathbf{G}_J'\mathbf{a}_J\right) \\
&= \mathbf{d}_1'\mathbf{d}_1 + ... + \mathbf{d}_J'\mathbf{d}_J + \mathbf{a}_J'\mathbf{a}_J \\
&= \|\mathbf{d}_1\|^2 + ... + \|\mathbf{d}_J\|^2 + \|\mathbf{a}_J\|^2
\end{aligned}
\tag{15}
$$

Equation (15) implies that the DWT is a energy preserving transformation allowing to decompose the variance of $\mathbf{x}$ on a scale-by-scale basis. This is a key a feature which has been exploited by Fan and Gençay (2010) to distinguish between a white noise and a unit root process. As discussed by Fan and Gençay (2010), a white noise process has more energy at the lowest scale detail coefficients $\mathbf{d}_1$ while declining towards the highest scale. In contrast, in the case of a unit root process the energy is basically concentrated in the scaling coefficients $\mathbf{a}_J$. Hence, a more (less) persistent series has more (less) energy at higher scales and less (more) at lower scales. We will also take advantage of this feature here and show how it can be used to improve factor model estimation.

2.2. Multiscale principal components analysis

In the previous section, we considered the case of a single series $\mathbf{x}$. Now, let $\mathbf{X}$ denote a data matrix $T \times N$. Suppose that one applies the DWT to $\mathbf{X}$, that is, $\mathcal{W}\mathbf{X}$. Firstly, one should note that the variance-covariance matrix of $\mathcal{W}\mathbf{X}$ is the same as that of $\mathbf{X}$ since

$$
\|\mathcal{W}\mathbf{X}\|^2 = \left(\mathcal{W}\mathbf{X}\right)'\mathcal{W}\mathbf{X} = \mathbf{X}'\mathcal{W}'\mathcal{W}\mathbf{X} = \mathbf{X}'\mathbf{X} = \|\mathbf{X}\|^2
\tag{16}
$$

An important implication of this result is that for principal component analysis, focusing on $\mathcal{W}\mathbf{X}$ or $\mathbf{X}$ does not change the analysis. In fact, (16) implies that the loadings obtained with principal components of $\mathbf{X}$ and $\mathcal{W}\mathbf{X}$ are identical. Moreover, the principal components of $\mathcal{W}\mathbf{X}$ are the wavelet transform of the the principal components of $\mathbf{X}$. To see this, let $\mathbf{Z}$ be the principal components matrix of $\mathbf{X}$ and $\mathbf{A}$ the corresponding loadings matrix, then

$$
\mathbf{X} = \mathbf{Z}\mathbf{A}' \Leftrightarrow \mathcal{W}\mathbf{X} = (\mathcal{W}\mathbf{Z})\mathbf{A}'
\tag{17}
$$

By extending the proof of (15) to the multivariate case and making use of (9) and (10) it can also be shown that

$$\|\mathcal{W}\mathbf{X}\|^2 = \|\mathbf{H}_1\mathbf{X}\|^2 + \ldots + \|\mathbf{H}_J\mathbf{X}\|^2 + \|\mathbf{G}_J\mathbf{X}\|^2 \qquad (18)$$

The above results are key to the subsequent analysis. Not only the variance-covariance of the wavelet transform is the same of the original data matrix but one can also decompose it terms of the contribution at multiple scales. Such a result allows one to conduct a principal component analysis at each scale independently of the other scales.

This leads to the so-called multiscale principal component analysis (as initially proposed by Bakshi (1998)). It consists in the following steps. First, the DWT is performed for each column of $\mathbf{X}$ which produces the matrices $\mathbf{D}_1,...,\mathbf{D}_J$ containing the detail coefficients and $\mathbf{A}_J$ containing the scaling coefficients. For each scale, one selects the appropriate number of principal components or surpress the detail j. Then, reconstruct a new matrix $\widetilde{\mathbf{X}}$ containing the main features of the original matrix $\mathbf{X}$. Finally, perform the principal component analysis of $\widetilde{\mathbf{X}}$. Note that, since, in general, $rank(\widetilde{\mathbf{X}}) = rank(\mathbf{X})$, the last step is needed to reduce dimensionality. Moreover, if no dimension reduction is performed at any scale, that is, all principal components are retained at each scale, then $\widetilde{\mathbf{X}}$ will be identical to $\mathbf{X}$.

To show how can multiscale principal component analysis enhance factor model estimation let us consider the following factor structure in matrix form as[3]

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{e} \qquad (19)$$

where $\mathbf{F}$ is the $T \times r$ matrix of non-observable factors, $\mathbf{\Lambda}$ is the $N \times r$ matrix of (unknown) factor loadings and $\mathbf{e}$ is the $T \times N$ matrix of idiosyncratic errors. When both $N \to \infty$ and $T \to \infty$, Stock and Watson (1998, 2002b), Bai and Ng (2002), Bai (2003) and Amengual and Watson (2007) have shown that, under slightly different sets of assumptions regarding the data generating processes of

---

3.   Without loss of generality, we consider the factor model within the static framework. A dynamic factor model with q factors can be written as a static factor model with r factors, where r is finite. Key results regarding dynamic factor models with large datasets can be found in Forni *et al.* (2000, 2004, 2005).

the factors and the idiosyncratic components[4], the space spanned by the true factors can be consistently estimated via the first principal components.

