BANK OF GREECE

EUROSYSTEM

Working Paper

A suggestion for a Dynamic Multi Factor Model (DMFM)

Heather D. Gibson
Stephen G. Hall
George S. Tavlas

282

JULY 2020

BANK OF GREECE
Economic Analysis and Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Published by the Bank of Greece, Athens, Greece

ISSN: 2654-1912 (online)

A SUGGESTION FOR A DYNAMIC MULTI FACTOR MODEL (DMFM)

Heather D. Gibson
Bank of Greece

Stephen G. Hall
University of Leicester, Bank of Greece and University of Pretoria

George S. Tavlas
Bank of Greece and Hoover Institution, Stanford University

ABSTRACT

We provide a new way of deriving a number of dynamic unobserved factors from a set of variables. We show how standard principal components may be expressed in state space form and estimated using the Kalman filter. To illustrate our procedure we perform two exercises. First, we use it to estimate a measure of the current-account imbalances among northern and southern euro-area countries that developed during the period leading up to the outbreak of the euro-area crisis, before looking at adjustment in the post-crisis period. Second, we show how these dynamic factors can improve forecasting of the euro-dollar exchange rate.

Keywords: Principal Components, Factor Models, Underlying activity, Forecasts

JEL Classification: E3, G01, G14, G21

Acknowledgements: We are very grateful to Apostolis Serletis, the Associate Editor, and two referees for constructive comments on an earlier draft. The views expressed in this paper are the authors' own and do not necessarily represent those of the institutions with which they are affiliated.

Correspondence:
George S. Tavlas
Bank of Greece
21 E Venizelos Ave,
Athens, 10250, Greece
Tel. no. +30 210 320 2370
Fax. no. +30 210 320 2432
Email: gtavlas@bankofgreece.gr

1. Introduction

There has been a long tradition of using either factor models (principal components) or dynamic factor models to: (i) derive measures of unobserved effects on key economic indicators; and (ii) to concentrate information for the purposes of forecasting. These models have been applied, for example, to measure underlying economic activity by Stock and Watson (1989) and Garrett and Hall (1996) and underlying inflation by Stock and Watson (1999) and González, Melo, Monroy and Rojas (2009). Applications to forecasting include studies by Artis, Banerjee and Marcellino (2001), Zaher (2007) and Ziegler and Eickmeier (2008). Key contributions to the development of factor models include Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2002a, 2002b). Factor models have become increasingly popular as a way of extracting information from data sets that consist of a fairly large cross-sectional element as well as a time series dimension.

In this paper, we focus on the two basic linear approaches to factor analysis that have been used in the literature. One approach is based on principal components and its variants. The second approach is based on the Kalman filter. Both approaches have advantages and disadvantages. The advantage of the principal component approach is that it is able to produce more than a single factor from the original series. A disadvantage of the approach is that it is inherently static in the sense that the component series at each point in time are only a function of the data at that point in time. A key advantage of the Kalman filter approach is that it is dynamic, in the sense that the factor produced will be smoothed since it is a function of the data, not only in the current period, but also both future and past values. Its disadvantage is that it can only be used to produce a single factor from a range of series. Hybrid approaches that start from principal components and then apply the Kalman filters to the analysis to smooth the resulting factors also exist; but these approaches are far from satisfactory.[1]

In this paper, we show: (i) how principal components may be represented in a state space form; and (ii) how this representation generalises to a dynamic multi factor model (DMFM). We apply this representation in two ways. First, we use it to model the growing structural imbalances in the current account to GDP ratios and their subsequent partial reversal which occurred between northern European countries (Germany, Austria and the

[1] Specifically, these approaches smooth the principal components in a arbitrary way -- for example, by filtering the components.

Netherlands) and southern European crisis countries (Greece, Ireland, Spain and Portugal) during the periods 2003:Q1 to 2008:Q1 and 2008:Q2 to 2018:Q1, respectively. Using our method of representation, we derive an underlying measure of these imbalances. Second, we show how these dynamic factors can improve our ability to forecast exchange rates; our focus here is the euro-U.S. dollar exchange rate.

The paper is structured as follows. Section 2 provides an overview of dynamic factor models and of our proposed procedure. Section 3 applies the procedure to the current accounts (relative to GDP) of the seven euro-area countries mentioned above over the two sub periods. Section 4 uses these dynamic factors to address the issue of forecasting the euro-U.S. dollar exchange rate. Section 5 presents our main conclusions. Finally, Appendix 1 provides a worked-out example that shows the exact equivalence between principal components and the Kalman filter model.

2. A Dynamic Multi Factor Model (DMFM)

2.1 Principle Components and Dynamic Factor Models

Let Y_t be a vector of R variables of interest i=1…R measured over T periods t=1…T, that is y_{it}. The objective of factor models is to summarise the information in Y_t in a smaller number of factors F_t, where F_t is a vector containing f_{jt} variables where j=1…J, and where J is less than or equal to R. Typically, we would want the number of useful factors to be considerably less than R, although principal components can produce up to R factors. The assumption we make throughout this paper is, therefore, that there exists a set of common factors underlying the observed data such that;

$$Y_t = \mathrm{M}F_t \tag{1}$$

Principal components proceeds by choosing the first factor to be a series that explains as much of the variation in y_{it} as possible. The second factor is then derived as a series that explains as much of the remaining variation y_{it} as possible, subject to being orthogonal to the first factor. The third factor is chosen so that it explains as much of the remaining variation as possible subject to being orthogonal to both of the first two factors, and so on. Formally, this is done by first normalizing the variables in the vector y_{it} so that they have a zero mean and unit variance; call this vector $\bar{Y}_t$. Then, the principal components, P, may be derived in matrix form as:

$$P = \bar{Y}W \tag{2}$$

where W is an RxR matrix. To construct the first principal component, the idea is that the first column of W is chosen so as to maximize the variance of the first principal component $p_1 = \bar{Y}w_1$. Such a linear combination is given by $\mathrm{var}(\bar{Y}w_1)$. Hence, the problem that is solved by principal components is to maximise the following equation with respect to the vector w_1 (see Joliffe and Cardima (2016)).

$$\max \mathrm{var}(\bar{Y}w_1) = w_1{}'Sw_1 \tag{3}$$

where S is the sample covariance of $\bar{Y}$. The solution to this problem is not unique because any linear scaling of w_1 will produce an equivalent answer, and so we need to identify a unique set of weights. The usual way to address this issue is to impose the constraint that $w_1 w_1{}' = 1$. We may then go on to sequentially solve for the remaining principal components, with the added constraint that each subsequent principal components is orthogonal to the ones that come before it. The full W matrix is, in fact, given as the eigenvectors of $\bar{Y}'\bar{Y}$. Typically, the information in the R variable comprising the $\bar{y}_{it}$ series[2] is explained by a relatively small number of factors, or principal components; as mentioned, these may be used either in forecasting or as measures of some underlying concept, such as economic activity. The disadvantage of this approach is that the factors are inherently static. For example, if we wished to model underlying economic activity, we might believe that the economy evolves smoothly. If the data are erratic, or even seasonal, then the principal components (or factors) will remain erratic or seasonal. Following Doz, Giannone and Reichlin (2011), it has been common practice to regress time series models on these factors in order to produce smoother versions of the factors and to make them dynamic. Doz, Giannone and Reichlin (2012) give this procedure a quasi-maximum likelihood interpretation. Our procedure, below, derives a full maximum likelihood estimator for the dynamic factors and, thus, should achieve the Cramer-Rao lower bound. It is, of course, the case that deriving a maximum likelihood estimator is not a necessary condition for achieving this result.

[2] For simplicity of notation we will henceforth drop the bar notation in $\bar{y}_t$ and will simply note when the variables are normalised.

