Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections

Marta Bańbura, European Central Bank
Domenico Giannone, Université Libre de Bruxelles, ECARES and CEPR
Michele Lenza, European Central Bank and ECARES

June 10, 2014

Abstract

This paper describes an algorithm to compute the distribution of conditional forecasts, i.e. projections of a set of variables of interest on future paths of some other variables, in dynamic systems. The algorithm is based on Kalman filtering methods and is computationally viable for large vector autoregressions (VAR) and dynamic factor models (DFM). For a quarterly data set of 26 euro area macroeconomic and financial indicators, we show that both approaches deliver similar forecasts and scenario assessments. In addition, conditional forecasts shed light on the stability of the dynamic relationships in the euro area during the recent episodes of financial turmoil and indicate that only a small number of sources drive the bulk of the fluctuations in the euro area economy.

JEL Classification: C11, C13, C33, C53

Keywords: Vector Autoregression, Bayesian Shrinkage, Dynamic Factor Model, Conditional Forecast, Large Cross-Sections.

* We would like to thank Todd Clark, Marek Jarociński and Lutz Kilian for their comments. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank and the Eurosystem. Please address any comments to Marta Bańbura marta.banbura@ecb.europa.eu, Domenico Giannone dgiannon@ulb.ac.be; or Michele Lenza michele.lenza@ecb.europa.eu.

1 Introduction

Vector autoregressions (VARs) are very flexible and general models and provide a reliable empirical benchmark for alternative econometric representations such as dynamic stochastic general equilibrium (DSGE) models, which are more grounded in theory but, at the same time, impose more structure on the data (see, for example, Christiano, Eichenbaum, and Evans, 1999).

Recent literature has shown that VARs are viable tools also for large sets of data (see Bańbura, Giannone, and Reichlin, 2010). In this paper, we construct a large VAR for the euro area and we apply it to unconditional forecasting as well as for conditional forecasts and scenarios. These, along with structural analysis (assessing the effects of structural shocks), have been the main applications of VARs. Whereas large VARs have been used for unconditional forecasting and structural analysis,[1] limited attention has been devoted as yet to conditional forecasting. This is because popular algorithms for deriving conditional forecasts have been computationally challenging for large data sets. We overcome this problem by computing the conditional forecasts recursively using Kalman filtering techniques.

Conditional forecasts and, in particular, scenarios are projections of a set of variables of interest on future paths of some other variables. This is in contrast to unconditional forecasts, where no knowledge of the future path of any variables is assumed. The prior knowledge, albeit imperfect, of the future evolution of some economic variables may carry information for the outlook of other variables. For example, future fiscal packages would affect the future evolution of economic activity and, thus, might provide important off-model information. Moreover, it may be of interest to assess the impact of specific future events on a set of variables, i.e. to conduct scenario analysis. Notable examples of the latter are the stress tests recently conducted in the US and the euro area in order to assess the vulnerability of their banking systems. For recent examples of conditional forecasts, see Lenza, Pill, and Reichlin (2010); Giannone, Lenza, and Reichlin (2010); Jarociński and Smets (2008); Bloor and Matheson (2011); Giannone, Lenza, Pill, and Reichlin (2012); Stock and Watson (2012a); Baumeister and Kilian (2011); Giannone, Lenza, Momferatou, and Onorante (2013).

[1] See e.g. Koop (2013), Giannone, Lenza, Momferatou, and Onorante (2013), Giannone, Lenza, and Reichlin (2012), Paciello (2011).

For VAR models, the conditional forecasts are typically computed by using the algorithm developed by Waggoner and Zha (1999). Due to computational burden, the latter approach can easily become impractical or unfeasible for high dimensional data and long forecast horizons, even if the computationally more efficient version of Jarociński (2010) is employed. However, many problems in macroeconomics and finance can only be addressed by looking at the joint dynamic behavior of a large number of time series. For example, business cycle research, as in the NBER tradition, typically involves the analysis of many macroeconomic variables. Professional forecasters and policymakers look at a variety of different indicators to predict key variables of interest and make their decisions. Investors analyze the joint behavior of many asset returns in order to choose their optimal portfolios. More in general, contemporary science relies more and more on the availability and exploitation of large data sets.

In this paper, building on an old insight by Clarida and Coyle (1984), we propose an algorithm based on Kalman filtering techniques to compute the conditional forecasts. Since the Kalman filter works recursively, i.e. period by period, this algorithm reduces significantly the computational burden and is particularly well suited for empirical approaches handling large data sets. Using a simulation smoother (see Carter and Kohn, 1994; de Jong and Shephard, 1995; Durbin and Koopman, 2001a, for examples of simulation smoothers) allows for the computation of the full distribution of conditional forecasts. The algorithm applies to any model which can be cast in a state space representation.

The interest in issues which are best addressed by considering large information sets raises a trade-off between excessive simplicity of the models – misspecification due to omitted variables – and their excessive complexity – many free parameters leading to large estimation uncertainty. Recent developments in macroeconometrics have suggested two approaches to deal with the complexity of large sets of data, without losing their salient features: Bayesian VARs (BVARs) and dynamic factor models (DFMs).

The aforementioned flexibility of VARs comes at the cost of a high number of free parameters to be estimated. Specifically, for a generic VAR(p) model for a vector of n variables $y_t = (y_{1,t}, \ldots, y_{n,t})'$:

$$y_t = c + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t \,, \qquad \varepsilon_t \sim WN(0, \Sigma) \,, \tag{1}$$

where $WN(0, \Sigma)$ refers to a white noise process with mean 0 and covariance matrix Σ, we

count: i) pn^2 parameters in autoregressive matrices, $A_1, \ldots, A_p$, that are of dimension $n \times n$ each; ii) $n(n+1)/2$ free parameters in the $n \times n$ covariance matrix of residuals Σ; iii) n parameters in the constant term c. The number of parameters proliferates as the number of variables in the model increases, making estimation unfeasible or unreliable. For example, when the number of variables in a VAR with 4 lags increases from 6, as in the original VAR model proposed by Sims (1980), to 20, 50 or 100, the total number of parameters to be estimated goes from 171 to, respectively, numbers in the order of 2, 10 and 50 thousands. Such a high number of parameters cannot be well estimated by ordinary least squares, for example, since the typical macroeconomic sample involves a limited number of data points (in the best case, $50-60$ years of data, i.e. $200-250$ data points with quarterly data). The problem of parameter proliferation that prevents econometricians from conducting reliable inference with large dimensional systems is also known as the "curse of dimensionality".

A solution to the curse of dimensionality in the VAR framework consists in adopting Bayesian shrinkage. The idea of this method is to combine the likelihood coming from the complex and highly parameterised VAR model with a prior distribution for the parameters that is naïve but enforces parsimony. As a consequence, the estimates are "shrunk" toward the prior expectations, which are typically equal to 0.[2] This approach can also be interpreted as a penalised maximum likelihood method, with a penalisation on large (in absolute value) coefficients.

The shrinkage methods have been advocated by early proponents of VARs as a macroeconometric tool (Litterman, 1979; Sims, 1980; Doan, Litterman, and Sims, 1984) but they were typically used for low dimensional systems. Recently, it has been shown that the idea of shrinkage works also for high dimensional systems and provides results that are very similar to those obtained by using the DFMs (see De Mol, Giannone, and Reichlin, 2008; Bańbura, Giannone, and Reichlin, 2010; Giannone, Lenza, and Primiceri, 2012). This is not surprising since, as shown by De Mol, Giannone, and Reichlin (2008), when applied to collinear variables, as are typically macroeconomic variables, the forecasts produced by factor models and Bayesian shrinkage tend to get closer, as the size of the sample and of the cross-section get larger.

[2]For an extensive discussion of shrinkage in various contexts see e.g. Stock and Watson (2012b) and Ng (2013).

Factor models exploit the fact that macroeconomic and financial time series are characterised by strong cross-sectional correlation. Under the assumption that most of the fluctuations are driven by a relatively limited set of common sources, factor models offer a parsimonious representation by summarizing the information from a large number of data series in few common factors. DFMs further parameterize the dynamics of the factors, typically assuming a VAR process. The estimation of factor models generally requires that the data are stationary. Assuming that stationarity is achieved via taking first differences[3], the DFM is defined as follows:

$$\begin{aligned} \Delta y_t &= \mu + \Lambda F_t + e_t, \\ F_t &= \Phi_1 F_{t-1} + \ldots + \Phi_p F_{t-s} + u_t\,, \qquad u_t \sim WN(0, Q)\,, \end{aligned}$$

where $F_t = (F_{1,t}, \ldots, F_{r,t})'$ is an r-dimensional vector of common factors, with r much smaller than n and Λ is an $n \times r$ matrix of factor loadings. Since the number of common factors r is typically small, the estimation of the VAR describing the dynamics of the common factors does not pose any problem. The residual $e_t = (e_{1,t}, \ldots, e_{n,t})'$ is the idiosyncratic component. The most common approach is to assume that the idiosyncratic component is cross-sectionally uncorrelated. This assumption gives rise to the "exact" factor model, which highlights the fact that the cross-correlation between the variables is fully accounted for by the common factors. Interestingly, recent literature has shown that factor models can be estimated with large data sets, i.e. even in situations in which the cross-sectional dimension n is much larger than the sample size T. In addition, the estimates are asymptotically valid also when the data generating process is not the "exact" but rather an "approximate" factor model, in the sense that the idiosyncratic components are weakly cross-correlated (see Forni, Hallin, Lippi, and Reichlin, 2000; Stock and Watson, 2002b; Bai and Ng, 2002; Bai, 2003; Forni, Hallin, Lippi, and Reichlin, 2004; Doz, Giannone, and Reichlin, 2012). Stock and Watson (2011) provide an exhaustive survey of the literature.

Factor models are appealing also because many popular economic models can be cast in their format. The typical theoretical macro model, indeed, includes only a handful of shocks driving the key aggregate variables in the economy. The arbitrage pricing theory (APT) is built upon the existence of a set of common factors underlying all returns. Moreover, the distinction

[3]Trending series are typically "logged" beforehand.

between common and idiosyncratic sources of fluctuations is often employed in international, regional and sectorial studies and represents a useful device to study macroeconomic implications of microeconomic behavior (see e.g. Kose, Otrok, and Whiteman, 2003; Foerster, Sarte, and Watson, 2011).

