Bayesian VARs with Large Panels*

Marta Banbura, Université Libre de Bruxelles, ECARES
Domenico Giannone, European Central Bank, ECARES and CEPR
Lucrezia Reichlin, European Central Bank, ECARES and CEPR

May 14, 2007

Abstract

This paper assesses the performance of Bayesian Vector Autoregression (BVAR) for models of different size. We consider standard specifications in the macroeconomic literature based on, respectively, three and eight variables and compare results with those obtained by larger models containing twenty or over one hundred conjunctural indicators. We first study forecasting accuracy and then perform a structural exercise focused on the effect of a monetary policy shock on the macroeconomy. Results show that BVARs estimated on the basis of hundred variables perform well in forecasting and are suitable for structural analysis.

JEL Classification: C11,C13,C33,C53

Keywords: Bayesian VAR, Forecasting, Monetary VAR, large cross-sections

* We would like to thank seminar participants at the New York Fed and at the second forecasting conference on empirical econometric methods at Duke University. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank. Support by the grants "Action de Recherche Concertée" Nb 02/07-281 and IAP-network in Statistics P5/24 is gratefully acknowledged. Please address any comments to Marta Banbura mbanbura@ulb.ac.be; Domenico Giannone dgiannon@ulb.ac.be; or Lucrezia Reichlin lucrezia.reichlin@ecb.int.

Replication files are available at:

http://homepages.ulb.ac.be/∼dgiannon/ or http://homepages.ulb.ac.be/∼lreichli/

1 Introduction

Vector Auto Regressive (VAR) models are standard tools in macroeconomics and are widely used for structural analysis and forecasting. In the early literature, Litterman (1986) and Doan, Litterman, and Sims (1984) showed how to specify priors on the coefficients and obtain successful forecasting results. More recently, priors have been designed to correspond to some features of macro models as, for example, in Del Negro and Schorfheide (2004). For structural analysis, perhaps the most successful application has been the analysis of the effect of monetary shocks to the economy (e.g. Leeper, Sims, and Zha, 1996; Sims and Zha, 1998). These applications are typically based on systems of small dimensions, matching the dimension of the typical structural macroeconomic model. The most common monetary VAR ranges from three variables (a measure of real activity, a price variable and the policy instrument), to about ten variables in the richest specification (as, for example, in Christiano, Eichenbaum, and Evans, 1999). The largest VAR in the literature contains about twenty variables (Leeper, Sims, and Zha, 1996).

This paper asks the question of whether Bayesian VARs can be estimated on a large number of variables, say 100 or more and, if yes, whether the forecasts are accurate and the impulse response functions of identified shocks interpretable and reasonable.

This problem is interesting because both the forecast and the structural analysis may be affected by informational assumptions (on this point, see Forni, Giannone, Lippi, and Reichlin, 2005; Giannone and Reichlin, 2006). For example, when identifying the monetary shock, it is important to condition on the relevant information set of the central bank, possibly containing many conjunctural indicators and financial variables. The empirical relevance of taking into account such information has been shown by Bernanke, Boivin, and Eliasz (2005), Favero, Marcellino, and Neglia (2005), Giannone, Reichlin, and Sala (2004) and Stock and Watson (2005b) in frameworks related to factor analysis. The literature based on factor models has also shown that large information helps in forecasting (Bernanke and Boivin, 2003; Boivin and Ng, 2005; D'Agostino and Giannone, 2006; Forni, Hallin, Lippi, and Reichlin, 2005, 2003; Giannone, Reichlin, and Sala, 2004; Marcellino, Stock, and Watson, 2003; Stock and Watson, 2002a,b).

Bayesian VARs, on the other hand, have not been studied for large data sets and this is puzzling. After all, Bayesian shrinkage is designed to solve the curse of dimensionality problem.

Our paper is a follow up to De Mol, Giannone, and Reichlin (2006) where we studied the Bayesian regression both empirically and asymptotically (as the dimension of the cross section n and that of the sample size T go to infinity). In that paper we showed that, under suitable conditions, a forecast based on the point estimates of Bayesian regression converges to the optimal forecast for n and T going to infinity along any path. Those results, moreover, suggested to increase the degree of shrinkage as n becomes larger.

In this paper we go beyond simple regression and study the VAR case. We use standard Litterman priors and perform both a forecasting and a structural exercise focusing on the effect of the monetary policy shock on the macroeconomy. We compare results for small systems typically used in the literature with those obtained from larger systems, which include the hundred and thirty one variables used by Stock and Watson (2005a) for forecasting based on principal components regression. We also compare our forecasting results with those obtained from VARs augmented by factors as in Bernanke, Boivin, and Eliasz (2005) (FAVAR model). Our results show that the Bayesian VAR is an appropriate tool for forecasting and structural analysis when it is desirable to condition on a large information set. Bayesian shrinkage is a natural way to deal with the curse of dimensionality. Given the progress in computing power (see Hamilton, 2006, for a discussion), estimation does not present any computational problem, requiring the inversion of a matrix which for our large data-set of a hundred and thirty variables is about 2000 by 2000 when including 13 lags in the VAR.

The paper is organized as follows. In Section 2 we describe the priors for the baseline BVAR model and the data. In Section 3 we perform the forecasting evaluation for all models and in Section 4 the structural analysis on the effect of the monetary policy shocks. Section 5 concludes while the Appendix shows results for a number of alternative specifications to verify the robustness of our findings.

2 Setting the priors for the VAR

Let $Y_t = (\, y_{1,t} \; y_{2,t} \; \ldots \; y_{n,t} \,)'$ be a potentially large set of time series. We consider the following VAR(p) model:

$$Y_t = c + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t \,, \tag{1}$$

where u_t is an n-dimensional Gaussian white noise with covariance matrix $\mathbb{E} u_t u_t' = \Psi$, $c = (c_1, \ldots, c_n)'$ is an n-dimensional vector of constants and $A_1, \ldots, A_p$ are $n \times n$ autoregressive matrices.

We estimate the model using the Bayesian VAR (BVAR) approach which helps to overcome the curse of dimensionality via the imposition of prior beliefs on the parameters. In setting the prior distributions, we follow standard practice and use the procedure developed in Litterman (1986) with modifications proposed by Kadiyala and Karlsson (1997) and Sims and Zha (1998).

Litterman (1986) suggests using a prior often referred to as the Minnesota prior. The basic principle behind it is that all the equations are "centered" around the random walk with drift, i.e. the prior mean can be associated with the following representation for Y_t:

$$Y_t = c + Y_{t-1} + u_t \,.$$

This amounts to shrinking the diagonal elements of A_1 toward one and the remaining coefficients in $A_1, \ldots, A_p$ toward zero. In addition, the prior specification incorporates the belief that the more recent lags should provide more reliable information than the more distant ones and that own lags should explain more of the variation of a given variable than the lags of other variables in the equation.