The Kalman filter approach is inherently dynamic from the start. Following the procedure originated by Stock and Watson (1989) and Garratt and Hall (1996), a state space form is set up under which a set of R measurement equations are specified as a function of an unobservable common factor. The R measurement equations are:

$$
\begin{aligned}
&y_{1t} = \lambda_1 f_t + \varepsilon_{1t} \\
&y_{2t} = \lambda_2 f_t + \varepsilon_{2t} \\
&. \\
&. \\
&. \\
&y_{Rt} = \lambda_R f_t + \varepsilon_{Rt} \\
&\varepsilon_{it} \sim N(0, \sigma_i) \quad i = 1 \ldots R
\end{aligned}
\tag{4}
$$

where the ε's are measurement errors, with zero means and constant variances, and the state equation is given by:

$$
\begin{aligned}
&f_t = \phi(L) f_{t-1} + e_t \\
&e_t \sim N(0, \varpi)
\end{aligned}
\tag{5}
$$

where e_t is the state equation error with zero mean and constant variance, and where $\phi(L)$ is a lag polynomial. The Kalman filter smoothing algorithm produces an optimal estimate of the factor f that explains as much of the movement in y_{it} as possible; the factor f is smoothed over time to allow for the dynamics of the process. In contrast to the various extensions to principal components, which produces smoothing in an *ad hoc* way, the Kalman filter produces the optimal level of smoothing.[3]

As mentioned, the main disadvantage of this approach is that it has not been possible to directly estimate more than one factor. Specifically, generalizing equation (1) to many factors produces a system that is not identified and, thus, cannot be estimated. It is possible to identify the model by imposing various constraints on the parameters of the model (see Harvey (1989)); indeed, this is, in effect, what we do below in a way that allows us to replicate what happens in principal components within the Kalman filter.

[3] See Cuthbertson, Hall and Taylor (1992).

2.2 Principal Components as a Special Case of the Kalman filter

Before turning to the dynamic model, it will be useful to show how the Kalman filter can exactly reproduce static principal components. That is, we will show that the Kalman filter can be used to derive precisely the same factors as those produced by principal components. To do this we set up the following state space form. The measurement equations are:

$$
\begin{aligned}
&y_{1t} = \lambda_1 f_t + \varepsilon_{1t} \\
&y_{2t} = \lambda_2 f_t + \varepsilon_{2t} \\
&. \\
&. \\
&. \\
&y_{Rt} = \lambda_R f_t + \varepsilon_{Rt} \\
&\varepsilon_{1t} \dots \varepsilon_{R_t} \sim N(0,1)
\end{aligned}
\tag{6}
$$

where, in contrast to model (2), all the errors have the *same* constant variance. The state equation is given by:

$$
\begin{aligned}
&f_t = e_t \\
&e_t \sim N(0,\sigma^2)
\end{aligned}
\tag{7}
$$

where f_t is the state variable, σ^2 is the variance of the state equation which is to be estimated, $y_1 .. y_R$ are the variables as before, which have been standardised so that they have a zero mean and unit standard error, following the first stage of principal components. We normalise the state variable on y_1 using an identification assumption (without loss of generality) and estimate the remaining λ_i. The error terms in the state equation have the same variance which we normalise to 1; this mimics the principle component approach of giving equal weights to all the series;[4] e_t is the error term in the state equation. Note that there are no dynamics in the state equation, which is unusual, but permissible, under the Kalman filter. The smoothed state variable will now give the maximum possible explanation of the variation in all the variables, which, intuitively, is exactly the same thing done by the first factor in the principal component procedure. More formally, the problem which the Kalman filter solves is to

[4] If different variances are assigned to each measurement equation then this is equivalent to weighted principal components where variables can have unequal weights in the construction of the principal components.

minimise the squared errors in the measurement equations (Jazwinski (1970) or Harvey (1989)), that is to

$$\min z = \sum_{t=1}^{T}\sum_{k=1}^{R}\varepsilon_{kt}^{2} \tag{8}$$

This is minimised with respect to the unknown parameters of the state space form, in this case $\lambda_2 \ldots \lambda_R$ *and* σ^2 .This is simply the dual of the principal components problem (3) -- if we maximise the variance of the factor we minimise the variance of the errors. We can see this is we define the sum of the squared normalised data as $\Sigma = \sum_{i=1}^{T}\sum_{j=1}^{R}\bar{y}_{ij}{}^{2}$ where $\bar{y}_{ij}$ are the elements of $\bar{Y}$ (as defined above (2)) then we can see that

$$\Sigma = w_1{}'Sw_1 + \varepsilon{}'\varepsilon \tag{9}$$

Σ , is fixed from the data. Principal components maximises the first term on the right hand side of (9), which, of course, minimises the second term. The Kalman filter minimises the second term, which maximises the first term.

Again, we have an identification problem in that the state variables are only unique up to a multiplicative factor; and, again, we need an arbitrary normalisation. In this case, the normalization often takes the form of setting $\lambda_1 = 1$. Apart from this normalisation scaling factor, the first principal component and the static state variable both contain the same information. Indeed, if the first state equation is normalised to the value of the first principal component's loading weight, then the two will be identical. Further, if we regress the first principal component on the first state variable, we will get an R^2 of exactly 1. Appendix 1 provides an example using artificial data that illustrates the equivalence between principal components and the Kalman filter model outlined above.

Now, in order to derive the second state variable under model (3), we create a set of variables based on the state variable as follows:

$$\begin{aligned}
\gamma_{11t} &= \lambda_1 f_t \\
\gamma_{12t} &= \lambda_2 f_t \\
&. \\
&. \\
&. \\
\gamma_{1Rt} &= \lambda_R f_t
\end{aligned} \tag{10}$$

where γ_{11t}, is the effect of the first state variable on y_{1t} ; that is, γ_{11t} is equal to the component extracted for y_{1t} Similarly, γ_{12t} is the component extracted for y_{2t}, etc. We now set up the following state space form to derive the second state variable:

$$
\begin{aligned}
&y_{1t} = f_{2t} + \gamma_{11t} + \varepsilon_{1t} \\
&y_{2t} = \lambda_2 f_{2t} + \gamma_{12t} + \varepsilon_{2t} \\
&. \\
&. \\
&. \\
&y_{Rt} = \lambda_R f_{2t} + \gamma_{1Rt} + \varepsilon_{Rt} \\
&\varepsilon_{1t} \ldots \varepsilon_{Rt} \sim N(0,1)
\end{aligned}
\tag{11}
$$

and the state equation is again given by:

$$
\begin{aligned}
&f_{2t} = e_t \\
&e_t \sim N(0,\sigma^2)
\end{aligned}
\tag{12}
$$

This formulation provides the best explanation of the variables after removing the effect of the first state variable. Again, it will be identical to the second principal component except for the scaling given by the normalisation. This process may be repeated to derive as many state variables as required, thus demonstrating the equivalence of the static Kalman filter approach and the principal component procedure. But, of course, there is little reason to perform this procedure since principal components are quicker and easier to perform than the iterated set of Kalman filter models.

2.3 The Dynamic Multiple Factor Model (DMFM)

To briefly summarize, we have shown how a state space form can exactly replicate principal components. The key restriction made to achieve this result is to make the state equation static. To generate a dynamic factor model, we relax that restriction. That is, what we need to do to generate a succession of dynamic factors is to repeat the process given above, but with a dynamic set of state equations.

To demonstrate, we begin by estimating the single factor Kalman filter, given as model (2) above. We then create a set of new variables which are given by the factor multiplied by its loadings:

$$
\begin{aligned}
&\gamma_{11t} = f_t \\
&\gamma_{12t} = \lambda_2 f_t \\
&. \\
&. \\
&. \\
&\gamma_{1Rt} = \lambda_R f_t
\end{aligned}
\tag{13}
$$

where the λ' s are the loading weights. Then, we modify the Kalman filter state equations by adding these variables to each state equation and perform a second Kalman filter estimation to estimate a second factor. Thus:

$$
\begin{aligned}
&y_{1t} = f_{2t} + \gamma_{11t} + \varepsilon_{1t} \\
&y_{2t} = \lambda_{22} f_{2t} + \gamma_{12t} + \varepsilon_{2t} \\
&. \\
&. \\
&. \\
&y_{Rt} = \lambda_{R2} f_{2t} + \gamma_{1Rt} + \varepsilon_{Rt} \\
&\varepsilon_{1t} \ldots \varepsilon_{Rt} \sim N(0,1)
\end{aligned}
\tag{14}
$$

This procedure produces a second factor which explains as much of the variation in the observed variables as possible not explained by the first factor. This produces a set of components that is analogous to principal components, except that the components are dynamic. That is, as in equation (3), the factors have a dynamic structure.