In our empirical application, we carry out a comprehensive comparison of the two techniques just described, on a large data set of quarterly euro area macroeconomic and financial variables. We consider two versions of the BVAR – with data in (log-)levels and in (log-)differences – and a dynamic factor model. First, we show that the three models produce quite accurate unconditional forecasts, compared to univariate benchmarks, and that the forecasts from the three approaches are very correlated. The latter finding lends empirical support to the theoretical argument that the approaches are tightly linked, complementing similar evidence already available for the US (see, for example De Mol, Giannone, and Reichlin, 2008; Giannone, Lenza, and Primiceri, 2012).

Then, we compare the two approaches also for what concerns scenarios and conditional forecasts. In particular, we study the economic developments associated to a scenario of an increase in world GDP as well as conditional forecasts based on the realised path of real GDP, consumer prices and the policy rate. We show that also the scenario analysis and the conditional forecasts computed for the three models provide similar insights. The fact that the results are not model specific is reassuring, since it indicates that the predictions of the models are reflecting genuine data features.

The results from the conditional forecasts yield support to two further conclusions. First, the fact that the conditional forecasts based on the three variables track, in general, quite closely the actual developments in most of the variables under analysis suggests that there are only a few "sources" that drive the bulk of the fluctuations in the euro area economy. Second, there appears to be some degree of stability in the economic relationships following the financial crisis as the the conditional forecasts for this period based on the parameters estimated with data until end of 2007 are relatively accurate, with the possible exception of some categories of loans and broad monetary aggregates (see Giannone, Lenza, and Reichlin, 2012, for an extensive discussion and interpretation of these results).

The structure of the paper is as follows. In section 2, we review the state-of-the-art techniques

for the estimation and inference for DFMs and BVARs and we expound the close relationship linking the two approaches. In section 3, we describe a Kalman filter based methodology to compute conditional forecasts. In section 4, we present and discuss the empirical results. Section 5 concludes. The appendix contains some implementation details and data descriptions.

2 Models for large data sets

2.1 Dynamic factor models

The general representation of the dynamic factor model described in the introduction is:

$$\begin{aligned} \Delta y_t &= \mu + \Lambda F_t + e_t, \\ F_t &= \Phi_1 F_{t-1} + \ldots + \Phi_s F_{t-s} + u_t\,, \qquad u_t \sim WN(0, Q)\,. \end{aligned} \tag{2}$$

Following Doz, Giannone, and Reichlin (2012) the model can be estimated by means of quasi-maximum likelihood methods. In this context, the estimation of the model is performed by maximising a likelihood function, under the assumption that data are Gaussian and that the factor structure is exact, i.e. the idiosyncratic errors are cross-sectionally orthogonal: $u_t \sim$ i.i.d.$\mathcal{N}(0, Q)$ and $e_t \sim$ i.i.d.$\mathcal{N}(0, \Gamma_d)$, where Γ_d is a diagonal matrix.

Doz, Giannone, and Reichlin (2012) have shown that this estimation procedure provides consistent estimates for approximate dynamic factor models under general regularity conditions (convergence in probability of the covariance matrix of the data and data stationarity). Remarkably, consistency is achieved without any constraint on the number of variables, n, relative to the sample size, T, under the assumption of weak cross-sectional dependence of the idiosyncratic term, e_t, and of sufficient pervasiveness of the common factors.

As the factors are unobserved, the maximum likelihood estimators of the parameters Λ, Γ_d, $\Phi_1, \ldots, \Phi_s, Q$, which we collect in θ, are, in general, not available in closed form. They can be obtained either via a direct numerical maximisation of the likelihood, which can be computationally demanding,[4] or, as in Doz, Giannone, and Reichlin (2012), via the Expectation-

[4] Jungbacker and Koopman (2008) show how to reduce the computational burden in case the number of observables is much larger than the number of factors.

Maximisation (EM) algorithm. The EM algorithm was proposed by Dempster and Rubin (1977) as a general solution to problems with incomplete or latent data. In the case of the DFM, the algorithm alternates between the use of the Kalman smoother to estimate the common factors given a set of parameters (E-step), and multivariate regressions (corrected for the uncertainty in the estimation of the common factors) to estimate the parameters given the factors (M-step), see e.g. Watson and Engle (1983) or Shumway and Stoffer (1982).

The algorithm can be initialised using the sample principal components. In what follows, we assume that data are standardised to have sample mean equal to zero and variance equal to one.[5] Denote by d_j, $j = 1, \dots n$, the eigenvalues of $\frac{1}{T}\sum_{t=1}^{T} \Delta y_t \Delta y_t'$ and by v_j, $j = 1, \dots n$, the associated eigenvectors, i.e.

$$\left(\frac{1}{T}\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right) v_j = v_j d_j, \quad j = 1, 2, \dots n\,,$$

with $v_j' v_j = 1$, $v_j' v_k = 0$ for $j \neq k$ and $d_1 \geq d_2 \geq \ldots \geq d_n$. The sample principal components of Δy_t are defined as $z_{jt} = \frac{1}{\sqrt{d_j}} v_j' \Delta y_t$. The principal components are ordered accordingly to their ability to explain the variability in the data as the total variance explained by each principal component is equal to d_j. The principal components transform cross-sectionally correlated data, Δy_t, into linear combinations $z_t = (z_{1,t}, ..., z_{n,t})' = H \Delta y_t$ where $H = \left(\frac{1}{\sqrt{d_1}} v_1, \dots, \frac{1}{\sqrt{d_n}} v_n\right)'$. These linear combinations are cross-sectionally uncorrelated, with unit variance $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' = I_n$.

The approximate factor structure is defined in terms of behavior of the eigenvalues of the population covariance matrix when the number of variables increases. Specifically, the first r eigenvalues of the population covariance matrix of Δy_t are assumed to grow with the dimension of the system, at a rate n. All the remaining eigenvalues remain instead bounded. It can be proved that these assumptions imply that the eigenvalues d_j of the sample covariance matrix will go to infinity at a rate n for $j = 1, \dots, r$, where r is the number of common factors. On the other hand, $d_{r+1}, \dots, d_n$ will grow at a rate given by $n/\sqrt{T}$ (see De Mol, Giannone, and

[5]The zero mean assumption is without loss of generality, since it is equivalent to concentrating out the mean. Since maximum likelihood estimates are scale invariant, rescaling the data does not affect the estimates. On the other hand, homogeneity of the scale across variables is convenient, since the algorithm for maximizing the likelihood is more efficient from a computational standpoint. In addition, working with standardised data is useful since the initialisation of the algorithm is based on principal components, which are not scale invariant. Once the estimates are obtained, the factor loadings, $\hat{\Lambda}$, and the covariance matrix of the idiosyncratic components, $\hat{\Gamma}_d$, can be obtained by simple rescaling.

Reichlin, 2008; Doz, Giannone, and Reichlin, 2011, 2012). Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002a,b) have shown that if data have an approximate factor structure[6], then the first r principal components can approximate well the space spanned by the unobserved common factors, when the sample size and the cross-sectional dimension are large.

The sample principal components offer thus good starting values for the common factors: $\hat{F}_t^{(0)} = z_t$.[7] The starting values for the parameters of the model, $\theta^{(0)}$, can then be estimated by means of OLS techniques, by treating the principal components as if they were the true factors. Once the parameters have been estimated, we can estimate a new set of factors by using the Kalman smoother: $\hat{F}_t^{(1)} = \mathrm{E}_{\theta^{(0)}}[F_t|\Delta y_1, \ldots, \Delta y_T]$. At this stage, we have the two-step procedure of Doz, Giannone, and Reichlin (2011). The quasi-maximum likelihood estimation via the EM algorithm consists essentially in iterating these steps until convergence. Details are reported in the appendix.

2.2 Bayesian vector autoregressions

The general representation of the VAR model described in the introduction is:

$$y_t = c + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma).$$

We consider conjugate priors belonging to the normal-inverse-Wishart family, where the prior for the covariance matrix of the residuals is inverse-Wishart and the prior for the autoregressive coefficients is normal. The priors are a version of the so-called Minnesota prior, originally due to Litterman (1979), which is centered on the assumption that each variable follows an independent random walk process, possibly with drift:

$$y_t = c + y_{t-1} + \varepsilon_t,$$

which is a parsimonious yet "reasonable approximation of the behavior of an economic variable".

[6] As stressed above, this amounts to assume that the idiosyncratic components are weakly cross-correlated.

[7] In fact, under the assumption that $\Phi_1 = \ldots = \Phi_s = 0$ and $\Gamma_d = \bar{\gamma} I_n$ (i.e. homoscedastic idiosyncratic components) the quasi-maximum likelihood solution is analytical, with the expected value for the factors proportional to the principal components of the data.

For the prior on the covariance matrix of the errors, Σ, we set the degrees of freedom equal to $n+2$, which is the minimum value that guarantees the existence of the prior mean, which we set as $\mathrm{E}[\Sigma] = \Psi$, where Ψ is diagonal.

The prior moments for the VAR coefficients are as follows:

$$\begin{aligned} E\left[(A_s)_{ij} \,|\Sigma, \lambda, \Psi\right] &= \begin{cases} 1 & \text{if } i=j \text{ and } s=1 \\ 0 & \text{otherwise} \end{cases} \\ cov\left[(A_s)_{ij}, (A_r)_{hm} \,|\Sigma, \lambda, \Psi\right] &= \begin{cases} \lambda^2 \frac{1}{s^2} \frac{\Sigma_{ih}}{\psi_j} & \text{if } m=j \text{ and } r=s \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

Notice that the variance of this prior is lower for the coefficients associated with more distant lags, and that coefficients associated with the same variable and lag in different equations can be correlated. Finally, the key hyperparameter λ controls the scale of all the variances and covariances, and effectively determines the overall tightness of this prior. The terms Σ_{ij}/ψ_j account for the relative scale of the variables. The prior for the intercept, c, is diffuse.[8]

We include an additional prior, which implements a so-called "inexact differencing" of the data. More precisely, rewrite the VAR equation in an error correction form:

$$\Delta y_t = c + \Pi y_{t-1} + B_1 \Delta y_{t-1} + \ldots + B_{\tilde{p}} \Delta y_{t-\tilde{p}} + \varepsilon_t.$$

where $\tilde{p} = p-1$, $B_s = -A_{s+1} - \ldots - A_p$, $s = 1, \ldots, \tilde{p}$ and $\Pi = A_1 + \ldots + A_p - I_n$.