However, the finite sample performance of the principal components estimator deteriorates when the explanatory power of the factors decreases vis-à-vis the explanatory power of the idiosyncratic errors. Based on a Monte Carlo analysis, Boivin and Ng (2006) and Bai and Ng (2008), show that the factor estimates and forecasts are adversely affected by decreasing the relative importance of the common component. Intuitively, when the importance of the idiosyncratic error is magnified, it becomes harder to disentangle the common from the idiosyncratic component in the data. This issue has been futher investigated within a weakly influential factor asymptotics framework to assess the finite sample properties of the principal components estimator in a context of a relatively weak explanatory power of the factors. Johnstone and Lu (2009) show the inconsistency of the principal components estimator for the one-factor model with *i.i.d* Gaussian factor and *i.i.d.* Gaussian idiosyncratic terms while Onatski (2012) extends the findings to more general dynamics.

The above discussion suggests that a possible way to improve the performance of the principal components estimator consists in downweighting the noise. As mentioned earlier, when discussing the multiscale decomposition of a time series, the noise tends to be reflected, to a larger extent, at the lowest scale. On the other hand, factors are typically smooth and are therefore better captured by the scaling function. Hence, to improve factor model estimation, we suggest to surpress the lowest detail components of $\mathbf{X}$ while retaining the first principal components extracted at higher scales. In other words, since at the lowest scale the signal-to-noise ratio is expected to be very low one can surpress it without loosing too much information while avoiding the problem of estimating via principal components in a context of weak factors. The data at the lowest scale should be basically uniformative for the factor structure. At higher scales, by retaining the first principal components on scale-by-scale basis it allows to be more sensitive to scale-varying signal features and can potentially enhance factor model estimation and forecasting.

---

4.   The typical assumptions allow for some heteroskedasticity and limited dependence of the idiosyncratic components in both the time and cross-section dimensions, as well as for moderate correlation between the latter and the factors.

3. A Monte Carlo study

In this section, we conduct a Monte Carlo simulation study to support the above suggested approach.

3.1. The setup

Let us define a factor model with a data generating process given by

$$x_{it} = \sum_{j=1}^{r} \Lambda_{ij} f_{jt} + e_{it} \tag{20}$$

with $i = 1, ..., N$ and $t = 1, ..., T$ or simply in vector notation

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + \mathbf{e}_t \tag{21}$$

for $t = 1, ..., T$,

$$A(L)\mathbf{F}_t = \mathbf{u}_t \tag{22}$$

with $\mathbf{u}_t$ i.i.d. $\mathcal{N}(0, I_r)$,

$$D(L)\mathbf{e}_t = \mathbf{v}_t \tag{23}$$

with $\mathbf{v}_t$ i.i.d. $\mathcal{N}(0, \mathcal{T})$,

$$A_{ij}(L) = \begin{cases} 1 - a_1 L - a_2 L^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for } i, j = 1, ..., r \tag{24}$$

where $a_1 = 2b\cos(\varphi)$ with $\varphi \in (0, \pi)$ and $a_2 = -b^2$,

$$D_{ij} = \begin{cases} 1 - dL & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for } i, j = 1, ..., N \tag{25}$$

$$\Lambda_{ij} \ \text{i.i.d.} \ \mathcal{N}(0, 1) \tag{26}$$

for $i = 1, ..., N$, $j = 1, ..., r$,

$$\mathcal{T}_{ij} = \tau^{|i-j|} \left(1 - d^2\right) \sqrt{\alpha_i \alpha_j} \tag{27}$$

for $i, j = 1, ..., N$ and

$$\alpha_i = \frac{\beta_i}{1-\beta_i} \frac{1-a_2}{(1+a_2)\left[(1-a_2)^2 - a_1^2\right]} \sum_{j=1}^{r} \Lambda_{ij}^2 \tag{28}$$

A similar model has been used, for example, in Stock and Watson (2002b), Doz *et al.* (2012) and Pinheiro *et al.* (2013). Note that, herein we generate the factors as AR(2) processes in such a way that the resulting factors may display a cyclical behaviour (see, for example, Bierens (2001) and Castro *et al.* (2013)). This is motivated by the body of empirical literature where factors extracted from macroeconomic datasets are used as business cycle indicators (see the seminal work of Stock and Watson (1989, 1991, 1993) as well as Forni *et al.* (2001), Altissimo et *al.* (2001), Rua and Nunes (2005), Valle e Azevedo *et al.* (2006), Altissimo *et al.* (2010) among others).

In particular, (24) generates factors with a cyclical pattern of $2\pi/\varphi$ time periods with parameter b controlling for the persistence. In the simulations, we set φ equal to $2\pi/23$ so as to generate a cyclical pattern with the average duration of post-war US business cycles according to NBER business cycle dating committee, that is, 23 quarters. As in Doz *et al.* (2012) simulation study, we considered a smooth process for the factors and set b equal to 0.9.