These prior beliefs are imposed by setting the following moments for the prior distribution of the coefficients:

$$\mathbb{E}[(A_k)_{ij}] = \begin{cases} \delta_i \,, & j = i, k = 1 \\ 0, & \text{otherwise} \end{cases} \quad , \qquad \mathbb{V}[(A_k)_{ij}] = \vartheta \frac{\lambda^2}{k^2} \frac{\sigma_i^2}{\sigma_j^2} . \tag{2}$$

The coefficients $A_1, \ldots, A_p$ are assumed to be a priori independent and normally distributed. As for the covariance matrix of the residuals, it is assumed to be diagonal, fixed and known: $\Psi = \Sigma$ where $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Finally, the prior on the intercept is diffuse.

Originally, Litterman sets $\delta_i = 1$ for all i, reflecting the belief that all the variables are characterized by high persistence. However, this prior is not appropriate for variables believed to be characterized by substantial mean reversion. For those we impose the prior belief of white noise by setting $\delta_i = 0$.

The hyperparameter λ controls the overall tightness of the prior distribution around the random walk or white noise and governs the relative importance of the prior beliefs with respect to the information contained in the data. For $\lambda = 0$ the posterior equals the prior and the data do not influence the estimates. If $\lambda = \infty$, on the other hand, posterior expectations coincide with the Ordinary Least Squares estimates. We argue that the overall tightness governed by λ should be chosen in relation to the size of the system. As the number of variables increases the parameters should be shrunk more in order to avoid overfitting. De Mol, Giannone, and Reichlin (2006) discuss this point in detail and show that, as intuition suggests, as the number of estimated parameters increases the overall tightness should increase as well.

The factor $1/k^2$ is the rate at which prior variance decreases with increasing lag length and σ_i^2/σ_j^2 accounts for the different scale and variability of the data. The coefficient $\vartheta \in (0, 1)$ governs the extent to which the lags of other variables are "less important" than the own lags.

In the context of the structural analysis we need to take into account possible correlation among the residuals of different variables. Consequently, Litterman's assumption of fixed and diagonal covariance matrix is somewhat problematic. To overcome this problem we follow Kadiyala and Karlsson (1997) and Robertson and Tallman (1999) and impose a Normal inverted Wishart prior which retains the principles of the Minnesota prior. This is possible under the condition that $\vartheta = 1$, which will be assumed in what follows. Let us write the VAR in (1) as a system of multivariate regressions (see e.g. Kadiyala and Karlsson, 1997):

$$\underset{T\times n}{Y} = \underset{T\times k}{X} \underset{k\times n}{B} + \underset{T\times n}{U}, \tag{3}$$

where $Y = (y_1, \ldots, y_T)'$, $X = (X_1, \ldots, X_T)'$ with $X_t = (Y'_{t-1}, \ldots, Y'_{t-p}, 1)$, $U = (u_1, \ldots, u_T)'$, and $B = (A_1, \ldots, A_p, c)'$ is the $k \times n$ matrix containing all the coefficients and $k = np + 1$. The Normal inverted Wishart prior has the form:

$$\Psi \sim iW(S_0, \alpha_0) \qquad \text{and} \qquad B|\Psi \sim N(B_0, \Psi \otimes \Omega_0), \tag{4}$$

where the prior parameters B_0, Ω_0, S_0 and α_0 are chosen so that prior expectations and variances of B coincide with those implied by equation (2) and the expectation of Ψ is equal to the fixed residual covariance matrix Σ of the Minnesota prior, for details see Kadiyala and Karlsson (1997).

We implement the prior (4) by adding dummy observations. It can be shown that adding T_d dummy observations Y_d and X_d to the system (3) is equivalent to imposing the Normal inverted Wishart prior with $B_0 = (X_d' X_d)^{-1} X_d' Y_d$, $\Omega_0 = (X_d' X_d)^{-1}$, $S_0 = (Y_d - X_d B_0)'(Y_d - X_d B_0)$ and $\alpha_0 = T_d - k - n - 1$. In order to match the Minnesota moments, we add the following dummy observations:

$$Y_d = \begin{pmatrix} \mathrm{diag}(\delta_1 \sigma_1, \ldots, \delta_n \sigma_n)/\lambda \\ 0_{n(p-1)\times n} \\ \cdots\cdots\cdots\cdots \\ \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \\ \cdots\cdots\cdots\cdots \\ 0_{1\times n} \end{pmatrix} \qquad X_d = \begin{pmatrix} J_p \otimes \mathrm{diag}(\sigma_1, \ldots, \sigma_n)/\lambda & 0_{np\times 1} \\ \cdots\cdots\cdots\cdots & \cdots \\ 0_{n\times np} & 0_{n\times 1} \\ \cdots\cdots\cdots\cdots & \cdots \\ 0_{1\times np} & \epsilon \end{pmatrix} \tag{5}$$

where $J_p = \mathrm{diag}(1, 2, \ldots, p)$. Roughly speaking, the first block of dummies imposes prior beliefs on the autoregressive coefficients, the second block implements the prior for the covariance matrix and the third block reflects the uninformative prior for the intercept (ϵ is a very small number). Although the parameters should in principle be set using only prior knowledge we follow common practice (see e.g. Litterman, 1986; Sims and Zha, 1998) and set the scale parameters σ_i^2 equal the variance of a residual from a univariate autoregressive model of order p for the variables y_{it}.

Consider now the regression model (3) augmented with the dummies in (5):

$$\underset{T_*\times n}{Y_*} = \underset{T_*\times k}{X_*} \; \underset{k\times n}{B} + \underset{T_*\times n}{U_*}, \tag{6}$$

where $T_* = T + T_d$, $Y_* = (Y', Y_d')'$, $X_* = (X', X_d')$ and $U_* = (U', U_d')'$. To insure the existence of the prior expectation of Ψ it is necessary to add an improper prior $\Psi \sim |\Psi|^{-(n+3)/2}$. In that case the posterior has the form:

$$\Psi|Y \sim iW\left(\tilde{\Sigma}, T_d + 2 + T - k\right) \qquad \text{and} \qquad B|\Psi, Y \sim N\left(\tilde{B}, \Psi \otimes (X_*' X_*)^{-1}\right), \tag{7}$$

with $\tilde{B} = (X_*' X_*)^{-1} X_*' Y_*$ and $\tilde{\Sigma} = (Y_* - X_* \tilde{B})'(Y_* - X_* \tilde{B})$. Note that the posterior expectation of the coefficients coincides with the OLS estimates of the regression of Y_* on X_*. It can be easily checked that it also coincides with the posterior mean for the Minnesota setup in (2). From the computational point of view, estimation is feasible since it only requires the inversion of a square matrix of dimension $k = np + 1$. For the large data-set of hundred and thirty variables and thirteen lags k is smaller than 2000.

The dummy variable implementation will prove useful for imposing additional beliefs. We will exploit this feature in Section 4.1.