A VAR in first differences implies the restriction $\Pi = 0$ (or $A_1 + \ldots + A_p = I_n$). We follow Doan, Litterman, and Sims (1984) and set a prior centered at 1 for the sum of coefficients on own lags for each variable, and at 0 for the sum of coefficients on other variables' lags. This prior introduces correlation among the coefficients on each variable in each equation. The tightness of this additional prior is controlled by the hyperparameter μ. As μ goes to infinity the prior becomes diffuse while, as it goes to 0, we approach the case of exact differencing, which implies the presence of a unit root in each equation.

The setting of the priors importantly depends on the hyperparameters λ, μ and Ψ, which reflect the informativeness of the prior distributions for the model coefficients. These heper-

[8]Koop (2013) considers non-conjugate priors which allow for exclusion of certain variables from some equations, however, he finds that these do not outperform simpler Minnesota priors in terms of forecast accuracy. Carriero, Clark, and Marcellino (2012) find that allowing for stochastic volatility helps to improve forecast accuracy. See Karlsson (2013) for a comprehensive overview of Bayesian methods for inference and forecasting with VAR models.

parameters have been usually set on the basis of subjective considerations or rules-of-thumb. Instead, we closely follow the theoretically grounded approach proposed by Giannone, Lenza, and Primiceri (2012). This involves treating the hyperparameters as additional parameters, in the spirit of hierarchical modelling. As hyperpriors (i.e. prior distributions for the hyperparameters), we use proper but quite disperse distributions. The implementation details are reported in the appendix.

2.3 Bayesian vector autoregression and dynamic factor model

The connection between Bayesian shrinkage and dynamic factor models is better understood by focusing on the data that have been transformed to achieve stationarity, Δy_t, and that have been standardised to have mean zero and unit variance.

The VAR in differences can be represented by:

$$\Delta y_t = B_1 \Delta y_{t-1} + \ldots + B_{\tilde{p}} \Delta y_{t-\tilde{p}} + \varepsilon_t.$$

Imposing that the level of each variable y_t follows an independent random walk process, is equivalent to imposing that its difference, Δy_t, follows an independent white noise process.

Consequently, the prior on the autoregressive coefficients can be characterised by the following first and second moments:

$$\begin{aligned} E\left[(B_s)_{ij} \,|\Sigma, \lambda, \Psi\right] &= 0 \\ cov\left[(B_s)_{ij}, (B_r)_{hm} \,|\Sigma, \lambda, \Psi\right] &= \begin{cases} \lambda^2 \frac{1}{s^2} \frac{\Sigma_{ih}}{\psi_j} & \text{if } m = j \text{ and } r = s \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

Since the variables are rescaled to have the same variance, the hyperparameter related to the scale can be set to be the same for all variables, i.e. $\Psi = \bar{\psi} I_n$.

Hence, the model can be rewritten in terms of the principal components described in section 2.1:

$$\Delta y_t = B_1 H^{-1} z_{t-1} + \ldots + B_{\tilde{p}} H^{-1} z_{t-\tilde{p}} + \varepsilon_t,$$

where $z_t = H \Delta y_t$ are the ordered principal components.

Interestingly, the prior set-up that imposes a uniform shrinkage on the parameters is equivalent to imposing a non-uniform degree of shrinkage on principal components:

$$\begin{aligned} E\left[\left(B_s H^{-1}\right)_{ij} | \Sigma, \lambda, \bar{\psi}\right] &= 0 \\ cov\left(\left(B_s H^{-1}\right)_{ij}, \left(B_r H^{-1}\right)_{hm} | \Sigma, \lambda, \bar{\psi}\right) &= \begin{cases} (\lambda^2 d_j)\frac{1}{s^2}\frac{\Sigma_{jh}}{\bar{\psi}} & \text{if } m = j \text{ and } r = s \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

In fact, the prior variance for the coefficients on the j^{th} principal component turns out to be proportional to the variance explained by the latter (d_j).

As discussed in section 2.1, if the data are characterised by a factor structure then, as the number of variables and the sample size increase, d_j will go to infinity at a rate n for $j = 1, \ldots, r$, where r is the number of common factors. On the other hand, $d_{r+1}, \ldots, d_n$ will grow at a slower rate which cannot be faster than $n/\sqrt{T}$. As a consequence, if λ goes to zero at a rate that is faster than that for the smaller eigenvalues and slower than for the largest eigenvalues, i.e. $\lambda^2 = \kappa\frac{\sqrt{T}}{n}\frac{1}{T^\delta}$ with $0 < \delta < 1/2$ and κ an arbitrary constant, then $\lambda^2 d_j$ will go to infinity for $j = 1, \ldots, r$. Hence the prior on the coefficients associated with the first r principal components will become flat. Instead, for $j > r$ $\lambda^2 d_j$ will go to zero, i.e. the coefficients related to the principal components associated with the bounded eigenvalues will be shrunk to zero.

De Mol, Giannone, and Reichlin (2008) have shown that, if the data are generated accordingly to a dynamic factor model and the hyperparameter λ is set according to the rate described above, the point forecasts obtained by using shrinkage estimators converge to the unfeasible optimal forecasts that would be obtained if the common factors were observed.

3 Conditional forecasts for state space representations

3.1 State space representation

Several univariate and multivariate time-series models may be cast in a state space representation. For the sake of notation, the generic state space form is defined as[9]:

[9]See Harvey (1989) for a thorough treatment of state space techniques. To simplify notation we abstract from exogenous variables as they are not included in our empirical model.

Measurement equation

$$Z_t = C_t S_t + v_t \tag{3}$$

Transition equation

$$S_t = G_t S_{t-1} + w_t \tag{4}$$

where $Z_t = (Z_{1,t}, Z_{2,t}, \ldots, Z_{k,t})'$ is a k-dimensional vector of observables, S_t an m-dimensional vector of potentially unobserved states, v_t and u_t two vectors of errors with: $\mathrm{E}[v_t] = 0$, $\mathrm{E}[v_t v_t'] = R_t$, $\mathrm{E}\left[v_t v_{t-s}'\right] = 0 \ \forall s \neq 0$, $\mathrm{E}[w_t] = 0$, $\mathrm{E}[w_t w_t'] = H_t$, $\mathrm{E}\left[w_t w_{t-s}'\right] = 0 \ \forall s \neq 0$ and $\mathrm{E}[v_t w_s'] = 0 \ \forall \ t, s$. Finally, C_t and G_t are two, respectively, $k \times m$ and $m \times m$ matrices of potentially time-varying coefficients.

The dynamic factor model in (2) can be cast in the representation (3)-(4) with $Z_t := \Delta y_t$, $C_t := (\Lambda, \ 0_{n \times r(s-1)}, \ I_n)$, $R_t := \Gamma_d$ and

$$S_t := \begin{pmatrix} F_t \\ \vdots \\ F_{t-s+1} \\ \mu \end{pmatrix}, \quad G_t := \begin{pmatrix} \Phi_1 & \Phi_2 & \ldots & \Phi_s & 0_{r \times n} \\ I_r & 0_r & \ldots & 0_r & \vdots \\ \vdots & \ddots & \ldots & \vdots & \vdots \\ 0_r & \ldots & I_r & 0_r & 0_{r \times n} \\ 0_{n \times r} & \ldots & \ldots & 0_{n \times r} & I_n \end{pmatrix}, \quad H_t := \begin{pmatrix} Q & \ldots & 0_{r \times n} \\ \vdots & \ddots & \vdots \\ 0_{n \times r} & \ldots & 0_n \end{pmatrix}.$$

For the VAR in (1), we have $Z_t := Y_t$, $C_t := (I_n, \ 0_{n \times np})$, $R_t := 0_n$ and

$$S_t := \begin{pmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \\ c \end{pmatrix}, \quad G_t := \begin{pmatrix} A_1 & A_2 & \ldots & A_p & I_n \\ I_n & 0_n & \ldots & 0_n & 0_n \\ \vdots & \ddots & \ldots & \vdots & \vdots \\ 0_n & \ldots & I_n & 0_n & 0_n \\ 0_n & \ldots & 0_n & 0_n & I_n \end{pmatrix}, \quad H_t := \begin{pmatrix} \Sigma & \ldots & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & \ldots & 0_n \end{pmatrix}.$$

For the implementation in differences the modifications are straightforward.

3.2 Conditional forecasts

Simulation smoothers (see Carter and Kohn, 1994; de Jong and Shephard, 1995; Durbin and Koopman, 2001a, for example) can be used to generate a draw of the state vector $S_t \, , t = 1, \ldots, T$ conditional on the observations $\{Z_t, t = 1, \ldots, T\}$ and (a draw of) the parameters, $C_t^{(j)}$, $G_t^{(j)}$, $R_t^{(j)}$, $H_t^{(j)}$:

$$S_{t|T}^{(j)} \sim p(S_t|Z_t, C_t^{(j)}, G_t^{(j)}, R_t^{(j)}, H_t^{(j)}, t = 1, \ldots, T), \quad t = 1, \ldots, T.$$

Let us now assume that for a subset of variables, $\mathcal{I}$, we are interested in obtaining conditional forecasts for $t > t_0$, conditional on their own past and on the past and future observations of the remaining variables, i.e. conditional on the information set $\Omega = \{Z_{l,t}, l \in \mathcal{I}, t \leq t_0, Z_{l,t}, l \notin \mathcal{I}, t = 1, \ldots, T\}$:

$$Z_{i,t|\Omega} \sim p(Z_{i,t}|\Omega), \quad i \in \mathcal{I}, t > t_0 .$$

In order to obtain such conditional forecasts, we adopt the solution proposed for forecasting with ragged edge data sets using a Kalman filter methodology, see e.g. Giannone, Reichlin, and Small (2005). In fact, the variables for which we do not assume the knowledge of a future path can be considered as time series with missing data. The Kalman filter allows to easily deal with such time series. Going more in details, we follow a standard approach (see e.g. Durbin and Koopman, 2001b, pp. 92-93) and apply the Kalman filter to a modified state space representation with Z_t, C_t and R_t replaced by $\tilde{Z}_t$, $\tilde{C}_t$ and $\tilde{R}_t$ respectively. The latter are derived from the former by removing the rows (and, for R_t, also columns) that correspond to the missing observations in Z_t.[10]

This insight is already sufficient in order to compute point conditional forecasts, by means of the Kalman smoother. In addition, assuming that the posterior distribution of the model parameters conditional on the data is available, the following algorithm (described for the generic iteration j) may be used in order to derive the distribution of the conditional forecasts:

Step 1: draw the parameters $C_t^{(j)}$, $G_t^{(j)}$, $R_t^{(j)}$ and $H_t^{(j)}$ from their posterior (either using explicit posterior distributions or some version of the Gibbs sampler).[11]

Step 2: draw the states $S_{t|\Omega}^{(j)}$ using a simulation smoother (Carter and Kohn, 1994; de Jong

[10]Giannone, Reichlin, and Small (2005) propose an equivalent solution. Instead of removing rows (and columns) of C_t and R_t that correspond to missing observations, they replace R_t with $\tilde{R}_t$ defined as follows:

$$(\tilde{R}_t)_{ij} = \begin{cases} (R_t)_{ij} & \text{if } Z_{i,t} \text{ and } Z_{j,t} \text{ are available} \\ \infty & \text{otherwise} \end{cases}$$

where, in practice, ∞ is a large number.