As regards the idiosyncratic components, the model allows a limited cross-correlation as $\mathcal{T}$ is a Toeplitz matrix with the parameter τ controlling for the degree of cross-correlation. The dynamics of the idiosyncratic components is governed by an AR(1) process with autoregressive parameter d. For the time being, we set $\tau = 0$ and $d = 0$ (see also Doz *et al.* (2012)).

The parameter β_i corresponds to the ratio between the variance of the idiosyncratic component e_{it} and the variance of the corresponding variable x_{it}. This parameter controls for the relative importance between the common and the idiosyncratic components. As found for the US dataset considered by Stock and Watson (2012), we set this ratio equal to 0.6 with the number of factors r equal to 5.

In the case of the wavelet-based multiscale approach, we consider a Daubechies wavelet filter of length 4 and $J = 1$ as in Fan and Gençay (2010). Hence, we obtain two components namely a detail and a smooth component, and corresponding coefficients, $\mathbf{D}_1$ and $\mathbf{A}_1$. As mentioned earlier, the suggested approach involves surpressing the lowest detail while computing the first r principal components for the highest scale and finally retaining r factors from the corresponding reconstructed data matrix.

Several measures are computed for the comparison between the standard principal components estimator and the multiscale approach and the results are based on 1000 sample draws. Following Stock and Watson (2002b), Doz *et al.* (2012) and Pinheiro *et al.* (2013) we compute the trace R^2 as

$$R^2_{\mathbf{F}, \widehat{\mathbf{F}}} = \frac{\widehat{E}\left[tr\left(\mathbf{F}'\widehat{\mathbf{F}}\left(\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\right)^{-1}\widehat{\mathbf{F}}'\mathbf{F}\right)\right]}{\widehat{E}\left[tr\left(\mathbf{F}'\mathbf{F}\right)\right]} \tag{29}$$

where $\widehat{E}\left[.\right]$ denotes the expectation estimated by averaging the relevant statistic over the 1000 draws and $\widehat{\mathbf{F}}$ are the estimated factors. This statistic is a measure of fit of the multivariate regression of the true factors on the estimated factors, and is commonly used because the factors are identified only up to a rotation. A value higher and closer to one denotes a better estimation of the space spanned by the true factors.

Besides assessing the improvement in terms of the estimation of the factor space, we also evaluate the potential forecasting gains that might result from the suggested wavelet-based multiscale approach. Hence, similarly to Stock and Watson (2002b), we consider the scalar variable to be forecasted as generated by

$$y_{t+1} = \beta'\mathbf{F}_t + \varepsilon_{t+1} \tag{30}$$

where β is a $r \times 1$ vector of 1s and ε_{t+1} *i.i.d.* $\mathcal{N}(0,1)$ and independent of the other errors in the above specified factor model. The out-of-sample forecast is given by $\widehat{y}_{t+1|T} = \sum_{j=1}^{r} \widehat{\beta}_j \widehat{f}_{jt}$ where $\widehat{\beta}$ are the OLS coefficients in the regression of y_{t+1} onto $\widehat{f}_{jt}$, $j = 1, ..., r$, $t = 1, ..., T-1$ and $\widehat{f}_{jt}$ denote the estimated factors either via principal components or via the wavelet-based multiscale approach. The mean squared forecast error (MSFE) is computed for both approaches.

3.2. Simulation results

Let us consider different sizes for the cross-section and sample length (N, $T = 25, 50, 75, 100, 150, 200$). The simulation results are presented in Table 1. In particular, we report the trace R^2 for the wavelet-based multiscale principal components (R^2_{WMSPC}), the relative trace R^2 (R^2_{WMSPC}/R^2_{PC}) and the relative MSFE ($MSFE_{WMSPC}/MSFE_{PC}$) between the two approaches.

	$N = 25$	$N = 50$	$N = 75$	$N = 100$	$N = 150$	$N = 200$
			R^2_{WMSPC}			
$T = 25$	0.72	0.75	0.77	0.77	0.78	0.77
$T = 50$	0.79	0.84	0.86	0.87	0.88	0.88
$T = 75$	0.81	0.87	0.89	0.90	0.91	0.92
$T = 100$	0.82	0.88	0.91	0.92	0.93	0.93
$T = 150$	0.83	0.90	0.92	0.93	0.94	0.95
$T = 200$	0.84	0.90	0.93	0.94	0.95	0.96
			R^2_{WMSPC}/R^2_{PC}			
$T = 25$	1.09	1.05	1.04	1.03	1.02	1.01
$T = 50$	1.12	1.07	1.05	1.03	1.02	1.02
$T = 75$	1.13	1.07	1.05	1.03	1.02	1.02
$T = 100$	1.14	1.07	1.05	1.03	1.02	1.02
$T = 150$	1.14	1.07	1.05	1.03	1.02	1.02
$T = 200$	1.14	1.07	1.05	1.03	1.02	1.02
			$MSFE_{WMSPC}/MSFE_{PC}$			
$T = 25$	0.69	0.82	0.84	0.86	0.86	0.90
$T = 50$	0.76	0.78	0.84	0.87	0.91	0.92
$T = 75$	0.72	0.78	0.82	0.84	0.90	0.94
$T = 100$	0.77	0.82	0.86	0.91	0.94	0.95
$T = 150$	0.79	0.83	0.84	0.90	0.93	0.91
$T = 200$	0.77	0.83	0.87	0.89	0.95	0.96