2.1 Data

We use the data set of Stock and Watson (2005a). This data set contains 131 monthly macro indicators covering broad range of categories including, inter alia, income, industrial production, capacity, employment and unemployment, consumer prices, producer prices, wages, housing starts, inventories and orders, stock prices, interest rates for different maturities, exchange rates, money aggregates. The time span is from January 1959 through December 2003. We apply logarithms to most of the series with the exception of those already expressed in rates. For most of the variables we use the random walk prior, that is we set $\delta_i = 1$. The white noise prior, $\delta_i = 0$, is used for variables characterized by mean reversion.[1] The description of the data set, including the information on the transformations and the specification of δ_i for each series, is provided in the Appendix.

We analyze VARs of different sizes. We first look at the forecast performance. Then we identify the monetary policy shock and study impulse response functions as well as variance decompositions. The variables of special interest include a measure of real economic activity, a measure of prices and a monetary instrument. As in Christiano, Eichenbaum, and Evans (1999), we use employment as an indicator of real economic activity measured by the number of employees on non-farm payrolls (CES002). The level of prices is measured by the consumer price index (PUNEW) and the monetary instrument is the Federal Funds Rate (FYFF).

We consider the following VAR specifications:

- *SMALL*. This is a small monetary VAR including the three key variables;

[1] Since Stock and Watson (2005a) transform the data to stationarity, we set the random walk prior whenever they take differences of the data. We set a white noise prior otherwise.

- *CEE*. This is the monetary model of Christiano, Eichenbaum, and Evans (1999). In addition to the key variables in *SMALL*, this model includes the index of sensitive material prices (PSM99Q) and monetary aggregates: non-borrowed reserves (FMRRA), total reserves (FMRNBA) and $M2$ money stock (FM2);

- *MEDIUM*. This VAR extends the CEE model by the following variables: Personal Income (A0M051), Real Consumption (A0M051R), Industrial Production (IPS10), Capacity Utilization (A0M082), Unemployment Rate (LHUR), Housing Starts (HSFR), Producer Price Index (PWFSA), Personal Consumption Expenditures Price Deflator (GDMC), Average Hourly Earnings (CES275), M1 Monetary Stock (FM1), Standard and Poor's Stock Price Index (FSPCOM) and Yields on 10 year U.S. Treasury Bond (FYGT10). The system contains in total 20 variables.

- *LARGE*. This specification includes all the 131 macroeconomic indicators of Stock and Watson's data set.

It is important to stress that since we compare models of different size, we need to have a strategy for settting the overall tightness as models become larger. The problem is to control for overfitting while keeping the models comparable. We achieve this by setting priors that ensure that the in-sample fit is constant across models.

3 Forecast evaluation

In this section we compare empirically forecasts resulting from different VAR specifications.

We compute point forecasts using the posterior mean of the parameters. We write $\hat{A}_j^{(\lambda,m)}, j = 1,..,p$ and $\hat{c}^{(\lambda,m)}$ for the posterior mean of the autoregressive coefficients and the constant term of a given model (m) obtained by setting the overall tightness equal to λ. The point estimates of the h-steps ahead forecasts are denoted by $Y_{t+h|t}^{(\lambda,m)} = \left(y_{1,t+h|t}^{(\lambda,m)}, ..., y_{n,t+h|t}^{(\lambda,m)}\right)'$, where n is the number of variables included in model m. The one-step ahead forecast is computed as $\hat{Y}_{t+1|t}^{(\lambda,m)} = \hat{c}^{(\lambda,m)} + \hat{A}_1^{(\lambda,m)} Y_t + ... + \hat{A}_p^{(\lambda,m)} Y_{t-p+1}$. Forecasts for $h > 1$ are computed recursively.

In the case of the benchmark model the prior restriction is imposed exactly, that is $\lambda = 0$. Corresponding forecasts are denoted by $Y_{t+h|t}^{(0)}$ and are the same for all models. Hence we drop the superscript m.

To simulate real-time forecasting we conduct an out-of-sample experiment. Let us denote by H the longest forecast horizon to be evaluated, and by T_0 and T_1 the beginning and the end of the evaluation sample, respectively. For a given forecast horizon h, in each period $T = T_0 + H - h, ..., T_1 - h$, we compute h-step-ahead forecasts, $Y_{T+h|T}^{(\lambda,m)}$, using only the information up to time T.

Out-of-sample forecast accuracy is measured in terms of Mean Squared Forecast Error (MSFE):

$$MSFE_{i,h}^{(\lambda,m)} = \frac{1}{T_1 - T_0 - H + 1} \sum_{T=T_0+H-h}^{T_1-h} \left(y_{i,T+h|T}^{(\lambda,m)} - y_{i,T+h} \right)^2 .$$

We report results for MSFE relative to the benchmark, that is

$$RMSFE_{i,h}^{(\lambda,m)} = \frac{MSFE_{i,h}^{(\lambda,m)}}{MSFE_{i,h}^{(0)}} .$$

Notice that a number smaller than one implies that the VAR model with overall tightness λ performs better than the naive prior model.

We evaluate the forecast performance of the VARs for the three key series included in all VAR specifications (Employment, CPI and the Federal Funds Rate) over the period going from $T_0 = Jan70$ until $T_1 = Dec03$ and for forecast horizons up to one year ($H = 12$). The order of the VAR is set to $p = 13$ and parameters are estimated using for each T the observations from the most recent 10 years (rolling scheme).[2]

The overall tightness is set to yield a desired average fit for the three variables of interest in the pre-evaluation period going from $Jan60$ ($t = 1$) until $Dec69$ ($t = T_0 - 1$) and then kept fixed for the entire evaluation period.

The in-sample fit is an in-sample measure of the 1-step-ahead mean squared forecast error evaluated using the training sample $t = 1, ..., T_0 - 1$. Precisely:

$$\text{msfe}_i^{(\lambda,m)} = \frac{1}{T_0 - p - 1} \sum_{t=p}^{T_0-2} (y_{i,t+1|t}^{(\lambda,m)} - y_{i,t+1})^2 ,$$

[2]Using all the available observations up to time T (recursive scheme) does not change the qualitative results. Qualitative results remain the same also if we set $p = 6$.

where the parameters are computed using the same sample $t = 1, ..., T_0 - 1$. We order the variables in all models so that the three key variables are in the first lines and we set the overall tightness for a given model (m) and for a given measure of fit (Fit) as:[3]

$$\lambda_m(Fit) = \arg\min_{\lambda} \left| \text{Fit} - \frac{1}{3}\sum_{i=1}^{3} \frac{\text{msfe}_i^{(\lambda,m)}}{\text{msfe}_i^{(0)}} \right|.$$

In the main text we report the results for the fit corresponding to that obtained from OLS estimation on the small model with $p = 13$. That is for

$$\text{Fit} = \frac{1}{3}\sum_{i=1}^{3} \frac{\text{msfe}_i^{(\lambda,m)}}{\text{msfe}_i^{(0)}} \Bigg|_{\lambda=\infty, m=\text{SMALL}}.$$

In the Appendix we present the results for a range of in-sample fits and show that they are qualitatively the same provided that the in-sample fit is not below 50%.

