Forecasting with Bayesian Multivariate Vintage-Based VARs

Andrea Carriero
Queen Mary University of London

Michael P. Clements
University of Warwick

Ana Beatriz Galvão
Queen Mary University of London

December, 2012
PRELIMINARY

Abstract

We propose modelling and forecasting macroeconomic variables that are subject to revision using a Bayesian vintage-based vector autoregressions. The prior incorporates the belief that data releases after the first few are likely to be efficient. The Bayesian approach allows the dimensionality increase from jointly modelling data revisions while keeping parameter uncertainty manageable. Our model is shown to provide markedly more accurate forecast of inflation than other models in the literature.

Keywords: Bayesian VARs, multiple-vintage models, forecasting, output growth, inflation. JEL code: C53.
1 Introduction

Economists who are asked to compute forecasts of macroeconomic variables such as output growth and inflation at a given point in time have to do so using data which they know will subsequently be revised.¹ That data are subject to revision raises issues to do with the appropriate criterion for assessing forecast accuracy. A researcher may subsequently choose to evaluate the accuracy of these forecasts against post-revision or fully-revised actual values, or against an earlier release of the ‘actual’ values. A justification for the former is that the post-revision data will typically provide the most accurate estimates of the true values of the variables, whereas the latter acknowledges that these values may contain the effects of methodological changes in the measurement system which could not reasonably have been foreseen. A more fundamental question than the issue of forecast evaluation is the potential impact of data revision on model specification and estimation. At any forecast origin, the more recent observations are only lightly revised relative to the more heavily-revised earlier data, and Koenig, Dolmas and Piger (2003) and Clements and Galvão (2012d) amongst others have shown that the traditional approach to (real-time) forecasting ignores this aspect to its detriment, and suggest alternative approaches.

The recent academic literature has proposed explicitly considering the multiple-vintages of data that the forecaster has access to at any point in time to model the data revisions process, see, e.g., Patterson (2003), Kishor and Koenig (2011), Cunningham, Eklund, Jeffery, Kapetanios and Labhard (2009), Garratt, Lee, Mise and Shields (2008, 2009). Some of the leading multiple-vintage models have been evaluated by Clements and Galvão (2012c). For our purposes, they present two key findings. Firstly, vintage-based VARs (described fully below) provide competitive forecasts of the post-revision values of data for which the forecaster has already observed an earlier release, i.e., they are able to forecast data revisions better than natural comparators (such as assuming revisions to already-released data are zero). Secondly, vintage-based VARs are not able to forecast post-revision values of future observations more accurately than simple autoregressive models.

Two potential reasons for the failure to improve forecasts of future observations are addressed in this paper. The first is that the vintage-based VARs typically used in the literature (Garratt, Lee, Mise and Shields, 2008; Clements and Galvão, 2012c; Clements and Galvão, 2012b) are univariate: they model the revisions to a single variable. This may be overly restrictive. For example, if we are

¹Landefeld, Seskin and Fraumeni (2008) describe the way in which US national accounts data are revised over time.
interested in modelling inflation, the Phillips Curve suggests a role for an activity variables such as output (Stock and Watson, 2007; Stock and Watson, 2009; Stock and Watson, 2010). Restricting the analysis to a single variable avoids the parameter proliferation from jointly modelling two or more variables and their data revisions, but may be costly in terms of forecasting future observations.

In this paper, we consider multivariate multiple-vintage models. The second reason is that even restricting the number of variables to one, as is the univariate multiple-vintage models typical in the literature, the effects of parameter-estimation uncertainty are likely to be large, and may seriously detract from forecast performance. To capture the nature of the Bureau of Economic Analysis revisions process, Clements and Galvão (2012c) model a vector of 14 elements, so that even for a first-order VAR their are $14^2$ slope parameters to estimate.

In this paper we employ a Bayesian approach which offers a solution to both these problems, permitting multiple-variable multiple-vintage models. On the one hand, the coefficients are shrunk towards prior values, attenuating the ‘curse of dimensionality’ that afflicts classical VARs when the number of variables increases beyond a handful. In this way it is possible to efficiently summarize the information contained in a large set of vintages. On the other hand, the nature of revisions lends itself to the specification of an economically meaningful prior, which is tailored to accommodate the joint modelling of the data revision process and the process for the revised values. In particular, the proposed Bayesian Vintage-based VAR (BVB-VAR) model has a prior for the VAR coefficients that incorporates the assumption that earlier revisions are predictable while later revisions are not because they primarily add new information.