TABLE 1. Monte Carlo results

Firstly, as the size of the cross-section and sample length increase, the trace R^2 for the wavelet-based multiscale principal components increases while approaching one when both N and T are large. Even for small N and T, the trace R^2 is relatively high. More importantly, the trace R^2 of wavelet-based multiscale approach is always higher than that of principal components for all N and T considered, as R^2_{WMSPC}/R^2_{PC} is always above one. That is, the wavelet-based multiscale approach delivers a better estimation of the space spanned by the true factors whatever the size of the cross-section and sample length. One should note that the gains vary depending on N and T. In particular, the relative improvement over the principal components estimator, in terms of the trace R^2, is higher for small N attaining up to 14 per cent. As one increases N the gains decrease but even for very large N and T one still records a gain of around 2 per cent.

On top of the previous findings, the wavelet-based multiscale approach also leads to noteworthy forecasting gains vis-à-vis principal components with the ratio $MSFE_{WMSPC}/MSFE_{PC}$ being below one for all N and T. In particular,

the forecasting improvement is higher for smaller datasets with the gains reaching more than 30 per cent. Note that, even for relatively large datasets the gains are substantial ranging from more than 20 per cent up to 10 per cent in most cases. Overall, the simulation results highlight the usefulness of the suggested approach for the estimation of the factor space and for factor-augmented forecasting.

	$N=25$	$N=50$	$N=75$	$N=100$	$N=150$	$N=200$
			R^2_{WMSPC}			
$T=25$	0.75	0.76	0.77	0.77	0.78	0.78
$T=50$	0.81	0.85	0.87	0.88	0.88	0.88
$T=75$	0.83	0.89	0.90	0.91	0.92	0.92
$T=100$	0.84	0.90	0.92	0.93	0.93	0.94
$T=150$	0.85	0.92	0.94	0.94	0.95	0.96
$T=200$	0.86	0.92	0.94	0.95	0.96	0.96
			R^2_{WMSPC}/R^2_{PC}			
$T=25$	1.07	1.04	1.03	1.02	1.01	1.01
$T=50$	1.10	1.05	1.03	1.02	1.01	1.01
$T=75$	1.12	1.06	1.03	1.02	1.01	1.01
$T=100$	1.12	1.05	1.03	1.02	1.01	1.01
$T=150$	1.13	1.05	1.03	1.02	1.01	1.01
$T=200$	1.13	1.05	1.03	1.02	1.01	1.01
			$MSFE_{WMSPC}/MSFE_{PC}$			
$T=25$	0.75	0.83	0.88	0.88	0.89	0.89
$T=50$	0.79	0.84	0.87	0.87	0.91	0.94
$T=75$	0.72	0.75	0.87	0.90	0.89	0.97
$T=100$	0.78	0.85	0.89	0.90	0.96	0.95
$T=150$	0.78	0.85	0.89	0.91	0.95	0.97
$T=200$	0.77	0.86	0.89	0.92	0.95	0.97

TABLE 2. Monte Carlo results with heteroskedastic idiosyncratic components

As a sensitivity analysis, we relax the assumption of homoskedastic idiosyncratic components. As in Doz *et al.* (2012) and Pinheiro *et al.* (2013), we consider heteroskedastic idiosyncratic components by setting β_i *i.i.d.* $\mathcal{U}([0.1, 0.9])$. From the results presented in Table 2, one can conclude that the above findings are basically unchanged. In Table 3, we present the simulation results allowing for moderate serial correlation and cross-correlation among idiosyncratic components by setting $\tau = 0.5$ and $d = 0.5$ (see Doz *et al.* (2012)) departing from the exact factor model. Although, in general, the trace R^2 is slightly lower and both the estimation and forecasting gains are smaller,

the wavelet-based approach continues to outperfom the principal components approach.