Table 1 presents the relative MSFE for forecast horizons $h = 1, 3, 6$ and 12. The specifications are listed in order of increasing size and the last row indicates the value of the shrinkage hyperparameter λ.

Table 1: BVAR, Relative MSFE, 1971-2003

		SMALL	CEE	MEDIUM	LARGE
h=1	CES002	1.14	0.67	0.54	0.46
	PUNEW	0.89	0.52	0.50	0.50
	FYFF	1.86	0.89	0.78	0.75
h=3	CES002	0.95	0.65	0.51	0.38
	PUNEW	0.66	0.41	0.41	0.40
	FYFF	1.77	1.07	0.95	0.94
h=6	CES002	1.11	0.78	0.66	0.50
	PUNEW	0.64	0.41	0.40	0.40
	FYFF	2.08	1.30	1.30	1.29
h=12	CES002	1.02	1.21	0.86	0.78
	PUNEW	0.83	0.57	0.47	0.44
	FYFF	2.59	1.71	1.48	1.93
λ		∞	0.262	0.108	0.035

Results show the following features:

[3]To obtain the desired magnitude of fit the search is performed over a grid for λ.

- Given that the in-sample fit is fixed, the required degree of shrinkage is larger the larger is the size of the model, i.e. the hyperparameter λ decreases with n. On this point, see also, De Mol, Giannone, and Reichlin (2006).

- Adding information helps to improve the forecast for all variables included in the table and across all horizons.

- Good results are already obtained with the medium size model containing twenty variables.

- Not surprisingly, the forecast of the federal funds rate does not improve over the simple random walk model beyond the first quarter. However, results for the first quarter are quite good.

3.1 Parsimony by lags selection

In VAR analysis there are alternative procedures to obtain parsimony. One alternative method to the BVAR approach is to implement information criteria for lag selection and then estimate the model by OLS. In what follows we will compare results obtained using these criteria to those obtained from the BVARs.

Table 2 presents the results for *SMALL* and *CEE*. We look at the forecast performance of the models with $p = 13$ or p selected at each evaluation period by the BIC criterion. For comparison, we also recall from Table 1 the results for the Bayesian estimation. We do not report the corresponding results for the larger models since for those the OLS estimation with $p = 13$ is unfeasible. However, we recall in the last column the results for the large model estimated in the Bayesian framework.

These results show that for the model *SMALL*, BIC selection results in the best forecast accuracy. For the larger *CEE* model, the traditional VAR with lags selected by BIC and the BVAR perform similarly. Both specifications are, however, outperformed by the large Bayesian VAR.

Table 2: OLS, Relative MSFE, 1971-2003

		SMALL			CEE			LARGE
		p = 13	p = BIC	BVAR	p = 13	p = BIC	BVAR	BVAR
h=1	CES002	1.14	0.73	1.14	7.56	0.76	0.67	0.46
	PUNEW	0.89	0.55	0.89	5.61	0.55	0.52	0.50
	FYFF	1.86	0.99	1.86	6.39	1.21	0.89	0.75
h=3	CES002	0.95	0.76	0.95	5.11	0.75	0.65	0.38
	PUNEW	0.66	0.49	0.66	4.52	0.45	0.41	0.40
	FYFF	1.77	1.29	1.77	6.92	1.27	1.07	0.94
h=6	EML	1.11	0.90	1.11	7.79	0.78	0.78	0.50
	PUNEW	0.64	0.51	0.64	4.80	0.44	0.41	0.40
	FYFF	2.08	1.51	2.08	15.9	1.48	1.30	1.29
h=12	CES002	1.02	1.15	1.02	22.3	0.82	1.21	0.78
	PUNEW	0.83	0.56	0.83	21.0	0.53	0.57	0.44
	FYFF	2.59	1.59	2.59	47.1	1.62	1.71	1.93

3.2 The Bayesian VAR and the Factor Augmented VAR (FAVAR)

Factor models have been shown to be successful at forecasting macroeconomic variables with a large number of predictors. It is therefore natural to compare forecasting results based on the Bayesian VAR with those produced by factor models where factors are estimated by principal components.

A comparison of forecasts based, alternatively, on Bayesian regression and principal components regression has recently been performed by De Mol, Giannone, and Reichlin (2006) and Giacomini and White (2006). In those exercises, variables are transformed to achieve stationarity which is a standard practice in the principal components literature. Moreover, the Bayesian regression is estimated as a single equation.

Here we want to perform an exercise in which factor models are compared with the standard VAR specification in the macroeconomic literature where variables are treated in levels and the model is estimated as a system rather than as a set of single equations. Therefore, for comparison with the VAR, rather than considering principal components regression, we will use a small VAR (with variables in levels) augmented by principal components extracted from the panel (in differences). This is the FAVAR method advocated by Bernanke, Boivin, and Eliasz (2005) and discussed by Stock and Watson (2005b).

More precisely, principal components are extracted from the entire set of 131 indicators. The variables are first made stationary by taking first differences wherever we have imposed a random walk prior $\delta_i = 1$. Then, as principal components are not scale invariant, variables are standardized and the factors are computed from the standardized panel, recursively at each point T in the evaluation sample.

We consider specifications with one and three factors and look at different lag selection strategies for the VAR. We consider the cases when $p = 13$, as in Bernanke, Boivin, and Eliasz (2005) and when it is selected by the BIC criterion. We also consider Bayesian estimation of the FAVAR (BFAVAR), taking $p = 13$ and choosing the shrinkage hyperparameter λ that results in the same in-sample fit as in the exercise summarized in Table 1.

Results are reported in Table 3 (the last column recalls results from the large Bayesian VAR for comparison).

Table 3: FAVAR, Relative MSFE, 1971-2003

		FAVAR 1 factor				FAVAR 3 factors				LARGE
		p = 13	p = BIC	BVAR		p = 13	p = BIC	BVAR		BVAR
h=1	CES002	1.36	0.54	0.70		3.02	0.52	0.65		0.46
	PUNEW	1.10	0.57	0.65		2.39	0.52	0.58		0.50
	FYFF	1.86	0.98	0.89		2.40	0.97	0.85		0.75
h=3	CES002	1.13	0.55	0.68		2.11	0.50	0.61		0.38
	PUNEW	0.80	0.49	0.55		1.44	0.44	0.49		0.40
	FYFF	1.62	1.12	1.03		3.08	1.16	0.99		0.94
h=6	CES002	1.33	0.73	0.87		2.52	0.63	0.77		0.50
	PUNEW	0.74	0.52	0.55		1.18	0.46	0.50		0.40
	FYFF	2.07	1.31	1.40		3.28	1.45	1.27		1.29
h=12	CES002	1.15	0.98	0.92		3.16	0.84	0.83		0.78
	PUNEW	0.95	0.58	0.70		1.98	0.54	0.64		0.44
	FYFF	2.69	1.43	1.93		7.09	1.46	1.69		1.93

The Table shows that the FAVAR is in general outperformed by the BVAR of large size and that therefore Bayesian VAR is a valid alternative to factor based forecasts.[4] We should also note that BIC lag selection generates the best results for the FAVAR while the traditional specification of BBE with $p = 13$ performs very poorly due to its lack of parsimony.