We can then proceed to another iteration to extract a third dynamic factor by again defining a new set of variables as:

$$
\begin{aligned}
&\gamma_{22t} = f_{2t} \\
&\gamma_{22t} = \lambda_{22} f_{2t} \\
&. \\
&. \\
&. \\
&\gamma_{2Rt} = \lambda_{R2} f_{2t}
\end{aligned}
\tag{15}
$$

and then estimating a standard Kalman filter with these variables added.

$$
\begin{aligned}
& y_{1t} = \lambda_{13} f_{3t} + \gamma_{11t} + \gamma_{21t} + \varepsilon_{1t} \\
& y_{2t} = \lambda_{23} f_{3t} + \gamma_{12t} + \gamma_{22t} + \varepsilon_{2t} \\
& . \\
& . \\
& . \\
& y_{Rt} = \lambda_{R3} f_{3t} + \gamma_{1Rt} + \gamma_{2Rt} + \varepsilon_{Rt} \\
& \varepsilon_{1t} \ldots \varepsilon_{Rt} \sim N(0,1)
\end{aligned} \tag{16}
$$

This will then produce a third dynamic factor and the process can be repeated as many times as we wish.

An obvious way to choose between these models would be in terms of a likelihood ratio test; we should continue producing more factors until the final factor produced does not produce a significant rise in the likelihood function following the addition of the last factor. Given that we have the likelihood function, it would also be straightforward to construct one of the standard information criteria, such as the AIC or the SBC criteria.

3. A DMFM for the Current-Account Positions of Northern and Southern Euro-area Countries

We utilize the above multiple dynamic factor model to derive an indicator of the degree of current-account imbalance between the groups of northern and southern euro-area countries. We expect the first dynamic factor (as the dominant one) to capture the growing imbalances that developed during the early part of the euro period up to the onset of the 2008 financial crises. Specifically, we focus on the current account to GDP ratios (CB) of three northern countries -- Germany Austria and the Netherlands -- and four southern countries which experienced a sovereign debt crises -- Portugal, Spain, Greece and Ireland.[5] The data are quarterly; the sample period is 2003:Q1 to 2018:Q4. The source of the current account and GDP data is Eurostat. Figures 1a and 1b illustrates the paths of the current account balances in periods before and after the outbreak of the crisis in 2008. The Figures reveal several distinct patterns. First, during the sub period 2003:Q1 to 2008:Q1, the current account positions of the three northern countries rose sharply; for example Germany's surplus rose from 2 per cent of GDP at the beginning of the period to about 6 per cent of GDP at the end

[5] The specific countries were selected because the three "northern" countries had the largest current-account surpluses relative to GDP in the euro area during 2001 and 2008, and the four "southern" countries had the largest deficits.

of the period. Second, whereas the current-account surpluses of Austria and Germany exhibited steady rises in the first sub period, the current-account surplus of the Netherlands exhibited more erratic behavior, rising from about 0 per cent at the beginning of the period, peaking at 8 ½ per cent in 2006:Q3, and then falling to 6 per cent in 2008:Q1. This circumstance is related to its position as a producer of natural resources and an exporter of crude oil and processed petroleum products. Its current account is heavily influenced by oil and gas prices. Third, each of the four southern countries experienced large deteriorations in their current-account deficits in the first sub period. Fourth, in the second sub period, the current-account surpluses of the three northern countries evolve in different ways -- the current-account surplus of Austria initially fell (from about 4 per cent to about 2 per cent) and then fluctuated around 2 per cent in a steady way; that of Germany rose from about 6 per cent to about 8 per cent steadily, while that of the Netherlands exhibited erratic behavior, rising from 4 per cent to over 10 per cent, then falling to 4 per cent, before rising to 8 per cent. Fifth, after 2008, the southern countries adjusted and reduced their deficits considerably, typically moving either to a balanced current account or to surpluses; the northern countries either maintained stable surpluses (Austria) or increased their surpluses in either a steady way (Germany) or a somewhat erratic way (the Netherlands).

Period from 2003:Q2 to 2008:Q1

We begin by examining the common factor structure underlying the seven countries over the first sub period when we believe the fundamental imbalances were building up. To do this we begin by seasonally adjusting the current account balance data[6] and standardising each variable. We then set up the following state space form; the measurement equations are[7]:

[6] It is critical to work with seasonally adjusted current account data, especially in the case of the southern European countries where, because of the importance of tourism, inter alia, the current account is highly seasonal.

[7] We suppress the obvious time subscript for notational simplicity

$$
\begin{aligned}
CB_{GE} &= \phi_1 + \varepsilon_1 \\
CB_{AU} &= \alpha_1 \phi + \varepsilon_2 \\
CB_{NL} &= \alpha_2 \phi + \varepsilon_3 \\
CB_{IR} &= \alpha_3 \phi_1 + \varepsilon_4 \\
CB_{PT} &= \alpha_4 \phi_1 + \varepsilon_5 \\
CB_{ES} &= \alpha_5 \phi_1 + \varepsilon_6 \\
CB_{GR} &= \alpha_6 \phi_1 + \varepsilon_7 \\
\varepsilon_1 &\ldots \varepsilon_7 \sim N(0, \sigma)
\end{aligned}
\tag{17}
$$

where CB is the observed ratio of the current-account position relative to GDP for Germany (GE), Austria (AU), the Netherlands (NL), Ireland (IR), Portugal (PT), Spain (ES) and Greece (GR). The state equation is:

$$
\begin{aligned}
\phi_{1t} &= \phi_{it-1} + e_t \\
e_t &\sim N(0,1)
\end{aligned}
\tag{18}
$$

This produces the following state variable, or the first dynamic factor, which is our index of structural imbalance (Figure 2). This shows a fairly smooth increasing level of imbalance over the period. The coefficients on the measurement equations are given in Table 1, where the German coefficient is normalised to 1.0 as an identification condition; hence, there is no Z-statistic available for Germany. The key result here is that the loading weights on Germany, Austria and the Netherlands are positive while those of the other four countries are all negative. This result implies that as the dynamic common factor rises, the German, Austrian and Netherlands current balances move further into surplus while the other four countries move further into deficit, demonstrating the growing divergence of the external positions between the northern and the southern countries during this period.

We can see how much of each of the current balance for each country is explained by this first dynamic factor by running a simple regression of the first dynamic factor on the CB for each country. This is shown in the final column of Table 1 as the R^2 from this regression. The imbalance index has a very high degree of explanatory power for all countries except the Netherlands. The low (*i.e.*, 0.01) R^2 for the Netherlands reflects the nature of the procedure used to obtain the factor. As discussed, the evolution of the current account-to-GDP ratio of the Netherlands has little in common (in terms of a common factor) with the other countries. Hence, the common factor found for the other countries has little effect on the Netherlands, producing the low R^2.

As mentioned, the first dynamic factor is identical to what is derived under the standard factor model. We can now proceed to derive the second dynamic factor, which *cannot* be derived under the standard model. As described above, this is done by adding the first factor multiplied by its coefficient to each of the measurement equations and then re-running the Kalman filter. This procedure produces the second dynamic factor shown in Figure 3. There is no clear trend in this factor, as we would expect, since it should be orthogonal to the first factor. The coefficients on the second dynamic factor no longer have the clear pattern found for the first factor. The coefficient on the second dynamic factor is again normalised to unity on Germany (Table 2). The coefficients of three countries, Austria, Ireland, and Portugal have the same (positive) sign as Germany while the other three countries have the opposite (negative) sign. This factor is, therefore, picking-up some differences among the countries rather than something they experience in common. The R^2 rises substantially for three countries: the Netherlands (from 0.01 to 0.78), Austria (from 0.64 to 0.83), and Portugal (from 0.60 to 0.78). The loading weight for the Netherlands also has a high Z statistic relative to most of the other countries. This factor would then seem to be mainly picking up something that is largely specific to the Netherlands and which does not have the same general applicability as the first dynamic factor. In short, there is clear and strong evidence that over this period there is an underlying common factor that links the rises in the current account surpluses of Germany and Austria, to the rising deficits in Spain, Greece and Ireland.