	$N = 25$	$N = 50$	$N = 75$	$N = 100$	$N = 150$	$N = 200$
			R^2_{WMSPC}			
$T = 25$	0.73	0.76	0.77	0.77	0.78	0.77
$T = 50$	0.76	0.82	0.84	0.85	0.87	0.87
$T = 75$	0.75	0.83	0.86	0.88	0.90	0.91
$T = 100$	0.73	0.83	0.87	0.89	0.91	0.92
$T = 150$	0.72	0.84	0.89	0.91	0.93	0.94
$T = 200$	0.72	0.85	0.89	0.92	0.94	0.94
			R^2_{WMSPC}/R^2_{PC}			
$T = 25$	1.04	1.02	1.01	1.01	1.00	1.00
$T = 50$	1.05	1.03	1.02	1.01	1.01	1.00
$T = 75$	1.06	1.03	1.02	1.01	1.01	1.00
$T = 100$	1.07	1.04	1.02	1.01	1.01	1.01
$T = 150$	1.08	1.04	1.02	1.01	1.01	1.01
$T = 200$	1.09	1.04	1.02	1.02	1.01	1.01
			$MSFE_{WMSPC}/MSFE_{PC}$			
$T = 25$	0.81	0.95	0.98	0.97	0.99	1.01
$T = 50$	0.93	0.91	0.95	0.96	0.99	0.97
$T = 75$	0.88	0.89	0.94	0.93	0.96	0.99
$T = 100$	0.89	0.93	0.94	0.97	0.97	0.98
$T = 150$	0.89	0.93	0.92	0.95	0.97	0.97
$T = 200$	0.86	0.91	0.92	0.94	0.97	0.98

TABLE 3. Monte Carlo results with serial and cross-correlation in the idiosyncratic components

4. Forecasting US GDP growth and inflation

To illustrate the empirical usefulness of the suggested wavelet-based multivariate multiscale principal components approach, we evaluate its performance for forecasting GDP growth and inflation in the United States.

4.1. Data

In particular, we resort to the large dataset of Stock and Watson (2012) which comprises 143 quarterly series for the United States spanning 49 years, from 1960 up to 2008. This macro dataset includes GDP and its

components, industrial production and capacity utilization, employment and hours, unemployment, housing starts, inventories and new orders, consumer prices and commodity prices, hourly earnings and unit labor costs, interest rates and spreads, monetary aggregates, exchange rates, stock prices and consumer expectations. As described in Stock and Watson (2012), the series are transformed by taking logarithms and/or differencing. In particular, first differences of logarithms are used for real variables, first differences are used for nominal interest rates and second differences of logarithms for prices. See Stock and Watson (2012) for a detailed list and further details on the series.

4.2. Design of the exercise

Based on the above mentioned dataset, Stock and Watson (2012) conduct an empirical comparison of various forecasting methods designed for a large number of predictors. They find that factor model forecasts using the first five principal components as predictors outperform all the alternative methods. Hence, this evidence suggests setting such a forecasting model as the benchmark to beat.

The factor model to be considered to forecast y is based on the least squares estimation of equation

$$y_{t+h} = \beta_0 + \sum_{j=1}^{r} \beta_j \widehat{f}_{jt} + \sum_{i=1}^{p} \gamma_i y_{t+1-i} + \varepsilon_{t+h} \quad (t = p, \cdots, T - h) \qquad (31)$$

where y_{t+h} denote the variable to be forecasted with a forecasting horizon of h periods. In the case of GDP, y_{t+h} is the h-period growth whereas for inflation, y_{t+h} is the h-period change in inflation. Regarding the factors, we set r equal to five as in Stock and Watson (2012) with $\widehat{f}_{jt}$ denoting the estimated factors either via principal components or via the wavelet-based multiscale approach.[5] As usual, the number p of autoregressive terms is determined by the standard BIC criterion.

We consider several forecasting horizons namely $h = 1, ..., 8$ and the out-of-sample period runs from 1985Q1 up to 2008Q4 as in Stock and Watson (2012)

---

5.   As mentioned in Stock and Watson (2012), of the 143 series in the dataset, 34 are high-level aggregates that are related by an identity to subaggregates in the dataset. Since including these higher-level aggregates does not add information, we also only use the 109 lower-level disaggregated series to compute principal components.

which corresponds to about half of the sample size. The h-step ahead forecasts for the variable y are obtained via a recursive forecasting exercise with recursive factor estimation, parameter estimation, model selection, and so forth.

The forecasting performance is evaluated by comparing the MSFE. In particular, we report the relative MSFE ($MSFE_{WMSPC}/MSFE_{PC}$) between the two approaches. To reinforce the findings, we assess the statistical significance of the forecasting gains by computing the Harvey *et al.* (1997) modified version of the Diebold and Mariano (1995) test.

4.3. Empirical results

In Table 4, we present the results for GDP growth and for both inflation and core inflation, based on the CPI all items and CPI excluding food and energy, respectively. The entries in the table correspond to the relative MSFE between the wavelet multiscale principal components approach and the standard principal components. Hence, a ratio lower than one means that the former outperforms the latter. In the case of the wavelet-based approach, as in the Monte Carlo study, we consider $J = 1$ and surpress the lowest detail while computing the first r principal components for the highest scale and retaining r factors from the corresponding reconstructed data matrix.[6]

Horizon	GDP growth	Inflation	Core inflation
$h = 1$	0.90	1.05	0.76***
$h = 2$	0.78	0.98	0.65***
$h = 3$	0.81	0.91*	0.64***
$h = 4$	0.80*	0.98	0.66***
$h = 5$	0.73**	0.94	0.66***
$h = 6$	0.71**	0.88	0.64***
$h = 7$	0.73**	0.92	0.68**
$h = 8$	0.72**	0.78	0.70**

TABLE 4. Out-of-sample forecasting evaluation

Note: The entries in the table denote the relative MSFE between the wavelet-based and principal components approach. The *,**,*** denote statistical significance of the forecasting gains at the 10, 5 and 1 per cent significance levels respectively.