[4]Results from De Mol, Giannone, and Reichlin (2006) show that the forecast accuracy of more traditional principal components regression based on stationary variables is comparable to that of Bayesian regression.

4 Structural analysis: impulse response functions and variance decomposition

We now turn to the structural analysis and estimate, on the basis of BVARs of different size, the impulse responses of different variables to a monetary policy shock.

To this purpose, we identify the monetary policy shock by using a recursive identification scheme adapted to a large number of variables. We follow Bernanke, Boivin, and Eliasz (2005), Christiano, Eichenbaum, and Evans (1999) and Stock and Watson (2005b) and divide the variables in the panel into two categories: slow- and fast-moving. Roughly speaking the former group contains real variables and prices while the latter consists of financial variables (the precise classification is given in the Appendix). The identifying assumption is that slow-moving variables do not respond contemporaneously to a monetary policy shock and that the information set of the monetary authority contains only past values of the fast-moving variables.

The monetary policy shock is identified as follows. We order the variables as $Y_t = (X_t, r_t, Z_t)'$, where X_t contains the n_1 slow-moving variables, r_t is the monetary policy instrument and Z_t contains the n_2 fast-moving variables and we assume that the monetary policy shock is orthogonal to all other shocks driving the economy. Let $B = CD^{1/2}$ be the $n \times n$ lower diagonal Cholesky matrix of the covariance of the residuals of the reduced form VAR (see eq. 1), that is $CDC' = \mathbb{E}[u_t u_t'] = \Psi$ and $D = \text{diag}(\Psi)$.

Let e_t be the following linear transformation of the VAR residuals: $e_t = (e_{1t}, \ldots, e_{nt}) = B^{-1}u_t$. The monetary policy shock is the row of e_t corresponding to the position of r_t, that is $e_{n_1+1,t}$.

The Structural VAR can hence be written as

$$\mathcal{A}_0 Y_t = \nu + \mathcal{A}_1 Y_{t-1} + \cdots + \mathcal{A}_p Y_{t-p} + e_t, \quad e_t \sim WN(0, D)\,,$$

where $\nu = C^{-1}c$, $\mathcal{A}_0 = C^{-1}$ and $\mathcal{A}_j = C^{-1}A_j, j = 1, \ldots, p$.

Our experiment consists in increasing contemporaneously the federal funds rate by one basis point.

Since we have just identification, the impulse response functions are easily computed following Canova (1991) and Gordon and Leeper (1994) by generating draws from the posterior of

$(A_1, \ldots, A_p, \Psi)$. For each draw Ψ we compute C and B and we can then calculate $\mathcal{A}_j, j = 0, \ldots, p$.

We report the results for the same overall shrinkage as given in Table 1. Estimation is based on the sample 1961-2002. The number of lags remains 13.

Figure 1 displays the impulse response functions for four models under consideration for the three variables of interest. The shaded regions indicate the confidence intervals corresponding to 90 and 68 percent confidence levels. Table 4 reports the percentage share of the monetary policy shock in the forecast error variance for chosen forecast horizons.

Figure 1: BVAR, Impulse response functions

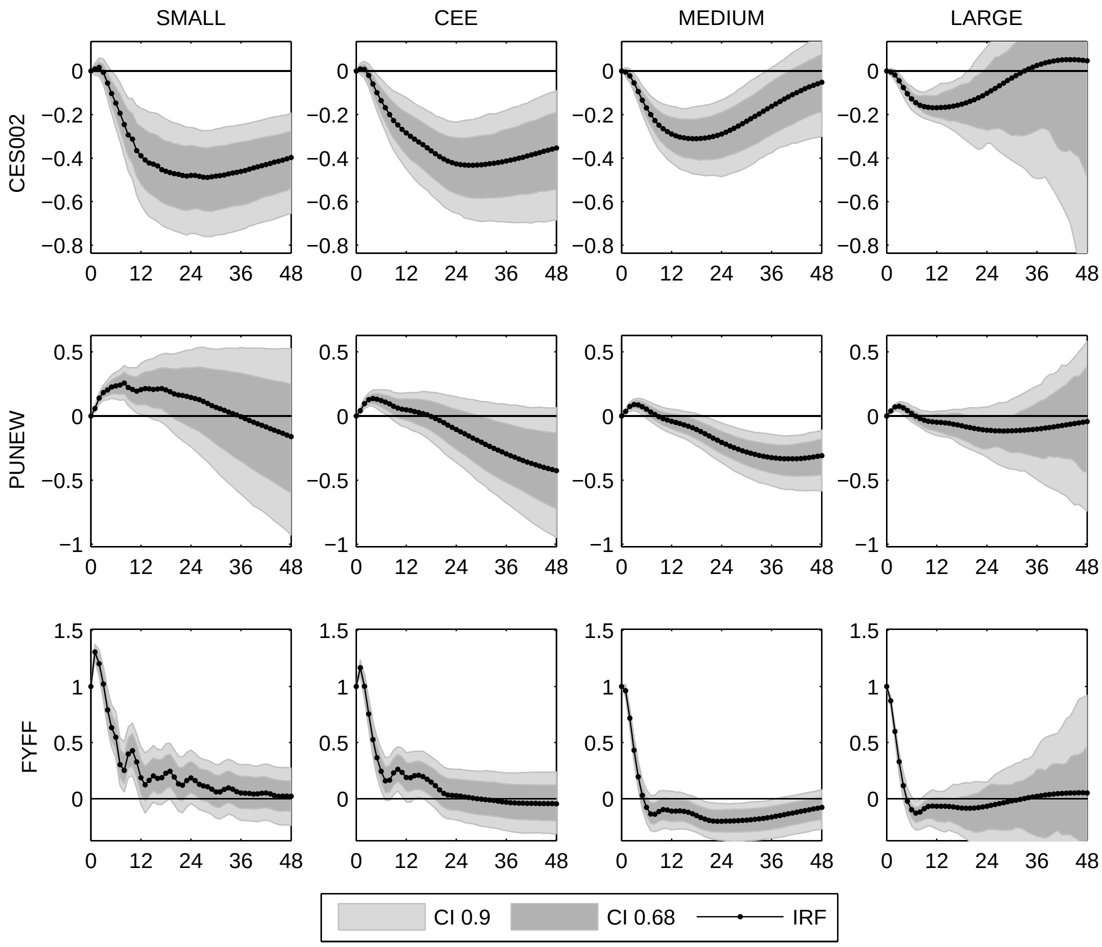

Figure 2 and 3 show the impulse responses for *MEDIUM* and *LARGE* for all 20 series included in *MEDIUM*. Impulse responses and variance decomposition for all the variables and models are reported in the Appendix.