Period from 2008:Q2 to 2018:Q1

We now turn to the second sub period. During this period the four crisis countries had to undertake sharp fiscal contractions, to eliminate their current account deficits; the upshot of those measures were reductions in domestic demand and sharp improvements in their current-account positions. We would, therefore, expect to see that their current accounts were moving in line with those of the northern countries, which were not subjected to contractionary policies and which tended to exhibit increases in their surpluses in the second period (Figure 1a). Again applying the dynamic factor analysis we derive the first factor which is presented in Figure 4. The first dynamic factor continues to have an upward slope although it levels off after 2013. The difference between this and the earlier period comes in the form of the loading weights shown in Table 3. In this case, all the coefficients are positive with the exception of Austria, meaning that the current balance to GDP ratios of all four crisis countries improve (*i.e.*, move either from deficits to surpluses, or from relatively-high deficits to lower deficits) over this period -- see Figure 1). However, as mentioned, the current-account surplus of Germany rose. The first dynamic factor has little explanatory power for Austria and the Netherlands, which experienced either slowly falling surpluses (Austria) or, behaved somewhat erratically (the Netherlands) relative to the other countries.

We now turn to the second dynamic factor presented in Figure 5. This factor seems to largely reflect differences among the three northern countries -- the erratic behaviour of the Netherlands in 2014 and 15 (see Figure 1) and the divergence of Germany from Austria over the first half of the period. We can see from the factor loadings in Table 4 that this factor almost exclusively explains developments in the north; this circumstance is evidenced in the sharp increases in the (cumulated) R^2 for each of the northern countries -- from 0.68 (with a single factor) to 0.81 (with two factors) for Germany; from 0.11 to 0.52 for Austria; and from 0.14 to 0.60 for the Netherlands. All the crisis countries have very small loading weights which are insignificant; and, the cumulated R^2 for these countries hardly rises from the simple R^2 in Table 3.

The broad conclusion from this section is that the first sub period was one of growing imbalances in which the first dynamic factor dominated developments in Austria and Germany, in a positive way, while it also dominated in the four crisis countries but in a negative way. This factor is then a good measure of the growing imbalances which developed between the north and the south and could usefully act as a summary variable of the phenomenon in econometric models. The second sub period did not see a reversal of this imbalance in the sense that Germany continued to have a rising surplus, while the Netherlands and Austria did not reduce their surpluses. The crisis countries turned around their current account deficits by internal adjustments to economic activity and domestic prices; the surplus countries did not contribute to narrowing imbalances. In each of the sub periods, our dynamic multi factor method substantially increased the explanatory power for three countries -- Austria, the Netherlands and Portugal in the first sub period, and Austria, Germany and the Netherlands in the second sub period -- compared to what was obtained under the standard method. Again, the factor could be used to summarise current account adjustment during the crisis. In the following section we provide an example of just such a use.

4. Forecasting the Euro-Dollar Exchange Rate

Since the seminal work of Meese and Rogoff (1983), who convincingly demonstrated that a simple random walk model could outperform virtually all other exchange rate models in forecasting accuracy, exchange rate forecasting has been a notable area of forecast failure. This finding has been replicated many times since the study by Meese and Rogoff -- see, for example, the surveys by Rossi (2013) and Caraiani (2017). Given this general failure, it seems appropriate to ask if the dynamic factors generated above can provide useful information in a simple forecasting exercise for the euro exchange rate. To provide context, it seems reasonable that current-account balances should influence the exchange rate. Yet, it would not be sensible to enter the current-account balances of all euro-area members into a forecasting equation for the euro because such a model would be highly over-parameterised. By deriving the dynamic common factors, however, we can concentrate the information contained in the large number of current-account variables into a much more parsimonious form.

To this end, we begin by estimating a simple autoregressive model of the log of the euro exchange rate against the US dollar (dollars per euro) over the two sub-periods specified above, and then we add the lagged dynamic factors derived earlier (here called DF1 and DF2, respectively). We then undertake one-step-ahead, static, forecasts and assess the forecasting performances of three model (discussed below) based on three criteria -- the root mean square error (RMSE), the mean absolute error (MAE), and Theil's inequality coefficient -- over the two periods used above to derive the dynamic factors. The exchange-rate data are from the ECB's statistical Data Warehouse.

The results of this exercise are reported in Table 5. For both periods we start from a simple AR(2) model which seems to be a good basic model, passing a range of standard diagnostics.[8] For the period 2003-2008, when we add the first dynamic factor alone, it is significant, but does not produce improvement in the forecast diagnostics (in fact, it produces a small deterioration in the RMSE and the MAE). We then add both dynamic factors; the first is significant, and the second, while not significant (with a t-stat of 1.2), does improve the forecasting performance of the equation; the RMSE, the MAE and the Theil inequality

[8] For each sub period, we lose two observations because of the AR(2) specifications. The diagnostics are available from the authors.

coefficient all fall substantially, indicating that the second dynamic factor produces a significant improvement in forecasting ability. The results are even stronger for the 2008-2018 period, where the first factor alone is not significant and adds little to the forecasting performance. When we include the two factors, however, both are highly significant and together produce a substantial improvement in forecast performance. To sum up, for both periods, the inclusion of two dynamic factors improve forecasting performance on the basis of the three criteria considered compared with both the baseline AR(2) model and with the single-factor model.

The intuition underlying this result is the following. While the first factor captures the broad, smooth, trend-like behaviour in the current balances, this effect is probably well captured by the AR(2) time series component of the model. Hence, the first factor adds little to the explanation of the exchange rate. The second factor, however, picks up sharp, sudden movements in the current balances, which would not be proxied by the simple time series model. Thus, this factor contributes significantly to the forecast. This example, therefore, demonstrates the importance of using more than one factor.

5. Conclusions

We have demonstrated that principle components can be generated from a state space representation using the Kalman filter, thus making the generalization to a dynamic multi factor model (DMFM) straightforward. We illustrated this approach by looking at the underlying dynamic factors for the current-account balance to GDP ratios for seven euro-area counties, deriving an index of current-account imbalance based on this approach for two sub periods. We then used these dynamic factors in a forecasting exercise for the euro-dollar exchange rate and demonstrated that these factors enhance the forecasting ability of a simple AR model. In addition, in both periods considered it is the second dynamic factor -- a factor which has previously not been possible to calculate -- that brings about the major improvement in the forecast diagnostics.

References

Artis, M., Banerjee, A. & Marcellino, M. (2001). Factor forecasts for the UK. *Working Papers 203*, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Cuthbertson K. Hall S.G. & Taylor M.P. (1992). *Applied econometric techniques*. University of Michigan Press, Ann Arbour.

Doz, C., Giannone, D. & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164 (1), 188-205.

Doz, Catherine, Giannone, Domenico and Reichlin, Lucrezia (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. & Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82 (4), 540-554.

Garratt A. & Hall S.G. (1996). Measuring underlying economic activity. *Journal of Applied Econometrics*, vol. II, p 135-151.

Giannone, D., Reichlin L. & Small, D. (2008). Nowcasting: the real time informational content of macroeconomic data releases. *Journal of Monetary Economics,* 55, 665– 676.

González E., Melo, L. F., Monroy, V. & Rojas, B. (2009). A dynamic factor model for the Colombian inflation. *Borradores de Economia*, 549, Banco de la Republica de Colombia.

Harvey A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge University Press.

Jazwinski A.H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press, isbn 0-12-381550-9

Joliffe I. T. and Cadima J. (2016). ‘Principal Component Analysis: A Review and Recent Developments’ *Philosophical Transactions* A374 Royal Society, dx.doi.org/10.1098/rsta.2015.0202

Meese R.A & Rogoff K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample. *Journal of International Economics*, 14 (1-2), 3-24.

Rossi B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51 (4), 1063-1119.

Stock J.H. and Watson M.W. (1989). New indexes of coincident and leading economic indicators. NBER Chapters in: *NBER Macroeconomics Annual* 1989, Volume 4, pages 351-409, National Bureau of Economic Research.

Stock J.H. & Watson M.W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2), 293-335.

Stock J.H. & Watson M.W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97 (460), 1167-1179.