---

6.   In addition, we considered the case of $J > 1$ with the number of factors determined on a scale-by-scale basis but the forecasting performance, in overall terms, does not improve over the case where $J = 1$.

Overall, one can see that for almost all variables and forecasting horizons, the wavelet based approach improves on the factor model based on standard principal components. Although the relative gains seem to be a bit lower for the one-quarter ahead horizon, in most cases one obtains noteworthy forecasting improvements. In particular, for GDP growth the gains attained are, on average, more than 20 per cent and near 30 per cent in some horizons. Even for the one-quarter ahead forecasts, where the improvement is the lowest, one obtains a 10 per cent gain. In the case of inflation, the ratio is, in general, lower than one but the magnitude of the gains are smaller and only around 10 per cent, on average. However, when one focus on core inflation, the forecasting improvements are striking reaching more than 30 per cent, on average. The gains range from 24 per cent for one-quarter ahead forecasts up to 36 per cent. These findings clearly support empirically the wavelet-based multiscale approach.

In the above forecasting exercise, we discarded the lowest detail by selecting no principal components at the lowest scale and set the number of factors to be retained informed by the work of Stock and Watson (2012). However, in practice, the number of factors to be considered is usually not known. To assess how one can cope with such a practical issue, we resort to the recently proposed criterion by Onatski (2010) which determines the number of factors based on the empirical distribution of the eigenvalues. Several reasons support the use of this criterion. On the one hand, Onatski (2010) shows that this criterion performs the best among a set of alternative criteria suggested in the literature, including the well-known Bai and Ng (2002) criteria. On the other hand, this criterion is particularly suitable in a context of weakly influential factors as shown by Onatski (2012). This is quite relevant in our case given that, as discussed earlier, one expects the explanatory power of the factors to be relatively small namely at the lowest scale.

Drawing on the above criterion, we find that no factors are selected at the lowest scale while the number of factors for the highest scale and for the reconstructed data is two. Note that the first finding provides additional empirical cross-validation for the above suggested procedure of surpressing the lowest detail whereas the fact that only two factors are retained is in line with the results of Onatski (2012) using US data. In Table 5, we present the corresponding relative forecasting behaviour.

Horizon	GDP growth	Inflation	Core inflation
$h = 1$	1.00	1.01	0.70***
$h = 2$	0.88	0.99	0.63**
$h = 3$	0.84	0.96	0.63**
$h = 4$	0.83	1.05	0.63**
$h = 5$	0.77	0.95	0.49**
$h = 6$	0.81	0.84	0.40**
$h = 7$	0.86	0.88	0.44*
$h = 8$	0.86	0.79	0.45*

TABLE 5. Out-of-sample forecasting evaluation with criterion determined number of factors

Note: The entries in the table denote the relative MSFE between the wavelet-based and principal components approach. The *,**,*** denote statistical significance of the forecasting gains at the 10, 5 and 1 per cent significance levels respectively.

The results in Tables 4 and 5 are broadly similar from a qualitatively point of view. As in the previously discussed exercise, we also find that for almost all variables and forecasting horizons, the wavelet based approach improves the forecasting performance. In particular, in the case of GDP growth, the forecasting gains are, on average, 15 per cent which is slightly lower than reported in Table 4. For inflation, the results are relatively close in overall terms. Finally, for core inflation, the improvement is even larger attaining, on average, 45 per cent, and being particularly visible for longer horizons. Hence, even if the number of factors selected at each scale is determined resorting to a criterion, the wavelet-based approach continues to deliver noteworthy forecasting gains.

5. Conclusions

In a context of growing data availability, the use of factor models has become widespread as it allows to take on board large datasets in a intuitive and parsimonious way. The estimation of the latent factors, which are subsequently used to obtain factor-augmented forecasts, is usually done via the principal components technique. Although it consistently estimates the space spanned by the true factors when the cross-section dimension and the number of observations tend to infinity, the finite sample performance of the principal components estimator deteriorates substantially when the explanatory power of the factors decreases vis-à-vis the explanatory power of the idiosyncratic

errors. Naturally, this impacts negatively on the forecasting performance of factor models.

Herein, we suggest a wavelet-based approach to cope with the above mentioned shortcoming. In particular, we propose estimating the factor model through a wavelet-based multiscale principal components analysis. Such an approach merges the benefits of principal components, which captures the relationship among the variables, and wavelet analysis, which enhances the decomposition of each variable dynamics. Based on a Monte Carlo simulation study, we show that it improves factor model estimation and forecasting performance.

Furthermore, we apply the suggested procedure to forecast GDP growth and inflation in the United States which are key variables for policymaking. We find that the wavelet-based approach delivers noteworthy forecasting gains over a wide range of forecasting horizons. On average, the forecasting improvement is more than 20 per cent in the case of GDP growth, around 10 per cent for inflation and more than 30 per cent for core inflation. These findings seem promising and reinforce the usefulness of wavelets to enhance forecasting performance.