Results show that, as we add information, impulse response functions slightly change in shape which suggests that realistic informational assumptions are important for structural analysis as well as for forecasting. In particular, it is confirmed that adding variables helps in resolving the price puzzle (on this point see also Bernanke and Boivin, 2003; Christiano, Eichenbaum, and Evans, 1999). Moreover, for larger models the effect of monetary policy on employment becomes less persistent, reaching a peak at about one year horizon. For the large model, the non-systematic component of monetary policy becomes very small, confirming results in Giannone, Reichlin, and Sala (2004) obtained on the basis of a factor model. It is also important to stress that impulse responses maintain the expected sign for all specifications.

However, one problem clearly emerges for the large specification: the confidence intervals of the impulse responses become very large for horizons beyond two years, eventually becoming explosive. This problem is the consequence of the fact that the variables have been introduced in levels, which, in the case of the large model, produces explosive draws. In the next section we show how to solve this problem by "inexact differencing", i.e. by setting priors constraining the sums of coefficients.

As for the variance decomposition we can see that, as the size of the model increases, the size of the monetary shock decreases. This is not surprising, given the fact that the forecast accuracy improves with size, but it highlights an important point. If realistic informational assumptions are not taken into consideration, we may mix structural shocks with miss-specification errors. Clearly, the assessment of the importance of the systematic component of monetary policy depends on the conditioning information set used by the econometrician and this may differ from the one that is relevant for policy decisions. Once the practically relevant assumption of rich available information is taken into account by the econometrician, the estimate of the size of the non-systematic component decreases.