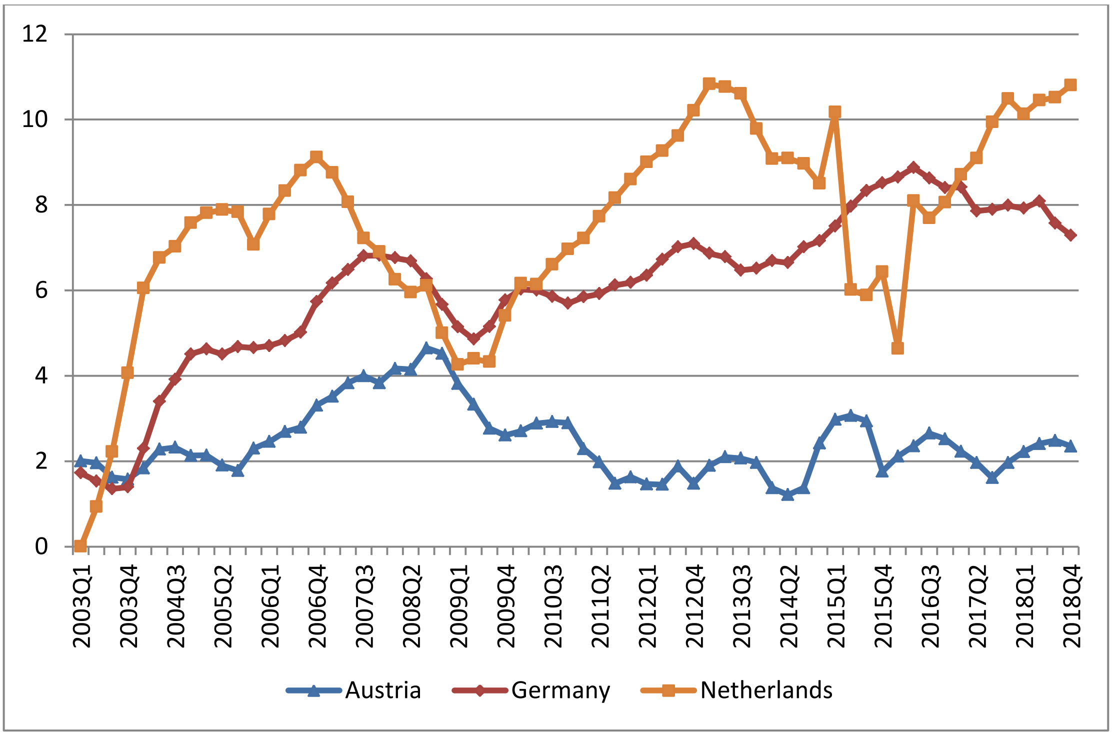
Stock J.H. & Watson M.W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economics Statistics*, 20 (2), 147-62.

Zaher, F. (2007). Evaluating factor forecasts for the UK: The role of asset prices. *International Journal of Forecasting*, 23 (4), 679-693.

Ziegler, C. & Eickmeier, S. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, 27, 237–265.

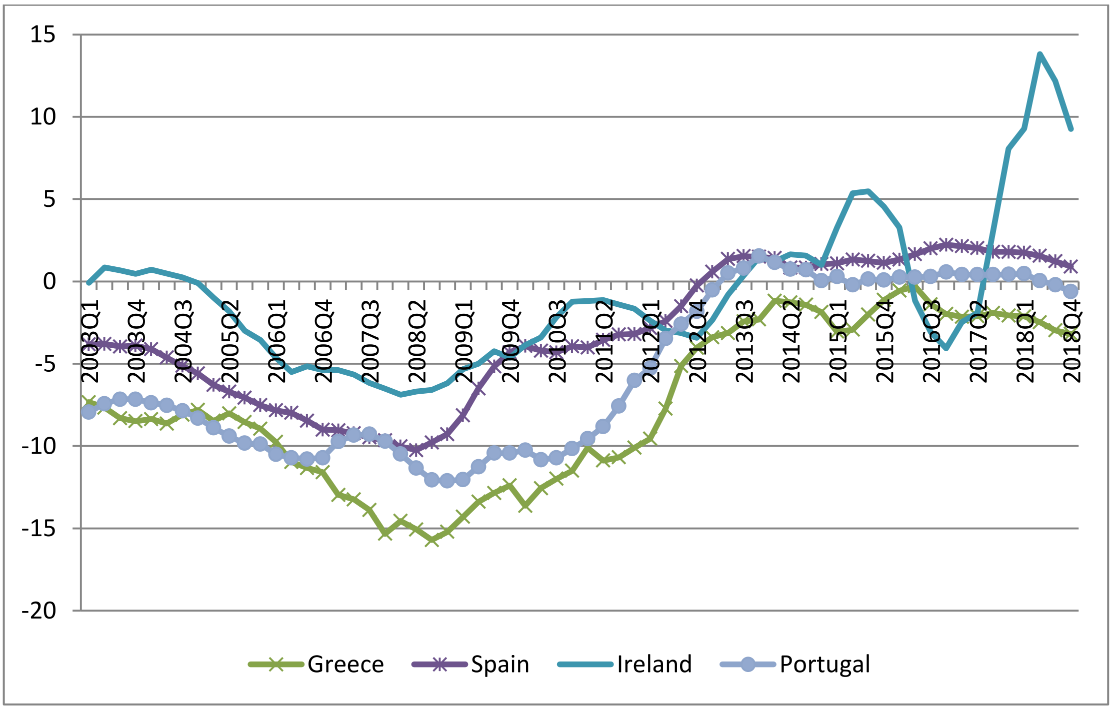
Figures and Tables

Figure 1a: Current account balances as % GDP in the North



Data source: Eurostat.

Figure 1b: Current account balances as % GDP in the South



Data source: Eurostat.