[11]In order to take the available future paths of selected variables into account when drawing the parameters, the initial values can be obtained from the "balanced" data set (up to t_0) and then steps 1-4 can be interated, with the latest conditional forecasts treated as data when drawing the parameters in Step 1.

and Shephard, 1995; Durbin and Koopman, 2001a) for the modified state space representation with the parameters $\tilde{C}_t^{(j)}$, $G_t^{(j)}$, $\tilde{R}_t^{(j)}$ and $H_t^{(j)}$.

Step 3: draw the noise for the observation equation from a conditional multivariate distribution $e_{i,t}^{(j)} \sim p(e_{i,t}^{(j)}|e_{l,t}^{(j)}, l \notin \mathcal{I}),\ \ i \in \mathcal{I}, t > t_0$.[12]

Step 4: compute $Z_{i,t|\Omega}^{(j)} = (C_t^{(j)})_{i\cdot}\, S_{t|\Omega}^{(j)} + e_{i,t}^{(j)},\ \ i \in \mathcal{I}, t > t_0$.

The algorithm can be modified in a straightforward manner for any pattern of "missing" observations in Z_t. Note that we can also easily condition on a linear combination of the observations. Suppose, in fact, that the aim is to condition on j linear combinations $W_t Z_t$ where W_t is a sequence of matrices of dimension $j \times k$. Then we set

$$\bar{Z}_t = \begin{pmatrix} W_t Z_t \\ Z_t \end{pmatrix}, \qquad \bar{C}_t = \begin{pmatrix} W_t C_t \\ C_t \end{pmatrix}, \qquad \bar{R}_t = \begin{pmatrix} W_t R_t W_t' & W_t R_t \\ R_t W_t' & W_t R_t W_t' \end{pmatrix}.$$

.

3.3 Comparison with the approach of Waggoner and Zha (1999)

The algorithm of Waggoner and Zha (1999) is a popular method to obtain conditional forecasts for VAR models. Roughly speaking, the methodology involves drawing directly vectors of ε_t, $t = t_0 + 1, \ldots, T$, in (1) which satisfy the required conditions. For the VAR described above and the pattern of variable availability discussed in section 3.2:

$$\Omega = \{Y_{l,t}, l \in \mathcal{I}, t \leq t_0, Y_{l,t}, l \notin \mathcal{I}, t = 1, \ldots, T\},$$

this would involve an inversion of a $q \times q$ matrix, where $q = (n - \#(\mathcal{I}))(T - t0)$ denotes the number of restrictions, and, more importantly, a QR decomposition of a $n(T-t0) \times n(T-t0)$ matrix (see e.g. Jarociński, 2010, for a detailed discussion).[13] When the number of variables and restrictions is large, this task could become computationally infeasible.[14]

In contrast, the application of the Kalman filter makes the problem "recursive", and the largest matrix to be inverted is of the size of the state vector, independently of the number of restrictions.

[12] See e.g. Greene (2002) pp. 872 for conditional normal distributions. For a VAR $e_{i,t}^{(j)} \equiv 0$.
[13] $\#(\mathcal{I})$ denotes the number of elements in $\mathcal{I}$.
[14] Jarociński (2010) proposes a way to decrease somewhat the computational complexity.

4 Empirical Results

4.1 Data set

Our data set includes 26 quarterly variables. Roughly, we include the most relevant real and nominal euro area aggregates and a set of international macroeconomic variables that proxies for global macroeconomic conditions (GDP and expenditure components, consumer and producer prices, labour market data, surveys, effective exchange rate, world economic activity, commodity prices), financial variables (short and long term interest rates, stock prices), credit (both to households and firms) and monetary aggregates (M1 and M3).

The sample covers the period from 1995Q1 to 2012Q4. Most of the data can be downloaded from the ECB Statistical Data Warehouse. The variables for which official data are not available over the entire sample, are back-dated by using the Area Wide Model data base (Fagan, Henry, and Mestre, 2005). The data appendix at the end of the paper provides the details, including the data transformations.

4.2 Model specifications

We include 3 common factors in the DFM, which are meant to roughly capture the information on real developments, prices and interest rates. Moreover, we include 4 lags in the VAR of the common factors in the DFM and in the BVAR in differences. In order to be consistent with the dynamics captured in the latter two approaches (which are specified on variables in differences), in the BVAR in levels we include 5 lags. In the BVAR approaches, we have to set-up the tightness of the prior distributions. As suggested in Giannone, Lenza, and Primiceri (2012), we follow a hierarchical approach and we treat the hyperparameters governing such tightness as random variables. Hence, we draw those hyperparameters from their posterior distribution.

4.3 Model validation: out-of-sample forecasting evaluation of unconditional forecasts

As a preliminary step, we gauge the accuracy of our empirical models in terms of out-of-sample unconditional forecasts. This preliminary step is particularly important because our models are specified on a large set of variables (26). This feature leads to a proliferation of parameters and, hence, potential instabilities due to estimation uncertainty might completely offset the gains obtained by limiting model misspecification. Out-of-sample forecasting, reflecting both parameter uncertainty and model misspecification, allows us to understand whether the benefits due to the generality of our models outweigh the costs.

For each of the three models, we produce the forecasts recursively for three horizons (1, 2 and 4 quarters). We focus on point forecasts. Our first estimation sample ranges from 1995Q1 to 2003Q4. Based on that sample, we set the hyperparameters of the BVAR models to the mode of their posterior distribution, we estimate the parameters of our three models and then we generate forecasts for 2004Q1 (one quarter ahead), 2004Q2 (two quarters ahead) and 2004Q4 (one year ahead). We then iterate the same procedure updating the estimation sample, one quarter at a time, until the end of the sample, i.e. 2012Q4. At each iteration, we also re-estimate the mode of the posterior distribution of the hyperparameters.

For each variable, the target of our evaluation is defined as $m^h_{i,t+h} = \frac{4}{h}[y_{i,t+h} - y_{i,t}]$. For variables specified in logs, this is approximately the average annualised growth rate over the next h quarters while, for variables not transformed in logs, this is the average annualised quarterly change over the next h quarters.

We compare our model with a simple benchmark model, a random walk with drift for the levels of the variables. This model is a particularly challenging benchmark over the monetary union sample and, in addition, it has also the appeal of being the prior model for the BVAR approaches. Hence, in case the BVAR models out-perform the benchmark model, this would suggest that they are able to extract valuable information from the sample.[15]

Table 1 reports the results of our analysis, for all models and variables. Results are cast in terms of ratios of the Mean Squared Forecast Errors (MSFE) of our three models with respect

[15]We also compared our models to a battery of univariate autoregressive models, another class of popular benchmark models, with very similar outcomes.

to the corresponding MSFE of the random walk. Hence, values smaller than one indicate that our model outperforms the random walk benchmark.

INSERT TABLE 1

The outcomes of the evaluation show that, in general, the BVARs and the DFM are able to outperform the benchmark, particularly for the short horizons. For the one- and two-quarter horizons, the models are more accurate than the random walk for most of the variables. Some failures to outperform the benchmark are not particularly surprising, since it is well known that it is very hard to beat the random walk model for financial and commodity prices, in general. Also for consumer prices (HICP) and the GDP deflator there is a relatively ample documentation of the difficulties to beat the random walk due to the steady anchoring of inflation expectations in the monetary union sample.[16]

As argued in section 2.3 and in De Mol, Giannone, and Reichlin (2008), there is a tight connection between the BVARs and the DFM approaches. Indeed, the out-of-sample performance of the three different approaches is quite similar. This reflects the fact that the forecasts from the three approaches are very correlated. Figure 1 (panels a-c) reports the bivariate correlation with the BVAR in levels of the DFM (black bar with stripes) and the BVAR in differences (red solid bar) for all the variables and forecast horizons.

INSERT FIGURE 1

The figure reveals the strong collinearity of the forecasts across approaches, providing empirical support for their theoretical connection highlighted in section 2.3 and in De Mol, Giannone, and Reichlin (2008).

4.4 Scenario analysis: an increase in world GDP

In this exercise we perform a scenario analysis to assess the effects associated with positive developments in the global economy, represented by a 0.1 percentage point stronger growth in global GDP.

[16]See, for example, Diron and Mojon (2005) and Giannone, Lenza, Momferatou, and Onorante (2013).

We compute the effects of the scenario by using our framework to produce conditional forecasts. Precisely, we estimate our models on the whole sample and generate two forecasts: an unconditional forecast given the sample $1, 2, \ldots, T$ (which provides a "baseline" scenario) and a conditional forecast in which the world GDP growth in $T + 1$ is set to the value of its own unconditional forecast plus 0.1 percentage points and all the remaining variables are left unconstrained (which we will refer to as "shocked scenario"). The scenario results for all variables are computed by taking the difference between the conditional and the unconditional forecasts described above. We explore the horizon of eight quarters.

Notice that the estimation of the scenario effects just described is akin to estimating a generalised impulse response to a change in world GDP (on this point, see Koop, Pesaran, and Potter, 1996; Pesaran and Shin, 1998).