Figure 2: BVAR, Impulse response function for model *MEDIUM*

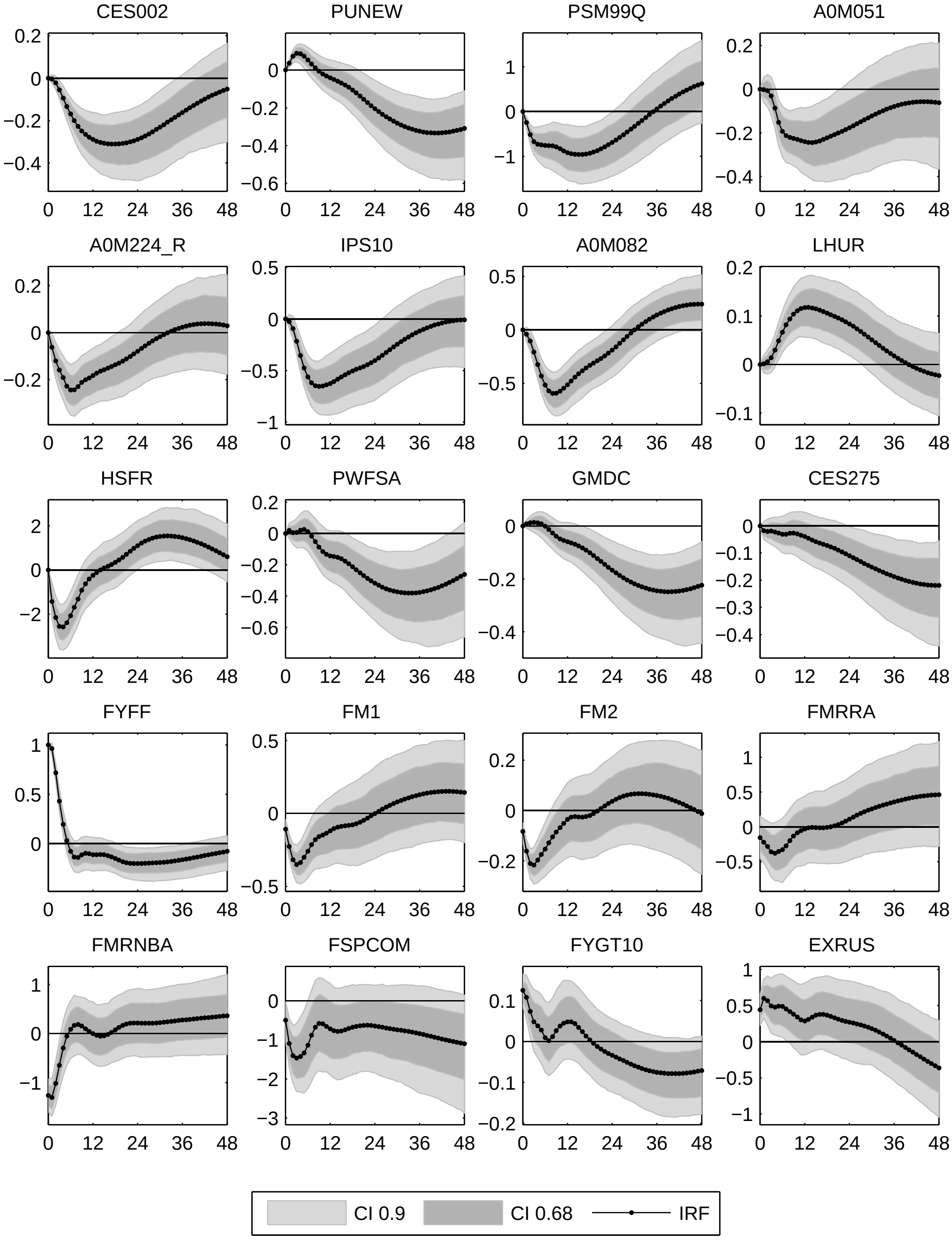

Figure 3: BVAR, Impulse response function for model *LARGE*

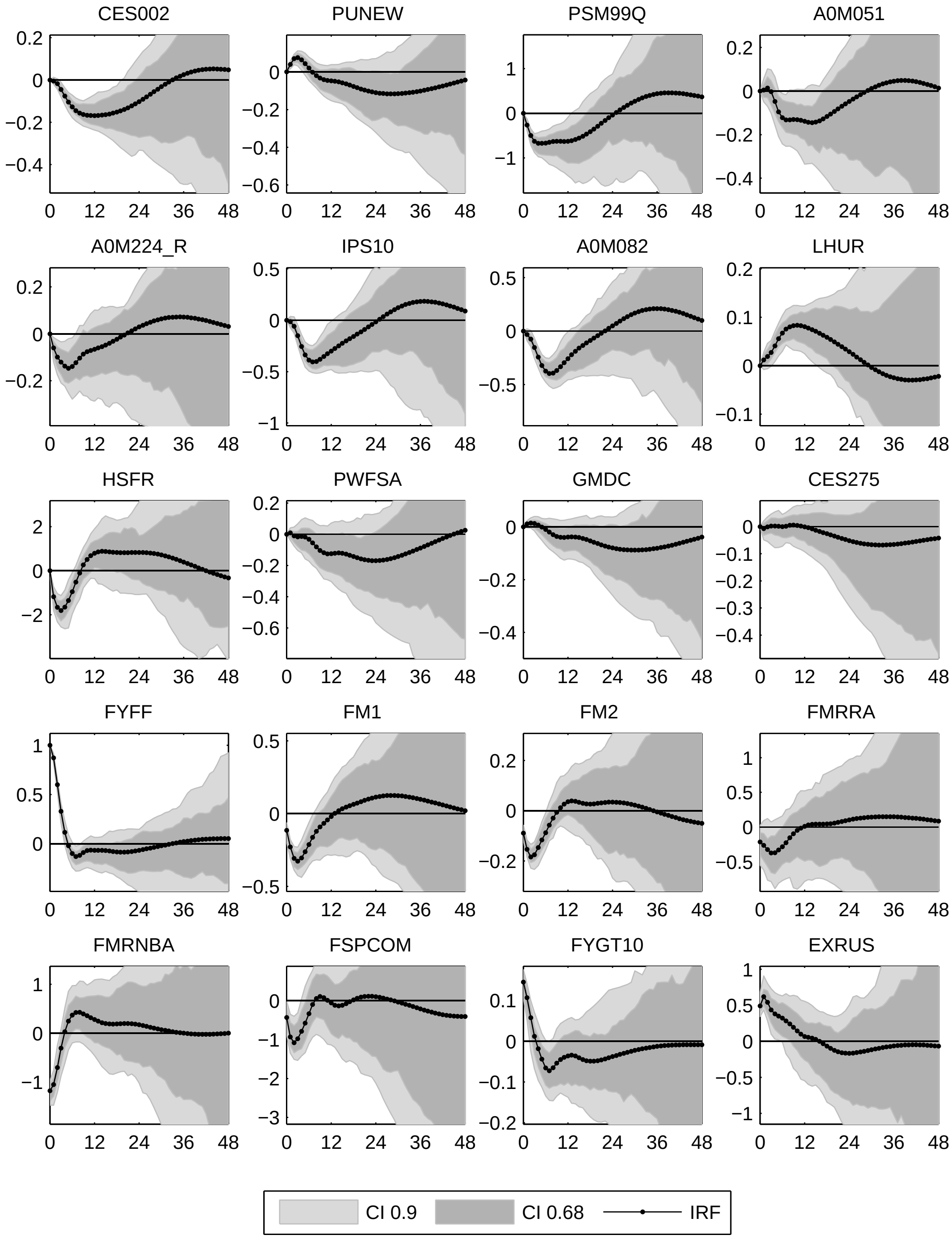

Table 4: BVAR, Variance Decomposition, 1961-2002

	Hor	SMALL	CEE	MEDIUM	LARGE
CES002	1	0	0	0	0
	3	0	0	0	0
	6	1	1	2	1
	12	5	4	8	3
	24	12	10	10	2
	36	18	15	7	1
	48	23	17	5	1
PUNEW	1	0	0	0	0
	3	3	2	1	1
	6	7	5	2	1
	12	6	2	1	1
	24	2	1	2	1
	36	1	1	5	1
	48	1	3	8	1
FYFF	1	99	96	93	53
	3	90	81	69	38
	6	74	60	43	22
	12	46	34	27	12
	24	26	19	18	7
	36	21	16	15	5
	48	18	14	13	4

5 Priors on the sums of coeffcients

This Section aims at finding a specification that solves the problem of explosive draws for the medium and long run. To this end, we add a prior on the sum of coefficients of the VAR. This is a simple modification of the Minnesota prior involving linear combinations of the VAR coefficients, see Doan, Litterman, and Sims (1984).

Let us rewrite the VAR of equation (1) in its error correction form:

$$\Delta Y_t = c - (I_n - A_1 - \cdots - A_p)Y_{t-1} + B_1 \Delta Y_{t-1} + \cdots + B_{p-1}\Delta Y_{t-p+1} + u_t\,. \tag{8}$$

A VAR in first differences implies the restriction $(I_n - A_1 - \cdots - A_p) = 0$. We follow Doan, Litterman, and Sims (1984) and set a prior that shrinks $\Pi = (I_n - A_1 - \cdots - A_p)$ to zero. This can be understood as "inexact differencing" and in the literature it is usually implemented by adding the following dummy observations (cf. Section 2):

$$Y_d = \mathrm{diag}(\delta_1\mu_1, \dots, \delta_n\mu_n)/\tau \qquad X_d = \begin{pmatrix} (1\ 2\ \dots\ p) \otimes \mathrm{diag}(\delta_1\mu_1, \dots, \delta_n\mu_n)/\tau & 0_{n\times 1} \end{pmatrix}. \quad (9)$$

The hyperparameter τ controls for the degree of shrinkage: as τ goes to zero we approach the case of exact differences and, as τ goes to ∞, we approach the case of no shrinkage. The parameter μ_i aims at capturing the average level of variable y_{it}. Although the parameters should in principle be set using only prior knowledge, we follow common practice[5] and set the parameter equal to the sample average of y_{it}.

Our approach is to set a loose prior with $\tau = 10\lambda$. Figure 4 reports impulse response functions for this choice and for a selected number of variables of interest computed using the large model. Qualitatively, results remain the same, but the explosive behavior of the bands from 24 lags onward is now corrected.

Impulse responses and variance decomposition for all the models and all the variables are provided in the Appendix. Concerning smaller models (*SMALL*, *CEE* and *MEDIUM*), imposing the inexact differencing does not change qualitatively the impulse responses. For a robustness check the Appendix also reports results for less informative priors such as $\tau = 100\lambda$ and more informative priors such as $\tau = \lambda$. We find that the qualitative results are robust with respect to the tightness of the prior.

Table 6 reports results from the forecast evaluation of the specification with the sum of coefficients prior. They show that imposing this additional prior improves forecast accuracy, confirming the findings of Robertson and Tallman (1999).

6 Summary and conclusions

This paper assesses the performance of Bayesian VAR on models of different size. We consider standard specifications in the literature with three and eight variables and also study VARs with twenty and a hundred and thirty variables. We examine both forecasting accuracy and structural analysis of the effect of a monetary policy shock.

[5]See for example Sims and Zha (1998).