Figure 2: The first dynamic factor

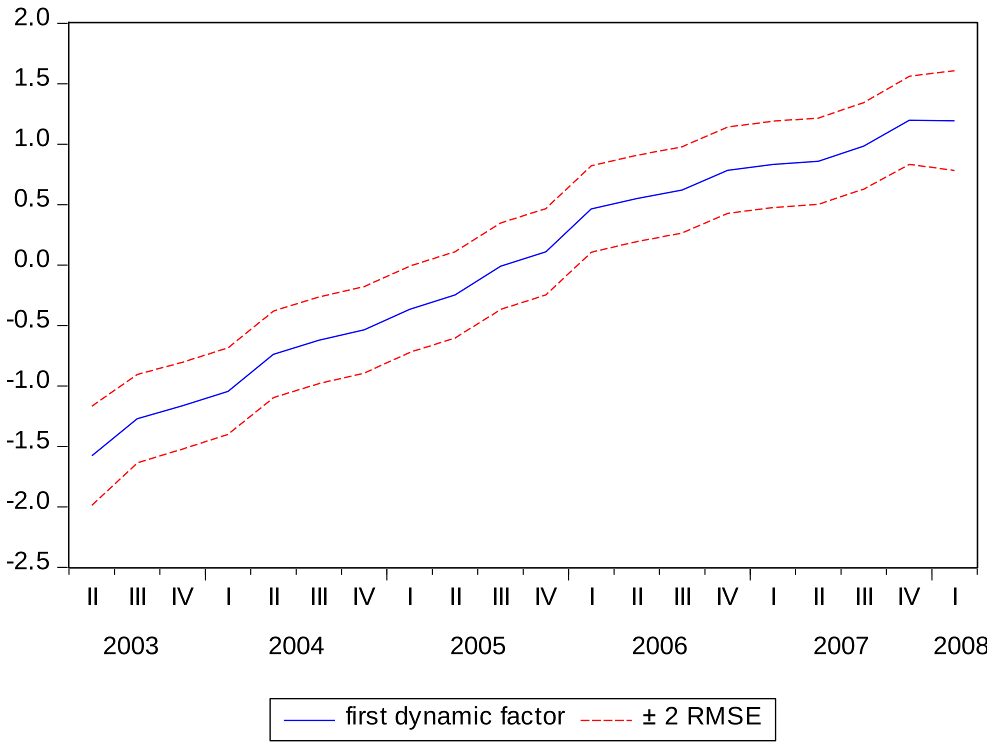


Figure 3: The second dynamic factor

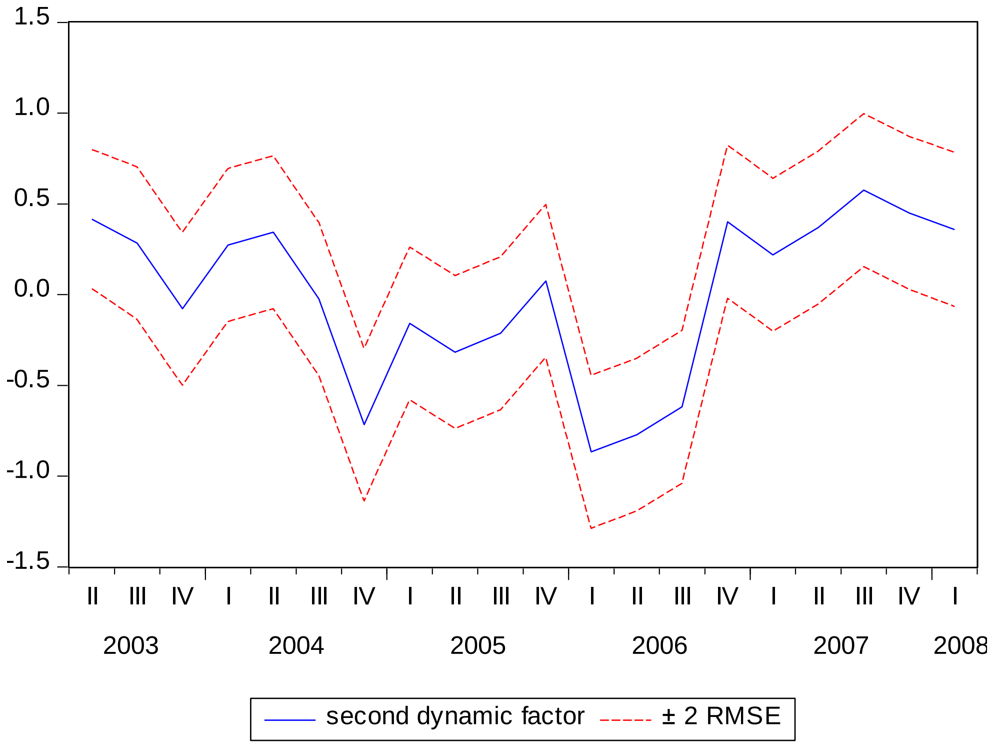


Figure 4: The first dynamic factor 2008-2018

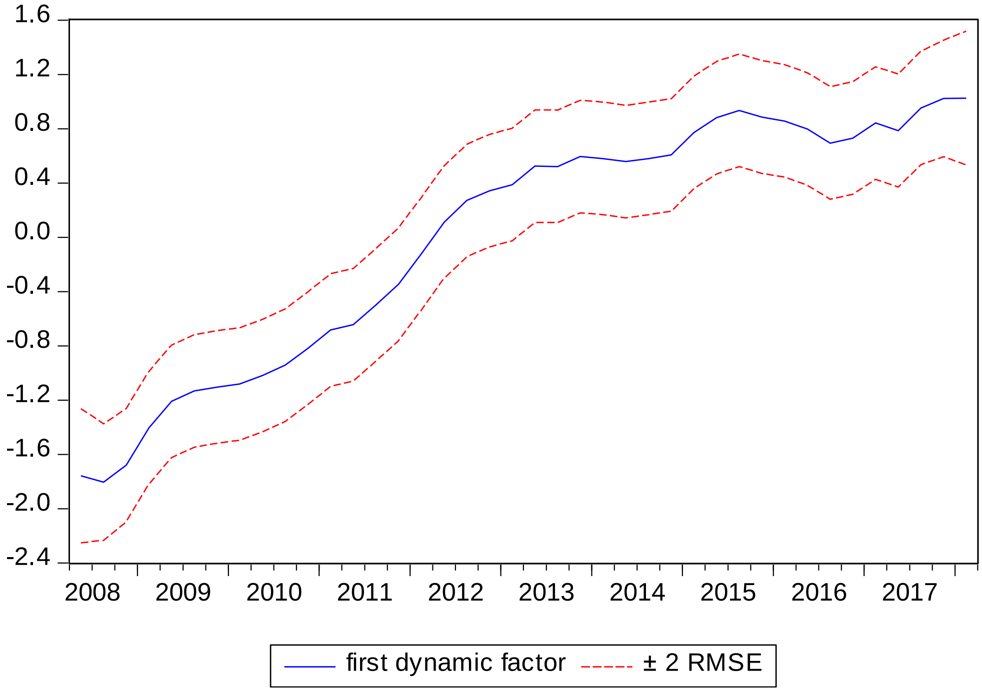


Figure 5: The second dynamic factor 2008-2018

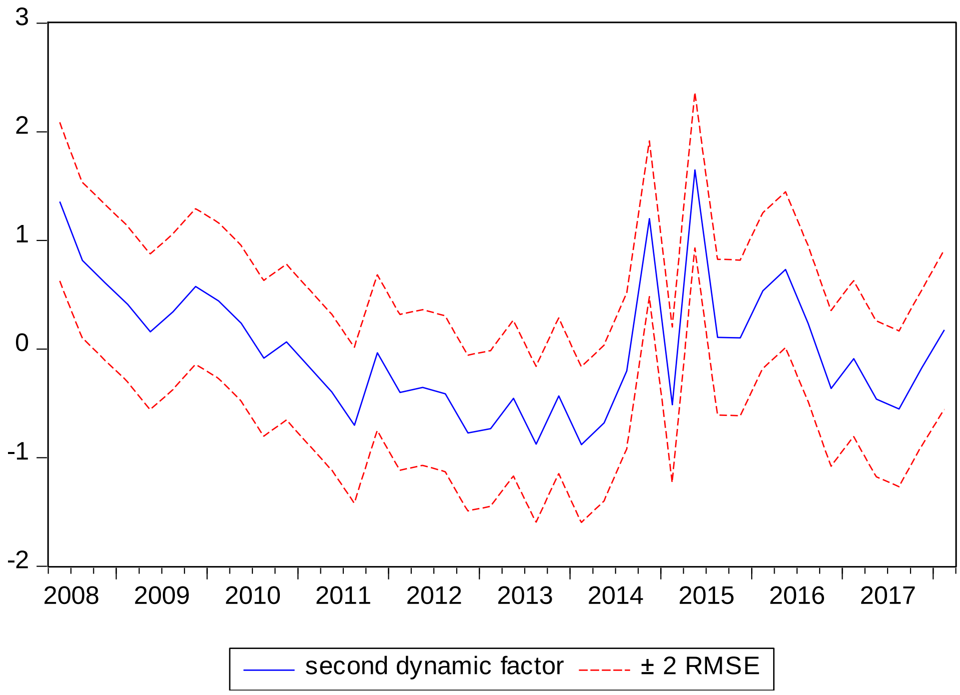


Table 1: Loading weights of the dynamic factor for each country.

	Coefficient	Z-Statistic	R^2
GE	1	-	0.82
AU	0.92	2.2	0.64
NL	0.41	1.9	0.01
IR	-1.0	3.9	0.9
PT	-0.75	3.2	0.6
GR	-0.99	3.5	0.7
ES	-1.0	1.5	0.96

The R^2 is derived from regressing the factor on the current-account balance (relative to GDP) for each country's current-account balance.

Table 2: Loading weights of the second dynamic factor for each country.

	Coefficient	Z-Statistic	Cumulated R^2
GE	1	-	0.83
AU	0.8	2.3	0.83
NL	-1.5	3.8	0.78
IR	0.17	0.3	0.9
PT	0.7	2.7	0.78
GR	-0.37	1.0	0.71
ES	-0.05	0.9	0.96

The R^2 is derived from a simple regression of the first and second factor onto each country's current-account balance.

Table 3: Loading weights of the dynamic factor for each country.

	Coefficient	Z-Statistic	R^2
GE	1	-	0.68
AU	-0.23	1.9	0.11
NL	0.04	1.7	0.14
IR	0.6	6.4	0.32
PT	1.0	3.1	0.93
GR	1.0	3.6	0.91
ES	0.96	4.0	0.92

The R^2 is derived from a simple regression of the factor on each country's current-account balance.

Table 4: Loading weights of the second dynamic factor for each country.