Figure 2 shows the responses of some selected variables for the three models.[17] In particular, we report the distribution of the scenario effects computed in the context of the BVAR model (shades of orange) in levels and the point estimates of the effects in the other two modelling approaches (DFM: dashed blue line; BVAR in differences: black solid line). All results are reported in terms of deviations of (log-)levels of the variables in the shocked scenario compared to the baseline.[18]

INSERT FIGURE 2

The three approaches provide similar scenario assessments for all variables, at least qualitatively but, generally, also quantitatively. This result confirms the view that, for the variables commonly used in macroeconometric studies, dynamic factor models and Bayesian shrinkage (irrespective of data transformation) are both valid alternative methods to deal with the curse of dimensionality.

Going more in details of the scenario analysis, the top left panel reports the developments in global real GDP, which is 0.1 percent higher on impact (as assumed in the scenario assumption), keeps on increasing for the first year and then tends to drop back to the level prevailing before the initial increase.

[17] The total set of responses is available upon request.

[18] For the variables that are modelled in logs, this approximately corresponds to percentage deviation from the baseline for the levels.

The euro area real economy (GDP, exports, imports and unemployment) closely mirrors the developments in global GDP. Consumer prices are also higher, reaching a peak after about one year.

The short-term interest rate, which reflects systematic monetary policy, reacts positively to stabilize the economy and then drops back toward the initial level. Long-term interest rates are not particularly affected. This implies that the term-spread (defined as long-term interest rates minus short-term interest rates) decreases on impact, to finally revert to initial values.

Credit aggregates, which are traditionally very cyclical, follow the same path as GDP. Moreover, loans to households are coincident with GDP, while loans to firms lag behind. The narrow monetary aggregate M1 decreases on impact, reaching a trough after about one year. To understand this pattern, notice that M1 is negatively related to the short-term interest rate, indicating that its response to world demand is mainly driven by the liquidity effect. The effect on M3 is instead mostly driven by the increases in the short-term monetary assets included in the M3-M1 component, which completely offset the decrease in M1.(see Giannone, Lenza, and Reichlin, 2012, for an extensive discussion on the cyclical properties of credit and monetary aggregates in the euro area and their relationships with short-term and long-term interest rates)

An implicitly maintained assumption in this exercise is that the forecast paths we examine involve shocks small enough not to violate the Lucas critique.[19] Indeed, the reliability of the results rests on the fact that the perturbations we induce in the system by means of the scenario assumptions are not too big to induce a substantial shift in the behavior of economic agents which could, in turn, change the economic structure and, hence, the estimated reduced form parameters.

4.5 Conditional forecasts

In this exercise we generate forecasts from the three models conditional on the realised paths for the following three variables: real GDP, HICP and the short-term interest rate.

The conditional forecasts are generated over the period 1997-2012. The first two years in the

[19]See, Kilian and Lewis (2011) and references therein for a discussion of this issue.

sample are used as initial conditions. The parameters are estimated over the sample 1995-2007. Thus the conditional forecasts for 1997-2007 can be considered as "in-sample" while those over 2008-2012 as "out-of-sample". The aim of this exercise (see Giannone, Lenza, and Reichlin, 2012; Stock and Watson, 2012a, for similar exercises) is twofold. First, the "in-sample" part (1997-2007) of the conditional forecasts can be compared with the observed developments in order to gauge whether knowing only the time series of real GDP, HICP and the short-term interest rate is sufficient in order to capture the salient features of the variables in our model. Second, comparing the "out-of-sample" part (i.e. from 2008 onward) of the conditional forecasts with the observed developments, we can also assess whether the turmoil associated to the financial and the sovereign debt crises was reflected in a change in the structural economic relationships in the euro area. In fact, a change in the economic relationships would likely lead to relevant inaccuracies of the conditional forecasts based on parameters representing the pre-2007 economic relationships.

Figure 3 shows the conditional forecasts from the three models for the same selected variables shown in Figure 2.[20] As in the previous exercise, the distribution is generated using the BVAR in levels and blue dashed and black solid lines correspond, respectively, to the conditional point forecasts of the DFM and the BVAR in differences. In addition, the green line indicates the actual outcomes.

INSERT FIGURE 3

Analogously to the previous exercise, the forecasts from the three models are similar for most of the variables, indicating that different methodologies capture similar cross-sectional and dynamic information. In addition, the conditional forecasts are close to the actual outcomes, in particular in the "in-sample" period. This fact suggests that "3 dimensions" are sufficient to capture the developments in most of the economy[21] (Giannone, Reichlin, and Sala, 2004, reach a similar conclusion for the US economy).

[20] Three additional variables replace the three variables shown in figure 2 which in this exercise were used as conditions. A complete set of results is available upon request

[21] Notable exceptions are wages, GDP deflator, government consumption and the effective exchange rate (not shown). For these variables, the conditional forecast distributions cover a relatively wide range of values and the central forecasts are often quite far from the outcomes.

Turning to the "out-of-sample" evidence, there is still a general similarity of the conditional forecasts across approaches. However, some differences appear between forecasts and observed developments for a few variables, indicating an instability in the relationships of these variables with the conditioning set. For example, notable differences appear in the developments in money and credit variables, whose actual developments were much more subdued than what would have been predicted based on the conditioning information.[22] For the variables where we have evidence of instability, we also notice some discrepancies in the forecasts across methods.

5 Conclusions

We have modelled the dynamic interactions among a large set of macroeconomic and financial indicators in the euro area by means of large dynamic factor models and large Bayesian vector autoregressions.

We find that both classes of models are reliable tools for analyzing large data sets, since they produce accurate unconditional forecasts and meaningful scenarios.

Interestingly, the predictions of the two model classes are not only equally reliable, but are also very similar, in general. The fact that the results are not model specific is reassuring since it indicates that the predictions of the models are reflecting genuine data features.

The robustness and reliability of dynamic factor models and Bayesian vector autoregressions for analyzing large macroeconomic data sets has been already established for the United States in relation to forecasting and impulse response function analysis (see e.g. Bańbura, Giannone, and Reichlin, 2010; Giannone, Lenza, and Primiceri, 2012). We document that the same holds true for the euro area and for conditional forecasts.

In addition, we have shown how to implement scenario analysis and, in general, to compute conditional forecasts in the context of large data sets. The procedure is computationally feasible, produces meaningful results and interesting insights. The methodology has been already used in a number of papers including Giannone, Lenza, Momferatou, and Onorante (2013), Giannone, Lenza, and Reichlin (2010, 2012), Giannone, Lenza, Pill, and Reichlin

[22] Giannone, Lenza, and Reichlin (2012) extensively discuss and interpret the anomalies in the developments in credit and money markets during the crisis.

(2012), Lenza, Pill, and Reichlin (2010) and Luciani (2012).

A Implementation details

A.1 Dynamic factor model

If the factors were observed the joint likelihood of the data and the factors would be easily maximised and the estimates of the parameters would correspond to ordinary least squares outcomes. Specifically Λ and Γ_d would be obtained by regressing Δy_t on F_t while the autoregressive parameters $\Phi_1, \ldots, \Phi_s$ and the covariance matrix Q would be obtained by regressing F_t on its lags $F_{t-1}, \ldots, F_{t-s}$.

As the factors are unobserved the likelihood of the data cannot be maximised explicitly. As an alternative to numerical optimisation methods, the EM algorithm alternates between computing the expectation of the joint likelihood of the data and the factors given the parameter estimates from the previous step (E-step) and deriving new estimates by maximising the expected likelihood (M-step). An interesting property is that at each step the likelihood of the data increases, insuring that a convergence to a local maximum is reached.

Maximising the expected likelihood given the parameters at the j^{th} iteration is achieved through substituting the sufficient statistics with their expectation. This amounts to replacing the unobserved factors with their expected value $\hat{F}_t^{(j)} = \mathrm{E}_{\theta^{(j-1)}}\left[F_t | y_1, \ldots, y_T\right]$, and correcting for estimation uncertainty which is measured as

$$\hat{V}_{l,t}^{(j)} = \mathrm{E}_{\theta^{(j-1)}}\left[(F_t - \hat{F}_t^{(j)})(F_{t-l} - \hat{F}_{t-l}^{(j)})' | y_1, \ldots, y_T\right].$$

Those quantities can be computed recursively using the Kalman smoother.

It is easily seen that the expected sufficient statistics are as follows:

$$\mathrm{E}_{\theta^{(j-1)}}\left[\Delta y_t \Delta y_t' | y_1, \ldots, y_T\right] = \Delta y_t \Delta y_t',$$

$$\mathrm{E}_{\theta^{(j-1)}}\left[\Delta y_t F_t' | y_1, \ldots, y_T\right] = \Delta y_t \hat{F}_t^{(j)'}$$

and

$$\mathrm{E}_{\theta^{(j-1)}}\left[F_t F_{t-l}' | y_1, \ldots, y_T\right] = \hat{F}_t^{(j)} \hat{F}_{t-l}^{(j)'} + \hat{V}_{l,t}^{(j)}.$$

As a consequence the M-step consists of the following equations, where for simplicity we consider the case $s = 1$[23]:[24]

$$\hat{\Lambda}^{(j)} = \left(\sum_{t=1}^{T} \Delta y_t \hat{F}_t^{(j)\prime}\right)\left(\sum_{t=1}^{T} \hat{F}_t^{(j)}\hat{F}_t^{(j)\prime} + \hat{V}_{0,t}^{(j)}\right)^{-1},$$

$$\hat{\Gamma}_d^{(j)} = \frac{1}{T}\text{diag}\left[\left(\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right) - \hat{\Lambda}^{(j)}\left(\sum_{t=1}^{T} \hat{F}_t^{(j)} \Delta y_t'\right)\right],$$

$$\hat{\Phi}_1^{(j)} = \left(\sum_{t=2}^{T} F_t^{(j)}\hat{F}_{t-1}^{(j)\prime} + \hat{V}_{1,t}^{(j)}\right)\left(\sum_{t=2}^{T} \hat{F}_{t-1}^{(j)}\hat{F}_{t-1}^{(j)\prime} + \hat{V}_{0,t-1}^{(j)}\right)^{-1}$$

and

$$\hat{Q}^{(j)} = \frac{1}{T-1}\left(\left(\sum_{t=2}^{T} F_t^{(j)}F_t^{(j)\prime} + \hat{V}_{0,t}^{(j)}\right) - \hat{A}^{(j)}\left(\sum_{t=2}^{T} \hat{F}_{t-1}^{(j)}\hat{F}_t^{(j)\prime} + \hat{V}_{1,t}^{(j)}\right)\right).$$

Principal components represent a good starting point for the EM algorithm. The initial estimates of the factor loadings are obtained by regressing Δy_t on the principal components:

$$\hat{\Lambda}^{(0)} = \left(\sum_{t=1}^{T} \Delta y_t \hat{F}_t^{(0)\prime}\right)\left(\sum_{t=1}^{T} \hat{F}_t^{(0)}\hat{F}_t^{(0)\prime}\right)^{-1}.$$

The variance of the idiosyncratic residuals is hence given by:

$$\hat{\Gamma}_d^{(0)} = \frac{1}{T}\text{diag}\left[\left(\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right) - \hat{\Lambda}^{(0)}\left(\sum_{t=1}^{T} \hat{F}_t^{(0)} \Delta y_t'\right)\right].$$

The total variance of the residual is given by: $\text{trace}(\hat{\Gamma}_d^{(0)}) = d_{r+1} + \ldots + d_n$.