Figure 4: BVAR, Impulse response functions, sum of coefficients prior

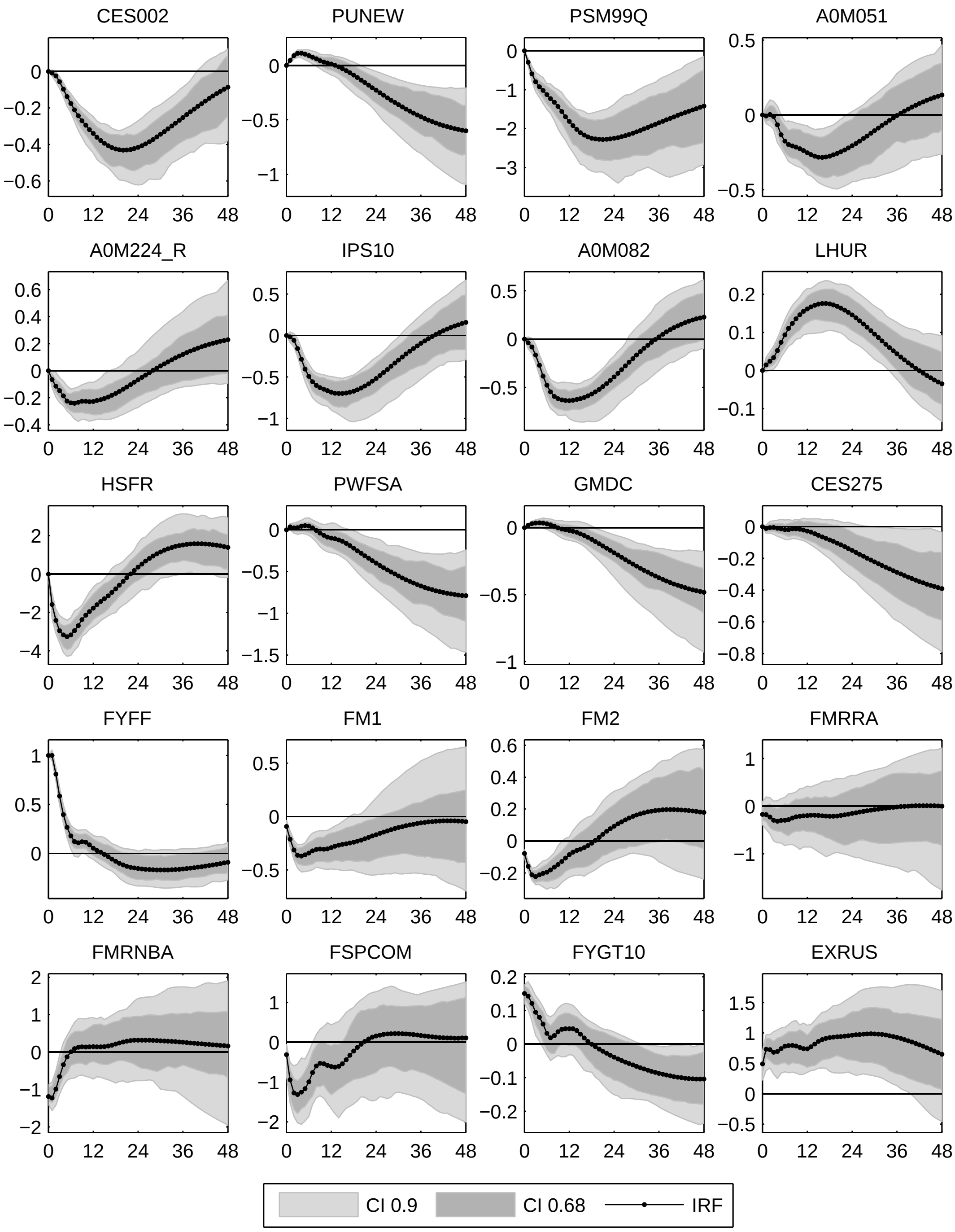

Table 5: BVAR, Relative MSFE, 1971-2003, $\tau = 10\lambda$

		SMALL	CEE	MEDIUM	LARGE
h=1	CES002	1.14	0.68	0.53	0.44
	PUNEW	0.89	0.57	0.49	0.49
	FYFF	1.86	0.97	0.75	0.74
h=3	CES002	0.95	0.60	0.49	0.36
	PUNEW	0.66	0.44	0.39	0.37
	FYFF	1.77	1.28	0.85	0.82
h=6	CES002	1.11	0.65	0.58	0.44
	PUNEW	0.64	0.45	0.37	0.36
	FYFF	2.08	1.40	0.96	0.92
h=12	CES002	1.02	0.65	0.60	0.50
	PUNEW	0.83	0.55	0.43	0.40
	FYFF	2.59	1.61	0.93	0.92

The setting of the prior follows standard recommendations in the Bayesian literature except for the fact that the overall tightness hyperparameter is set in relation to the model size where we follow the recommendation by De Mol, Giannone, and Reichlin (2006) and increase the overall shrinkage with the cross-sectional dimension of the model.

Overall, results show that a standard Bayesian VAR model is an appropriate tool for large panels of data and constitutes a valid alternative to factor models for dealing with the curse of dimensionality problem. In fact, Bayesian VAR and factor models provide alternative strategies to restrict the parameter space in large models. Given our results, it is indeed surprising that BVAR has not been applied to panels of large data typically involving the information used routinely by central bankers and private agents.

References

Bernanke, B., J. Boivin, and P. Eliasz (2005): "Measuring Monetary Policy: A Factor Augmented Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.

Bernanke, B. S., and J. Boivin (2003): "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50(3), 525–546.

Boivin, J., and S. Ng (2005): "Understanding and Comparing Factor-Based Forecasts," *International Journal of Central Banking*, 3, 117–151.

Canova, F. (1991): "The Sources of Financial Crisis: Pre- and Post-Fed Evidence," *International Economic Review*, 32(3), 689–713.

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.

D'Agostino, A., and D. Giannone (2006): "Comparing alternative predictors based on large-panel factor models," Working Paper Series 680, European Central Bank.

De Mol, C., D. Giannone, and L. Reichlin (2006): "Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?," CEPR Discussion Papers 5829.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARS," *International Economic Review*, 45(2), 643–673.

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

Favero, C. A., M. Marcellino, and F. Neglia (2005): "Principal components at work: the empirical analysis of monetary policy with large data sets," *Journal of Applied Econometrics*, 20(5), 603–620, available at http://ideas.repec.org/a/jae/japmet/v20y2005i5p603-620.html.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2005): "Opening the Black Box: Structural Factor Models with large cross-sections," Manuscript, Université Libre de Bruxelles.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003): "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?," *Journal of Monetary Economics*, 50, 1243–55.

——— (2005): "The Generalized Dynamic Factor Model: one-sided estimation and forecasting," *Journal of the American Statistical Association*, 100, 830–840.

Giacomini, R., and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, forthcoming.

Giannone, D., and L. Reichlin (2006): "Does information help recovering structural shocks from past observations?," *Journal of the European Economic Association*, 4(2-3), 455–465.

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.

Gordon, D. B., and E. M. Leeper (1994): "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy*, 102(6), 1228–47.

Hamilton, J. D. (2006): "Computing power and the power of econometrics," Manuscript, University of California, San Diego.

Kadiyala, K. R., and S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12(2), 99–132.

Leeper, E. M., C. Sims, and T. Zha (1996): "What Does Monetary Policy Do?," *Brookings Papaers on Economic Activity*, (4), 1–63.

Litterman, R. (1986): "Forecasting With Bayesian Vector Autoregressions – Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

Marcellino, M., J. H. Stock, and M. W. Watson (2003): "Macroeconomic forecasting in the Euro area: Country specific versus area-wide information," *European Economic Review*, 47(1), 1–18.

Robertson, J. C., and E. W. Tallman (1999): "Vector autoregressions: forecasting and reality," *Economic Review*, (Q1), 4–18.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–68.

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.

——— (2005a): "An Empirical Comparison Of Methods For Forecasting Using Many Predictors," Manuscript, Princeton University.

——— (2005b): "Implications of Dynamic Factor Models for VAR Analysis," Unpublished manuscript, Princeton University.