References

Aguiar-Conraria, L. and Soares, M. (2011a) "Oil and the macroeconomy: using wavelets to analyze old issues", *Empirical Economics*, vol. 40(3), 645-655.

Aguiar-Conraria, L. and Soares, M. (2011b) "Business cycle synchronization and the Euro: A wavelet analysis", *Journal of Macroeconomics*, vol. 33(3), 477-489.

Aguiar-Conraria, L. and Soares, M. (2014) "The continuous wavelet transform: Moving beyond uni- and bivariate analysis", *Journal of Economic Surveys*, vol. 28(2), 344–375.

Altissimo, F., Bassanetti, A., Cristadoro, R.,Forni, M., Hallin, M., Lippi, M., Reichlin, L. and Veronese, G. (2001) "A Real Time Coincident Indicator for the Euro Area Business Cycle", CEPR discussion paper no. 3108.

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M. and Veronese, G. (2010) "New Eurocoin: Tracking Economic Growth in Real Time", *The Review of Economics and Statistics*, 92(4), 1024-1034.

Amengual, D. and Watson, M. (2007) "Consistent estimation of the number of dynamic factors in a large N and T panel", *Journal of Business and Economic Statistics*, 25(1), 91-96.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., Rünstler, G. (2011) "Short-term forecasts of euro area GDP growth", *Econometrics Journal*, 14(1), 25-44.

Artis, M., Banerjee, A., Marcellino, M. (2005) "Factor forecasts for the UK", *Journal of Forecasting* 24, 279-298.

Bai, J. and Ng, S. (2002) "Determining the number of factors in approximate factor models", *Econometrica*, 70(1), 191-221.

Bai, J.and Ng, S. (2008) "Large dimensional factor analysis", Foundations and Trends in Econometrics 3 (2), 89-163.

Bai, J. (2003) "Inferential theory for factor models of large dimensions", *Econometrica*, 71(1), 135-171.

Bakshi, B. (1998) "Multiscale PCA with application to MSPC monitoring", *AIChE Journal*, 44, 1596–1610.

Banerjee, A. and Marcellino, M. (2006) "Are there any reliable leading indicators for US inflation and GDP growth?", *International Journal of Forecasting*, 22, 137-151.

Barhoumi K., Darné, O. and Ferrara, L. (2010) "Are disaggregate data useful for forecasting French GDP with dynamic factor models?", *Journal of Forecasting*, 29, 1-2, 132-144.

Barhoumi K., Darné, O. and Ferrara, L. (2013) "Testing the number of factors for dynamic factor modelling: An empirical assessment for forecasting purpose", *Oxford Bulletin of Economics and Statistics*, 75, 1, 64-79.

Bierens, H. J. (2001) "Complex unit roots and business cycles: are they real?", *Econometric Theory* 17, 962-983.

Boivin, J., Ng, S. (2006) "Are more data always better for factor analysis?", *Journal of Econometrics* 132, 169-194.

Camba-Méndez, G. and Kapetanios, G. (2005) "Forecasting euro area inflation using dynamic factor measures of underlying inflation", *Journal of Forecasting*, 25, 491-503.

Castro, T. B, Rodrigues, P. M. M. and Taylor, A.M. R. (2013) "The impact of persistent cycles on zero frequency unit root tests" *Econometric Theory*, 29, 1289-1313.

Crowley, P. (2007) "A guide to wavelets for economists", *Journal of Economic Surveys*, 21, 207-264.

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

Doz, C., Giannone. D. and Reichlin, L. (2012) "A quasi-maximum likelihood approach for large, approximate dynamic factor models", *The Review of Economics and Statistics*, 94(4), 1014-1024.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) "The generalized dynamic factor model: Identification and estimation", *Review of Economics and Statistics* 82(4), 540-554.

Forni, M., Hallin, M. , Lippi, M. and Reichlin, L. (2001) "Coincident and Leading Indicators for the Euro Area", *The Economic Journal*, 111, 62-85.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004) "The generalized factor model: Consistency and rates", *Journal of Econometrics* 119, 231-255.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005) "The generalized dynamic factor model, one sided estimation and forecasting", *Journal of the American Statistical Association* 100, 830-840.

Gallegati, M. (2008) "Wavelet analysis of stock returns and aggregate economic activity", *Computational Statistics & Data Analysis*, vol. 52(6), 3061-3074.

Gallegati, M. (2012) "A wavelet-based approach to test for financial market contagion", *Computational Statistics & Data Analysis*, vol. 56(11), 3491-3497.

Gallegati, M. (2014) "Making leading indicators more leading: A wavelet-based method for the construction of composite leading indexes", *Journal of Business Cycle Measurement and Analysis*, vol. 2014(1), 1-21.

Gallegati, M., Gallegati, M., Ramsey, J. and Semmler, W. (2011) "The US Wage Phillips Curve across Frequencies and over Time", *Oxford Bulletin of Economics and Statistics*, vol. 73(4), 489-508.