	Coefficient	Z-Statistic	Cumulated R^2
GE	1	-	0.81
AU	0.74	4.5	0.52
NL	-0.9	4.4	0.60
IR	-0.002	0.01	0.33
PT	-0.0001	0.002	0.94
GR	-0.06	0.01	0.92
ES	-0.1	0.4	0.92

The R^2 is derived from a simple regression of the first and second factor on each country's current-account balance.

Table 5: Forecasting the Dollar-Euro Exchange Rate

Period	2003q2-2008q1			2008q3-2018q1		
	AR(2)	One factor	Two factors	AR(2)	One factor	Two factors
constant	0.03(1.2)	0.09(2.44)	0.102(2.25)	0.037(2.3)	0.07(2.8)	0.093(4.2)
$LEURO_{t-1}$	1.15(4.9)	1.03(4.55)	0.984(4.3)	1.24(8.44)	1.19(8.17)	1.07(8.06)
$LEURO_{t-2}$	-0.24(1.1)	-0.39(1.84)	-0.386(1.6)	-0.41(2.85)	-0.49(3.23)	-0.473(3.7)
DF1	-	0.027(2.07)	0.03(2.2)	-	-0.02(1.66)	-0.035(3.1)
DF2	-	-	0.02(1.2)	-	-	-0.033(3.4)
RMSE	0.040623	0.043328	0.025974	0.036716	0.035343	0.0300568
MAE	0.032881	0.037143	0.022152	0.029788	0.028798	0.025673
Theil inequality	0.087914	0.087297	0.051694	0.072202	0.06947	0.060017

Dependent variable is the log of dollar-euro exchange rate. 't' statistics are in parenthesis. Forecast diagnostics are based on one step ahead static forecasts.

Data sources: Eurostat and ECB Statistical Data Warehouse.

Appendix 1

A step by step application showing the equivalence of the Kalman filter and Principal Components

In this appendix we give a worked example which shows the equivalence of our Kalman filter model and principal components. We take an artificial data set of 1000 observations and two variables where we construct the data in such a way that there is a common factor underlying both series but also a considerable amount of noise in each series. The example has been worked in EVIEWS and the EVIEWS workfile is available from the authors upon request.

We set up an EVIEWS workfile which is undated and has 1000 observations.

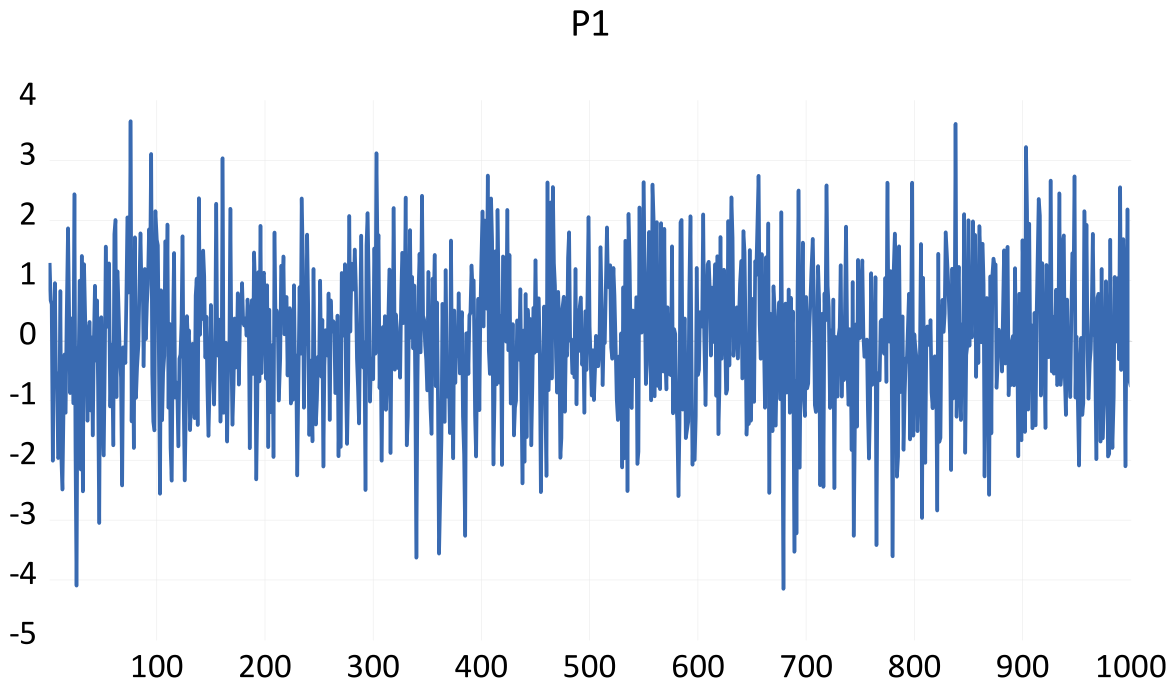
1. We generate three variables which are standard normal random numbers with N(0,1), $x_1, x_2 \text{ and } x_3$.
2. We then generate two variables from these three random variables as follows:

$$y_1 = x_1 + x_2$$
$$y_2 = x_1 + x_3$$

which means that both variables have a common factor but also quite a large idiosyncratic part.

3. We the apply standard principal components, using the EVIEWS procedure and derive the first principal component, P_1, which is graphed in Figure A1.

Figure A1: First Principal Component



4. We now turn to the Kalman filter procedure. We begin by creating a state space object (Sspace):

   4.1 The first step is to create two new series which are normalised to have a zero mean and unit variance. This is done simply as:

   $$ys_i = (y_i - \bar{y}_i) / SE_{y_i}$$

   This is important because we are trying to mimic what happens in PCA and the first stage here is to normalize each variable.

   4.2 We then set up the following state space form:

```
@signal y1s = sv1 + [var = 1]
@signal y2s =c(1)*sv1 + [var = 1]

@state sv1 =  [var = exp(c(2))]
```

   There are two signal equations for the two observed variables. The variance for the two signal equations is set to 1 to mimic what happens in principal components. The coefficient of the first signal equation is normalised to 1. There is an estimated parameter (c(1)) in the second signal equation and we estimate the variance of the state equation (c(2)). This is then estimated by maximising the likelihood function to produce the results in the Table A1.

Table A1: Kalman Filter Estimation

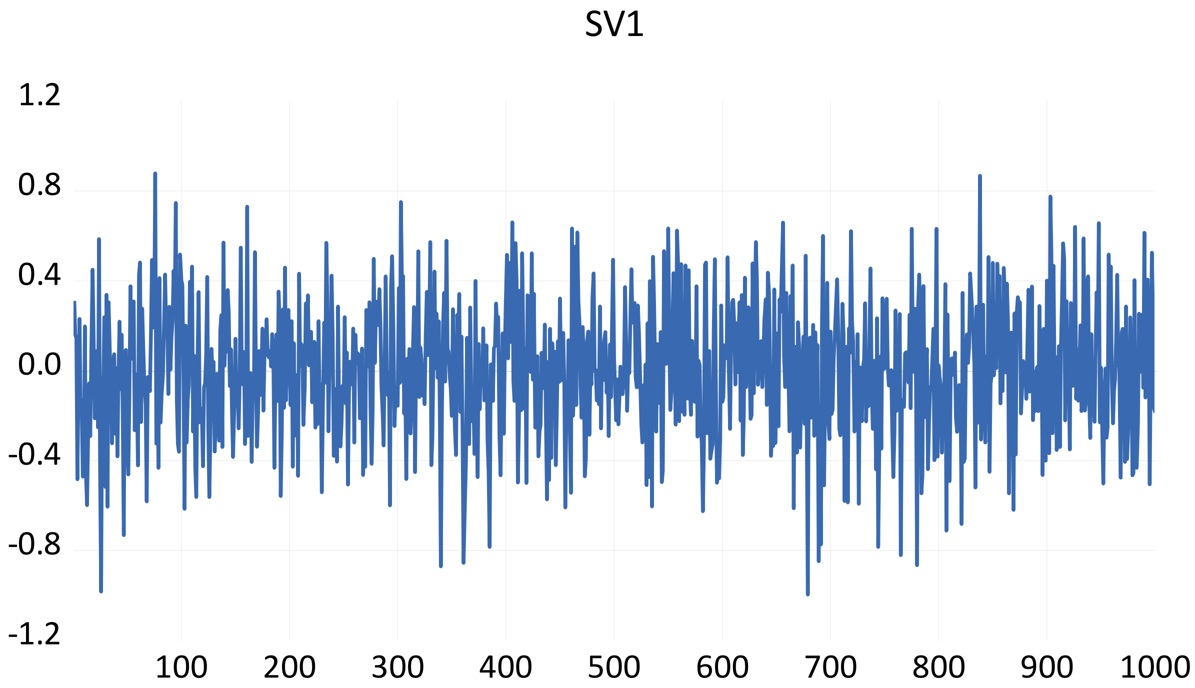
Sspace: SS01
Method: Maximum likelihood (BFGS / Marquardt steps)
Included observations: 1000
Convergence achieved after 16 iterations
Coefficient covariance computed using outer product of gradients