Turning to the estimation of the VAR model for the common factors, the OLS estimates, treating the factors as known, can be obtained as follows:

$$\hat{\Phi}_1^{(0)} = \left(\sum_{t=2}^{T} F_t^{(0)}\hat{F}_{t-1}^{(0)\prime}\right)\left(\sum_{t=2}^{T} \hat{F}_{t-1}^{(0)}\hat{F}_{t-1}^{(0)\prime}\right)^{-1}$$

and

$$\hat{Q}^{(0)} = \frac{1}{T-1}\left(\left(\sum_{t=2}^{T} F_t^{(0)}F_t^{(0)\prime}\right) - \hat{\Phi}_1^{(0)}\left(\sum_{t=2}^{T} \hat{F}_{t-1}^{(0)}\hat{F}_t^{(0)\prime}\right)\right).$$

[23] Extending it for more general situations is straightforward.

[24] Bańbura and Modugno (2014) show how to modify the following formulas in case some of the observations in y_t are missing.

It is important to stress that this algorithm is more efficient when the scale of all the variables is similar. Hence, although the QML estimates are scale invariant, it is useful to standardise variables beforehand. The scale can be re-attributed accordingly once the likelihood is maximised. Standardisation is also useful for assuring a good initialisation since principal components are not scale invariant.

A.2 Bayesian vector autoregression

In this section we summarise the the procedures derived by Giannone, Lenza, and Primiceri (2012).

Consider the VAR model of section 2.2:

$$\begin{aligned} y_t &= c + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t, \quad t = 1, ..., T\,, \\ \varepsilon_t &\sim N(0, \Sigma)\,, \end{aligned}$$

and rewrite it as

$$\begin{aligned} Y &= X\beta + \epsilon\,, \\ \epsilon &\sim N(0, \Sigma \otimes I_{T-p})\,, \end{aligned}$$

where $y \equiv (y_{p+1}, ..., y_T)'$, $Y \equiv vec(y)$, $x_t \equiv \left(1, y'_{t-1}, ..., y'_{t-p}\right)'$, $x \equiv (x_{p+1}, ..., x_T)'$, $X \equiv I_n \otimes x$, $\varepsilon \equiv (\varepsilon_{p+1}, ..., \varepsilon_T)'$, $\epsilon \equiv vec(\varepsilon)$, $B \equiv (c, A_1, ..., A_p)'$ and $\beta \equiv vec(B)$. Finally, denote the number of regressors for each equation by $k \equiv np + 1$.

For expositional convenience we will focus first on the implementation of the Minnesota prior. Later in the section, we will describe how to implement the sum-of-coefficient prior.

The Minnesota prior on (β, Σ) is given by the following normal-inverse-Wishart distribution:

$$\begin{aligned} \Sigma|\Psi &\sim IW(\Psi, d)\,, \\ \beta|\Sigma, \Psi, \lambda &\sim N(b, \Sigma \otimes \Omega_{\Psi,\lambda})\,. \end{aligned}$$

The posterior is given by:

$$\Sigma|\Psi, \lambda, Y \sim IW\left(\Psi + \hat{\varepsilon}'\hat{\varepsilon} + \left(\hat{B} - \hat{b}\right)' \Omega_{\Psi,\lambda}^{-1} \left(\hat{B} - \hat{b}\right), T - p + d\right), \tag{5}$$

$$\beta|\Sigma, \Psi, \lambda, Y \sim N\left(\hat{\beta}, \Sigma \otimes \left(x'x + \Omega_{\Psi,\lambda}^{-1}\right)^{-1}\right), \tag{6}$$

where $\hat{B} \equiv \left(x'x + \Omega_{\Psi,\lambda}^{-1}\right)^{-1}\left(x'y + \Omega_{\Psi,\lambda}^{-1}\hat{b}\right)$, $\hat{\beta} \equiv vec\left(\hat{B}\right)$, $\hat{\varepsilon} \equiv y - x\hat{B}$, $\hat{\epsilon} \equiv vec\left(\hat{\varepsilon}\right)$, and $\hat{b}$ is a $k \times n$ matrix obtained by reshaping the vector b in such a way that each column corresponds to the prior mean of the coefficients of each equation (i.e. $b \equiv vec(\hat{b})$).

We follow Giannone, Lenza, and Primiceri (2012) and set an almost flat, but proper, hyper-prior. For λ we choose Gamma distribution with mode equal to 0.2 and standard deviation equal to 0. Our prior on Ψ is an inverse-Gamma with scale and shape equal to $(0.02)^2$.

The posterior for the hyperparameters is $p(\Psi, \lambda|Y) \propto p\left(Y|\Psi, \lambda\right) p(\Psi, \lambda)$, where $p\left(Y|\Psi, \lambda\right)$ is the marginal likelihood, which takes the following form (see Giannone, Lenza, and Primiceri, 2012):

$$\begin{aligned} p\left(Y|\Psi, \lambda\right) &= \left(\frac{1}{\pi}\right)^{\frac{n(T-p)}{2}} \frac{\Gamma_n\left(\frac{T-p+d}{2}\right)}{\Gamma_n\left(\frac{d}{2}\right)} \times \\ &\quad \left|\Omega_{\Psi,\lambda}\right|^{-\frac{n}{2}} \cdot \left|\Psi\right|^{\frac{d}{2}} \cdot \left|x'x + \Omega_{\Psi,\lambda}^{-1}\right|^{-\frac{n}{2}} \times \\ &\quad \left|\Psi + \hat{\varepsilon}'\hat{\varepsilon} + \left(\hat{B} - \hat{b}\right)' \Omega_{\Psi,\lambda}^{-1}\left(\hat{B} - \hat{b}\right)\right|^{-\frac{T-p+d}{2}}. \end{aligned} \tag{7}$$

Draws from the joint posterior of the parameters and hyperparameters can be easily derived by using the following algorithm. Since the marginal likelihood conditional on the hyperparameters is available in closed form, the hyperparameters can be drawn using the Metropolis-Hastings algorithm. For any draw of the hyperparameters Ψ and λ, the covariance matrix of the residuals Σ and the autoregressive parameters β can be drawn from their distributions conditional on Ψ and λ.

A.2.1 Implementing the sum-of-coefficient prior

The sum of coefficient prior is implemented by using the following dummy observations:

$$Y_\mu = \text{diag}(\bar{y}_0)\mu; \quad X_\mu = [\mathbf{0}_n, \mathbf{1}_p' \otimes \text{diag}(\bar{y}_0)]\mu\,,$$

where $\bar{y}_0$ is the average of the first p initial observations, $\mathbf{1}_p$ is a $p \times 1$ vector of ones and $\mathbf{0}_n$ is a $n \times 1$ vector of zeros.

These dummy observations are added to the data and the procedure described above is performed on the augmented data set $Y_\mu^* = \left(Y'\ Y_\mu'\right)'$ and $X_\mu^* = \left(X'\ X_\mu'\right)'$. The only correction

that has to taken into account concerns the marginal likelihood which should be computed on the original data only. As derived in Giannone, Lenza, and Primiceri (2012), this is equivalent to taking the ratio between the marginal likelihood of the augmented data set relative to the marginal likelihood of the dummy observations:

$$p(Y|\Psi, \lambda, \mu) = p(Y_\mu^*|\Psi, \lambda)/p(Y_\mu|\Psi, \lambda).$$

B Description of the data set

No	Code	Variable	Transformation (BVAR in levels)
1	YER	Real GDP	Log-levels
2	PCR	Real private consumption	Log-levels
3	ITR	Real gross investment	Log-levels
4	GCR	Real government consumption	Log-levels
5	XXR	Real exports	Log-levels
6	MXR	Real imports	Log-levels
7	LNN	Total employment	Log-levels
8	URX	Unemployment rate	Raw
9	ESI	Economic sentiment indicator	Raw
10	YWR	World GDP	Log-levels
11	HICP	Harmonised consumer prices index	Log-levels
12	PPIXC	Producer price index	Log-levels
13	YED	GDP deflator	Log-levels
14	MXD	Imports deflator	Log-levels
15	WRN	Compensation per employee	Log-levels
16	EEN	Nominal effective exchange rate	Log-levels
17	POILU	Price of oil in USD	Log-levels
18	PCOMU	Non-energy commodity prices in USD	Log-levels
19	STN	Short term interest rate	Raw
20	LTN	Long term interest rate	Raw
21	SHEUR	Share prices - Total market euro area	Log-levels
22	FFR	Federal Funds Rate	Raw
23	M1	M1 notional stocks	Log-levels
24	M3	M3 notional stocks	Log-levels
25	LOAH	Loans to Households	Log-levels
26	LOAC	Loans to non-financial corporations	Log-levels

Note: In the BVAR in differences and in the DFM specification we take the first difference of the variables transformed as in the VAR in levels. The producer price index refers to total industry excluding construction. Loans to households are the sum of consumer loans, loans for house purchases and other loans.