Gallegati, M., Ramsey, J. and Semmler, W. (2014) "Interest rate spreads and output: A time scale decomposition analysis using wavelets", *Computational Statistics & Data Analysis*, vol. 76(C), 283-290.

Giannone, D., Reichlin, L. and Small, D. (2008) "Nowcasting: The real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55, 665-676.

Harvey, D., Leybourne, S., Newbold, P. (1997) "Testing the equality of prediction mean square errors", *International Journal of Forecasting* 13, 281–291.

Johnstone, I.M. and Lu, A.Y. (2009) "On consistency and sparsity for principal components analysis in high dimensions", *Journal of the American Statistical Association* 104, 682–693.

Mallat, S. G. (1989) "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (7), 674-693.

Marcellino, M., Stock, J., Watson, M. (2003) "Macroeconomic forecasting in the euro area: country specific versus euro wide information", *European Economic Review* 47, 1-18.

Onatski, A. (2010) "Determining the number of factors from empirical distribution of eigenvalues", *Review of Economics and Statistics* 92(4), 1004-1016.

Onatski, A. (2012) "Asymptotics of the principal components estimator of large factor models with weakly influential factors", *Journal of Econometrics* 168, 244-258.

Percival, D. and Walden, A. (2000) Wavelet methods for time series analysis, Cambridge University Press.

Pinheiro, M., Rua, A. and Dias, F. (2013) "Dynamic factor models with jagged edge panel data: Taking on board the dynamics of the idiosyncratic components", *Oxford Bulletin of Economics and Statistics*, 75(1), 80-102.

Ramsey, J. and Zhang, Z. (1996) "The application of wave form dictionaries to stock market index data", In: Kratsov, Y., Kadtke, J. (Eds.), Predictability of complex dynamical systems, Springer.

Ramsey, J. and Zhang, Z. (1997) "The analysis of foreign exchange data using waveform dictionaries", *Journal of Empirical Finance* 4, 341-372.

Ramsey, J.and Lampart, C.(1998a) "Decomposition of economic relationships by time scale using wavelets", *Macroeconomic Dynamics* 2 (1), 49-71.

Ramsey, J. and Lampart, C. (1998b) "The decomposition of economic relationships by time scale using wavelets: expenditure and income", *Studies in Nonlinear Dynamics and Econometrics* 3 (1), 23-42.

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*, vol. 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 Nunes, L. C. (2005) "Coincident and leading indicators for the euro area: A frequency band approach", *International Journal of Forecasting*, vol. 21, no 3, 503-523.

Rua, A., and Nunes, L.C. (2009) "International comovement of stock market returns: A wavelet analysis", *Journal of Empirical Finance* 16, 632-639.

Rua, A., and Nunes, L.C. (2012) "A wavelet-based assessment of market risk: The emerging markets case", *Quarterly Review of Economics and Finance*, vol. 52, 84-92.

Runstler, G., Barhoumi, K., Cristadoro, R., Reijer, D., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Benk, S., and Van Nieuwenhuyze, C. (2009) "Short-term forecasting of GDP using large monthly data sets: A pseudo real-time forecast evaluation exercise", *Journal of Forecasting*, 28, 7, 595-611.

Schumacher, C. (2007) "Forecasting German GDP using alternative factor models based on large datasets", *Journal of Forecasting*, 26, 271-302.

Schumacher, C. (2010) "Factor forecasting using international targeted predictors: The case of German GDP", *Economics Letters*, 107, 2, 95-98.

Schumacher, C. and Breitung, J. (2008) "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data", *International Journal of Forecasting*, 24, 368-398.

Stock, J. and Watson, M. (1989) "New Indexes of Coincident and Leading Economic Indicators" in O. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual*, MIT Press, 351-394.

Stock, J. and Watson, M. (1991) "A Probability Model of the Coincident Economic Indicators" in K. Lahiri and G.H. Moore (eds.), *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, 63-89.

Stock, J. and Watson, M. (1993) "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience" in J.H. Stock and M.W. Watson (eds.), *Business Cycles, Indicators and Forecasting*, University of Chicago Press, 255-284.

Stock, J. and Watson, M. (1998) "Diffusion Indexes", NBER Working Paper no. 6702.

Stock, J. and Watson, M. (1999) "Forecasting inflation", *Journal of Monetary Economics*, 44, 293-335.

Stock J. and Watson, M. (2002a) "Macroeconomic forecasting using diffusion indices", *Journal of Business and Economics Statistics*, 20, 147-162.

Stock, J. and Watson, M. (2002b) "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association* 97, 1167-1179.

Stock, J. and Watson, M. (2012) "Generalized Shrinkage Methods for Forecasting Using Many Predictors", *Journal of Business and Economic Statistics*, 30(4), 481-493.

Valle e Azevedo, J., Koopman, S. J. and Rua, A. (2006) "Tracking the business cycle of the Euro area: a multivariate model-based band-pass filter", *Journal of Business & Economic Statistics*, vol. 24, no 3, 278-290.

WORKING PAPERS

2013

2014

2015

2016