	Coefficient	Std. Error	z-Statistic	Prob.
Estimated parameter, C(1)	1.000	0.219	4.566	0.00
Variance of state equation, C(2)	-1.357	0.258	-5.254	0.00
	Final State	Root MSE	z-Statistic	Prob.
SV1	0.000	0.507	0.000	1.00

Log likelihood	-2787.14	Akaike info criterion	5.58
Parameters	2	Schwarz criterion	5.59
Diffuse priors	0	Hannan-Quinn criter.	5.58

4.3 We then form the smoothed state series sv1, which we present in Figure A2.

Figure A2: Smoothed State Series from Kalman Filter Estimation



4.4 This state variable now contains exactly the same information as the first principal components. We can demonstrate this by regressing SV1 on P1.

Table A2: Equivalence of Principal Components with Kalman Filter technique

Dependent Variable: P1
Method: Least Squares
Included observations: 1000

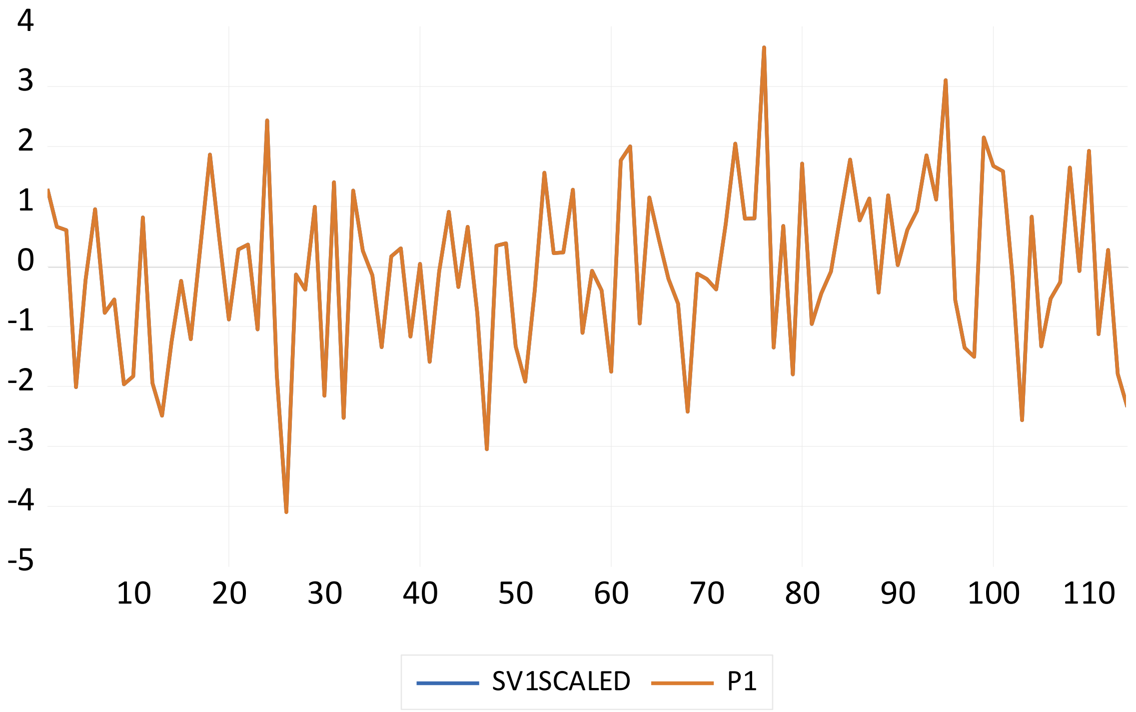
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.88E-07	1.41E-07	-1.332	0.18
SV1	4.164	4.78E-07	8711630.	0.00

R-squared	1.00	Mean dependent var	-3.86E-17
Adjusted R-squared	1.00	S.D. dependent var	1.231947
S.E. of regression	4.47E-06	Akaike info criterion	-21.797
Sum squared resid	1.99E-08	Schwarz criterion	-21.787
Log likelihood	10900.26	Hannan-Quinn criter.	-21.793
F-statistic	7.59E+13	Durbin-Watson stat	1.003
Prob(F-statistic)	0.00		

It should be noted that the R-squared is exactly equal to 1 as stated in the main text of the paper.

Finally, we rescale the state variable by 1.2921020/0.310325 to make the first observation of SV1 equal the first principal component. We then graph the two variables together. For clarity we show the first 110 observations only in the Figure A3.

Figure A3: The Two Series Compared



It is impossible to see any difference between the two scaled series. Looking at the spread sheet, the two series are equivalent to at least the 4th decimal point.

BANK OF GREECE WORKING PAPERS

265. Katseli T. L., A. Theofilakou and K-M Zekente, "Central Bank Independence and Inflation Preferences: New Empirical Evidence on the Effects on Inflation", July 2019.

266. Panagiotou S. and H. Thanopoulou, "Tonnage Tax Revisited: The Case of Greece During a Shipping Crisis and an economic crisis period", July 2019.

267. Economides, G. and A. Xepapadeas, "The Effects of Climate Change on a Small Economy", September 2019.

268. Bragoudakis, Z., S. Degiannakis and G. Filis, "Oil and Pump Prices: Is There Any Asymmetry in the Greek Oil Downstream Sector?, September 2019.

269. Dellis, K., "Financial System Heterogeneity and FDI Flows: Evidence from OECD Economies", September 2019.

270. Gibson H., S. Hall, G. Tavlas and P. Petroulas, "On the effects of the ECB's funding policies on bank lending and the demand for the euro as an international reserve", September 2019.

271. Karadima, M. and E. Louri-Dendrinou "Non-Performing Loans in the Euro Area: Does Market Power Matter?", September 2019.

272. Papandreou, A. "Stranded Assets and the Financial System", November 2019.

273. George S. Tavlas, "A Reconsideration of the Doctrinal Foundations of Monetary-Policy Rules: Fisher versus Chicago", November 2019.

274. Bragoudakis, Z. and D. Sideris, "Asymmetric Price Adjustment and the Effects of Structural Reforms in a Low Income Environment: the Case of the Gasoline Market in Greece", December 2019.

275. Panayiotis C. A., and S. Anyfantaki, "Financial Literacy and its Influence on Consumers' Internet Banking Behaviour, December 2019.

276. Balfoussia, H., H. Dellas, and D. Papageorgiou, "Fiscal Distress and Banking Performance: the Role of Macroprudential Regulation," December 2019.

277. Anastasiou, D., Z. Bragoudakis and S. Giannoulakis, "Perceived Vs Actual Financial Crisis and Bank Credit Standards: Is There Any Indication Of Self-Fulfilling Prophecy?", January 2020.

278. Gibson, D. H., S. G. Hall, P. Petroulas, V. Spiliotopoulos and G.S. Tavlas, "The Effect of Emergency Liquidity Assistance (ELA) on Bank Lending During the Euro Area Crisis", January 2020.

279. Tavlas, G., "On the Controversy over the Origins of the Chicago Plan for 100 Percent Reserves", June 2020.

280. Milionis E. A, N.G. Galanopoulos, A Study of the Effect of Data Transformation and «Linearization» on Time Series Forecasts. A Practical Approach, June 2020.

281. Gibson, D. H., S.G. Hall, D. GeFang, P. Petroulas, G.S. Tavlas, Did the Absence of A Central Bank Backstop in the Sovereign Bond Markets Exacerbate Spillovers During the Euro-Area Crisis?, July 2020.