TABLES AND FIGURES

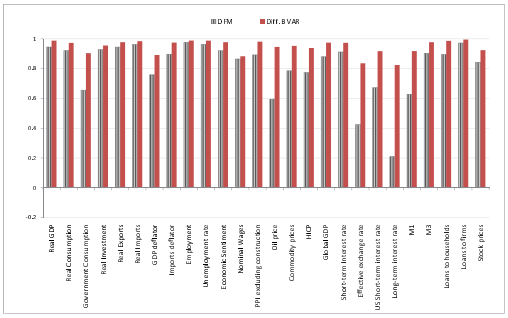
Table 1. Ratio of MSFE relative to random walk in levels

Variables	VAR in levels			VAR in differences			Dynamic Factor model		
	H=1	H=2	H=4	H=1	H=2	H=4	H=1	H=2	H=4
Real GDP	**0.60**	**0.84**	1.16	**0.56**	**0.80**	1.07	**0.56**	**0.70**	**0.83**
Real Consumption	**0.36**	**0.20**	**0.26**	**0.46**	**0.31**	**0.36**	**0.55**	**0.41**	**0.64**
Government Consumption	**0.56**	**0.43**	**0.35**	**0.59**	**0.44**	**0.38**	**0.85**	**0.85**	**0.91**
Real Investment	**0.54**	**0.63**	**0.68**	**0.53**	**0.56**	**0.61**	**0.43**	**0.46**	**0.51**
Real Exports	**0.71**	1.03	1.45	**0.71**	**0.98**	1.43	**0.72**	**0.99**	1.02
Real Imports	**0.51**	**0.64**	**0.77**	**0.51**	**0.64**	**0.84**	**0.54**	**0.65**	**0.73**
GDP deflator	**0.76**	**0.95**	1.63	1.20	1.35	1.58	1.25	1.38	1.81
Imports deflator	**0.79**	1.21	1.86	**0.78**	1.31	2.27	**0.91**	1.53	2.08
Employment	**0.17**	**0.22**	**0.38**	**0.16**	**0.22**	**0.39**	**0.17**	**0.19**	**0.33**
Unemployment rate	**0.41**	**0.58**	**0.89**	**0.37**	**0.52**	**0.82**	**0.30**	**0.39**	**0.61**
Economic Sentiment	**0.54**	**0.75**	**0.85**	**0.60**	**0.93**	1.27	**0.73**	**0.97**	1.17
Nominal Wages	1.06	1.36	1.98	1.12	1.04	1.28	**0.87**	**0.96**	1.44
PPI excluding construction	**0.75**	1.14	1.71	**0.70**	1.14	1.85	**0.97**	1.58	2.26
Oil price	1.30	1.62	1.65	1.37	1.83	2.05	1.93	2.89	3.43
Commodity prices	1.16	1.43	1.51	1.20	1.59	1.74	1.23	1.55	1.63
HICP	**0.99**	1.39	2.34	1.00	1.35	1.98	1.55	2.27	3.25
Global GDP	**0.66**	**0.72**	**0.84**	**0.66**	**0.72**	**0.73**	**0.60**	**0.62**	**0.63**
Short-term Interest rate	**0.72**	1.08	1.81	**0.61**	**0.98**	2.00	**0.71**	1.18	1.74
Effective exchange rate	1.34	1.50	2.20	1.54	1.60	2.94	1.29	1.43	1.76
US Short-term interest rate	1.02	1.27	1.31	**0.78**	**0.93**	**0.96**	**0.87**	**0.88**	**0.81**
Long-term interest rate	1.02	1.19	1.83	1.01	1.29	2.20	1.08	1.25	1.62
M1	**0.60**	**0.73**	1.37	**0.69**	**0.67**	1.27	**0.89**	1.22	1.74
M3	**0.44**	**0.43**	**0.56**	**0.54**	**0.55**	**0.71**	**0.54**	**0.49**	**0.55**
Loans to households	**0.09**	**0.11**	**0.17**	**0.12**	**0.18**	**0.31**	**0.25**	**0.31**	**0.50**
Loans to firms	**0.06**	**0.08**	**0.22**	**0.07**	**0.11**	**0.28**	**0.15**	**0.20**	**0.40**
Stock prices	**0.93**	1.10	1.53	**0.97**	1.13	1.89	**0.90**	1.12	1.50

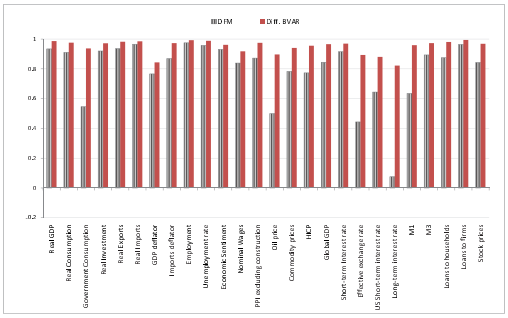
Note: The table reports the ratio of Mean Squared Forecast Errors (MSFE) of the BVAR in levels, BVAR in differences and the DFM over the MSFE of the random walk in levels with drift (the model that would prevail if we assumed a dogmatic prior). The ratios are reported for the horizons of one, two and four quarters ahead. Values smaller than one (in bold) indicate that the MSFE of a specific model is lower than the corresponding MSFE of the random walk model.

Figure 1:
Correlation of DFM and BVAR in differences forecasts with BVAR in levels forecasts

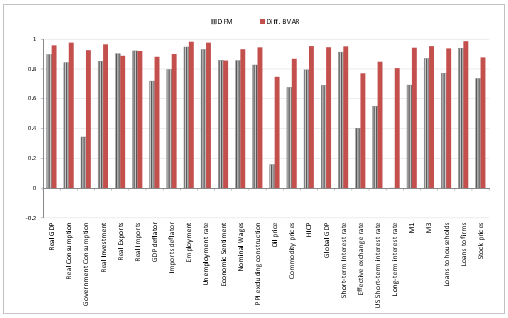
(a) One quarter ahead



(b) Two quarters ahead

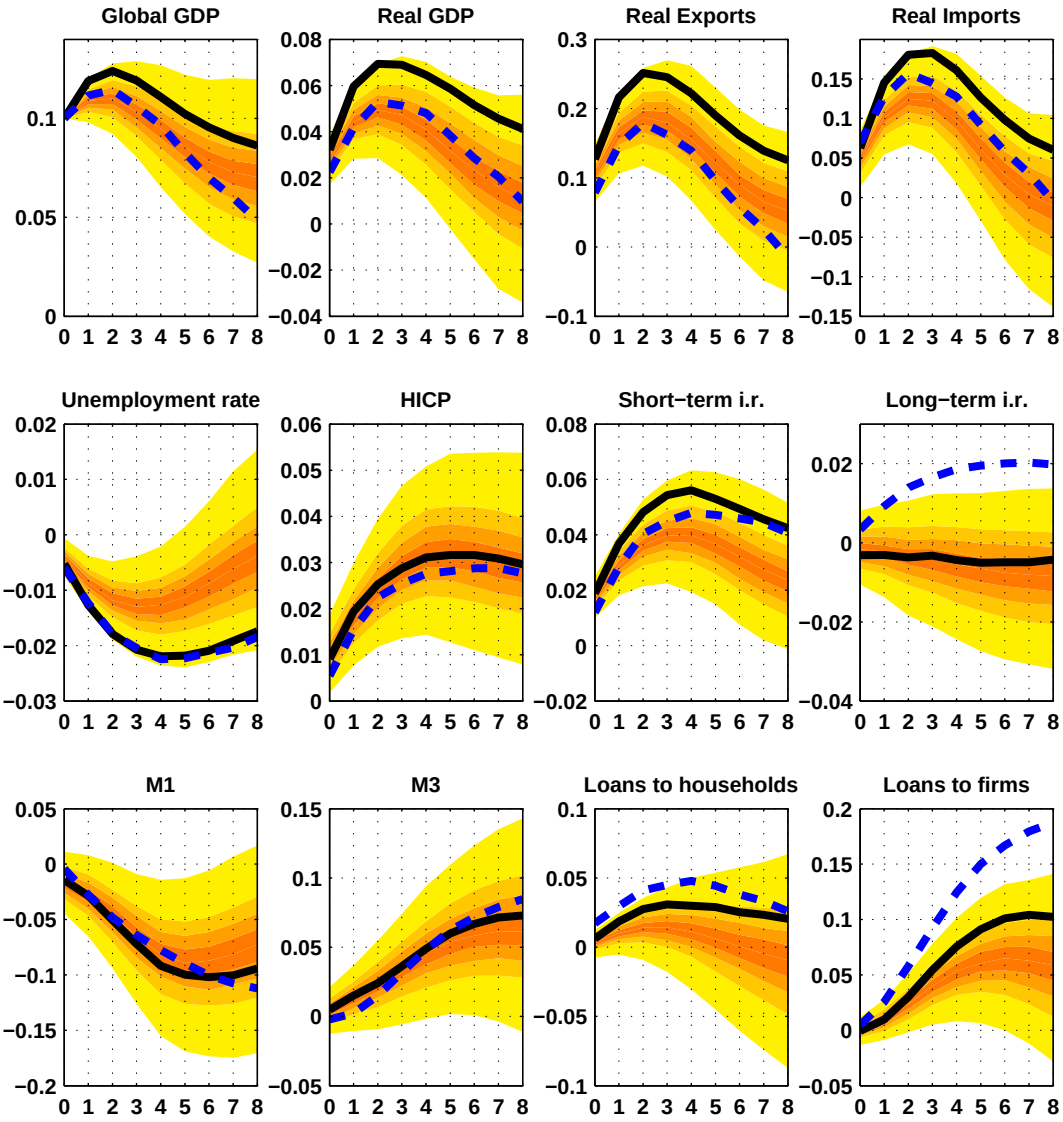


(c) Four quarters ahead



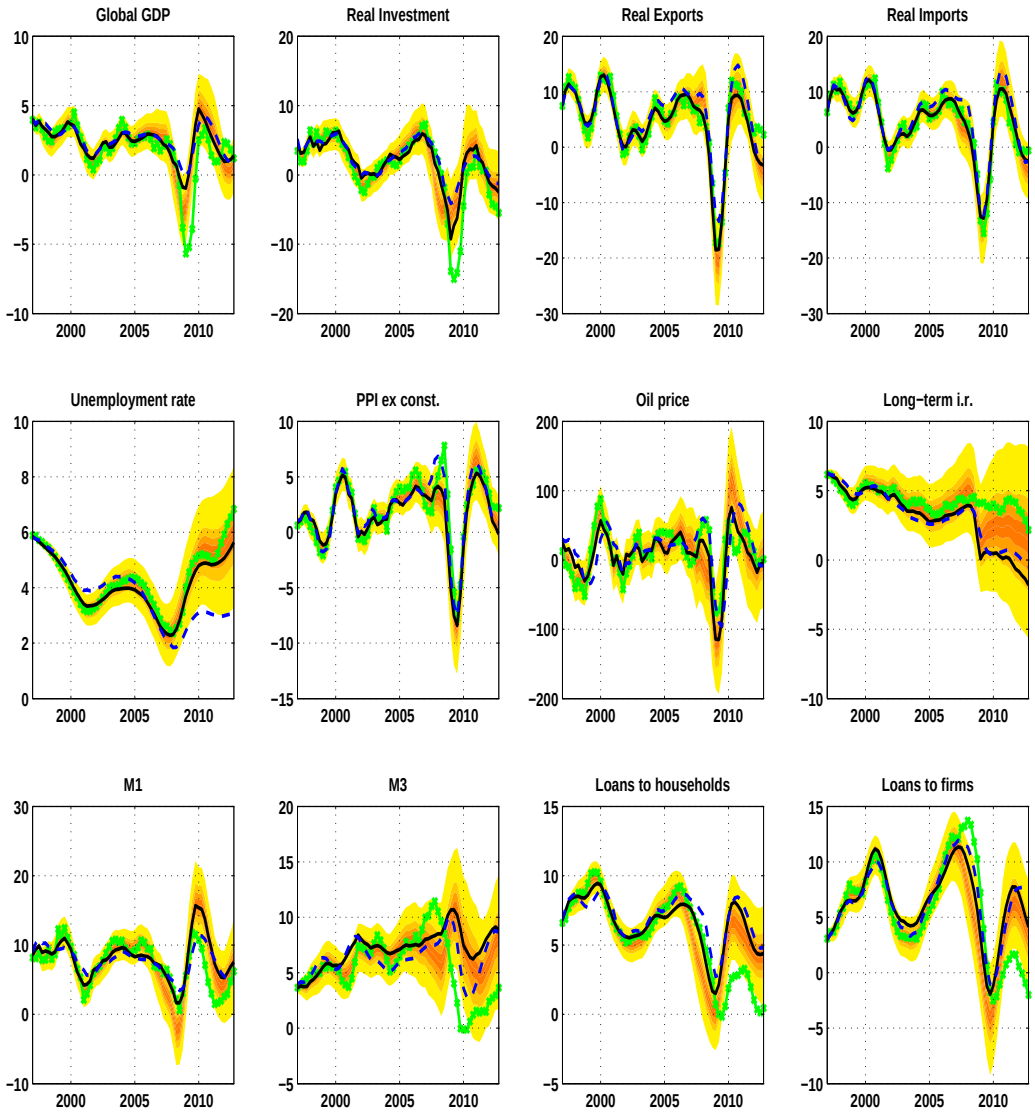
Note: For each variable on the horizontal axis, we report the correlation between the forecasts from the DFM and BVAR in levels (bars with white stripes) and between forecasts from the BVAR in differences and BVAR in levels (bars with red solid fill).

Figure 2: Scenario analysis: an increase in world GDP



Note: The scenario responses are shown in terms of percentage deviation of variables levels in the shocked scenario relative to the baseline scenario (except for unemployment and the interest rates, for which we show deviations). Shades of orange: distribution of the scenario responses in the BVAR in levels, excluding the lower and higher 5% quantiles. Dashed blue line: point estimate of the scenario responses in the DFM model. Solid black line: point estimate of the scenario responses in the BVAR in differences. The latter is computed as the mean of the distribution of the scenario responses in the BVAR in differences.

Figure 3: Conditional Forecasts



Note: Shades of orange: distribution of the conditional forecasts in the BVAR in levels, excluding the lower and higher 5% quantiles. Dashed blue line: point estimate of the conditional forecasts in the DFM model. Solid black line: point estimate of the conditional forecasts in the BVAR in differences. The latter is computed as the mean of the distribution of the scenario responses in the BVAR in differences. Green line with crosses: actual values. The variables are all reported in terms of annual percentage changes, except the unemployment rate and the long-term interest rate which are in levels. Compared to Figure 1, we report total investment, producer price index and the oil price in place of real GDP, HICP and the short-term interest rate, which are our conditioning assumptions.

References

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135–171.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

Bańbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92.

Bańbura, M., and M. Modugno (2014): "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, 29(1), 133160.

Baumeister, C., and L. Kilian (2011): "Real-Time Analysis of Oil Price Risks Using Forecast Scenarios," CEPR Discussion Papers 8698, C.E.P.R. Discussion Papers.

Bloor, C., and T. Matheson (2011): "Real-time conditional forecasts with Bayesian VARs: An application to New Zealand," *The North American Journal of Economics and Finance*, 22(1), 26–42.

Carriero, A., T. E. Clark, and M. Marcellino (2012): "Common drifting volatility in large Bayesian VARs," Discussion paper.

Carter, C., and P. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrica*, 81, 541–553.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 2, pp. 65–148. Elsevier.

Clarida, R. H., and D. Coyle (1984): "Conditional Projection by Means of Kalman Filtering," NBER Technical Working Papers 0036, National Bureau of Economic Research, Inc.

de Jong, P., and N. Shephard (1995): "The simulation smoother for time series models," *Biometrika*, 2, 339–350.

De Mol, C., D. Giannone, and L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?," *Journal of Econometrics*, 146(2), 318–328.

Dempster, A., N. L., and D. Rubin (1977): "Maximum Likelihood Estimation From Incomplete Data," *Journal of the Royal Statistical Society*, 14, 1–38.

Diron, M., and B. Mojon (2005): "Forecasting the central banks inflation objective is a good rule of thumb," Working Paper Series 0564, European Central Bank.

Doan, T., R. Litterman, and C. A. Sims (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

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

——— (2012): "A QuasiMaximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94(4), 1014–1024.

Durbin, J., and S. Koopman (2001a): "An efficient and simple simulation smoother for state space time series analysis," Computing in Economics and Finance 2001 52, Society for Computational Economics.

Durbin, J., and S. J. Koopman (2001b): *Time Series Analysis by State Space Methods*. Oxford University Press.

Fagan, G., J. Henry, and R. Mestre (2005): "An area-wide model for the euro area," *Economic Modelling*, 22(1), 39–59.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119(1), 1 – 38.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalized Dynamic Factor Model: identification and estimation," *Review of Economics and Statistics*, 82, 540–554.

——— (2004): "The generalized dynamic factor model: consistency and rates," *Journal of Econometrics*, 119, 231–245.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2013): "Short-term inflation projections: a bayesian vector autoregressive approach," *International Journal of Forecasting*, forthcoming.

Giannone, D., M. Lenza, H. Pill, and L. Reichlin (2012): "The ECB and the Interbank Market," *Economic Journal*, 122(564), F467–F486.

Giannone, D., M. Lenza, and G. E. Primiceri (2012): "Prior Selection for Vector Autoregressions," NBER Working Papers 18467, National Bureau of Economic Research, Inc.

Giannone, D., M. Lenza, and L. Reichlin (2010): "Business Cycles in the Euro Area," in *Europe and the Euro*, NBER Chapters, pp. 141–167. National Bureau of Economic Research, Inc.

——— (2012): "Money, credit, monetary policy and the business cycle in the euro area," CEPR Discussion Papers 8944, C.E.P.R. Discussion Papers.

Giannone, D., L. Reichlin, and L. Sala (2004): "Monetary Policy in Real Time," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, pp. 161–200. MIT Press.

Giannone, D., L. Reichlin, and D. Small (2005): "Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases," Finance and Economics Discussion Series 2005-42, Board of Governors of the Federal Reserve System (U.S.).

Greene, W. H. (2002): *Econometric Analysis*. Prentice Hall.

Harvey, A. (1989): *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Jarociński, M. (2010): "Conditional forecasts and uncertainty about forecast revisions in vector autoregressions," *Economics Letters*, 108(3), 257–259.

Jarociński, M., and F. R. Smets (2008): "House prices and the stance of monetary policy," *Review*, (Jul), 339–366.

Jungbacker, B., and S. J. Koopman (2008): "Likelihood-based Analysis for Dynamic Factor Models," Tinbergen Institute Discussion Papers 08-007/4, Tinbergen Institute.

Karlsson, S. (2013): "Forecasting with Bayesian Vector Autoregressions," in *Handbook of Economic Forecasting, Volume 2B*, ed. by G. Elliott, and T. Timmermann. North Holland, Elsevier.

Kilian, L., and L. T. Lewis (2011): "Does the Fed Respond to Oil Price Shocks?," *Economic Journal*, 121(555), 1047–1072.

Koop, G., M. H. Pesaran, and S. M. Potter (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74(1), 119–147.

Koop, G. M. (2013): "Forecasting with Medium and Large Bayesian VARS," *Journal of Applied Econometrics*, 28(2), 177–203.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003): "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, 93, 1216–1239.

Lenza, M., H. Pill, and L. Reichlin (2010): "Monetary policy in exceptional times," *Economic Policy*, 25, 295–339.

Litterman, R. B. (1979): "Techniques of forecasting using vector autoregressions," Discussion paper.

Luciani, M. (2012): "Monetary Policy and the Housing Market: A Structural Factor Analysis," Working Papers ECARES ECARES 2012-035, ULB – Universite Libre de Bruxelles, Journal of Applied Econometrics, forthcoming.

Ng, S. (2013): "Variable Selection in Predictive Regressions," in *Handbook of Economic Forecasting, Volume 2B*, ed. by G. Elliott, and T. Timmermann. North Holland, Elsevier.

Paciello, L. (2011): "Does Inflation Adjust Faster to Aggregate Technology Shocks than to Monetary Policy Shocks?," *Journal of Money, Credit and Banking*, 43(8), 1663–1684.

Pesaran, H., and Y. Shin (1998): "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, 58(1), 17 – 29.

Shumway, R., and D. Stoffer (1982): "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), pp. 1–48.

Stock, J. H., and M. W. Watson (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 147–162.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes.," *Journal of Business and Economics Statistics*, 20, 147–162.

——— (2011): "Dynamic Factor Models," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry. Oxford University Press.

——— (2012a): "Disentangling the Channels of the 2007-2009 Recession," NBER Working Papers 18094, National Bureau of Economic Research, Inc.

——— (2012b): "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business & Economic Statistics*, 30(4), 481–493.

Waggoner, D. F., and T. Zha (1999): "Conditional Forecasts In Dynamic Multivariate Models," *The Review of Economics and Statistics*, 81(4), 639–651.

Watson, M. W., and R. F. Engle (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, 23, 385–400.