Based on an estimated BVB-VAR model we can compute the variance decomposition of shocks arising both from future values and data revisions. This helps us to measure the relative impact of data and forecasting uncertainties.

We design a real-time forecasting exercise to address two interesting questions. First, we look at the potential improvements from using a bivariate BVB-VAR relative to a univariate unrestricted VB-VAR. The forecasting exercise is formulated to allow one to disentangle the gains from the use of an additional explanatory variable from those resulting from the adoption of a Bayesian approach. Second, we assess the forecasting improvements from jointly modelling the data revision process when forecasting in real time. We compare the BVB-VAR forecasting performance with a benchmark Bayesian VAR (as in Kadiyala and Karlsson (1997), Sims and Zha (1998), Bańbura, Giannone and Reichlin (2010), and Carriero, Clark and Marcellino (2012)). The Bayesian VAR is estimated with (only) the latest-available vintage at each forecast origin, essentially disregarding
the data revision process. An autoregressive model estimated with real-time data is also included in the comparison since it represents a simple forecasting model that takes into the account the nature of real-time data in the estimation, following Koenig et al. (2003) and Clements and Galvão (2012d).

The plan of the rest of the paper is as follows. Section 2 describes the forecasting models. Section 3 describes the results of the forecast comparison. Section 4 offers some concluding remarks.

2 The Forecasting Model

2.1 Univariate vintage-based VAR

This section describes the univariate V-VAR of Clements and Galvão (2012c): see the former for the relationship between the model described here and that of Garratt et al. (2008). The models are based on the quarterly data vintages made available in the Real-Time Dataset for Macroeconomists (Stark and Croushore, 2002). We write the first release of the variable \( y \) as \( y_{t+1} \). The superscript on \( y \) denotes the data vintage, and the subscript the time period to which the observation refers. So \( y_{t+1} \) is the estimate of \( y_t \) available around the 15th day of the middle month of the following quarter (quarter \( t+1 \)). So \( y_{t+2} \) is the estimate of \( y_t \) available around the 15th day of the middle month of quarter \( t+2 \), and so on. Suppose there are revisions for the next \( q \) quarters after the first release, but thereafter the observation is unrevised (i.e., \( y_{t+q+i} = y_{t+q} \) for \( i > 0 \)). Then we can model the vintage \( t+1 \) values of observations \( t+q+1 \) through \( t \) as a VB-VAR:

\[
y^{t+1} = \mathbf{c} + \sum_{i=1}^{p} \mathbf{\Gamma}_i y^{t+1-i} + \mathbf{\varepsilon}^{t+1}
\]

(1)

where \( y^{t+1-i} = [y_{t+1-i}, y_{t+1-i}, \ldots, y_{t-q+1-i}]' \), \( i = 0, 1 \ldots, p \), and \( \mathbf{c} \) is \( q \times 1 \). \( \mathbf{\varepsilon}^{t+1} \) is \( q \times 1 \). The variance-covariance matrix of the disturbances (\( \Sigma_\varepsilon = E(\mathbf{\varepsilon}^{t+1}\mathbf{\varepsilon}^{t+1'}) \)) captures the correlations between data published in the same vintage. Even when \( p = 1 \), as in Clements and Galvão (2012c), with \( q = 14 \) there are already a large number of parameters to estimate. A restricted V-VAR model imposes a large number of restrictions, based on the belief that after a small number of revisions, further revisions are unpredictable. Suppose that after \( n-1 \) revisions, the next estimate \( y_{t+n+1}^{t+n+1} \) is an efficient forecast in the sense that the revision from \( y_{t+n}^{t+n} \) to \( y_{t+n+1}^{t+n+1} \) is unpredictable, i.e., \( E [ (y_{t+n+1}^{t+n+1} - y_{t+n}^{t+n}) | y^{t+n} ] = 0 \), whereas \( E [ (y_{t+i+1}^{t+i+1} - y_{t+i}^{t+i}) | y^{t+i} ] \neq 0 \) for \( i < n \). We can impose this restriction on the VAR with \( p = 1 \), where it translates to \( E (y_{t-n}^{t+1} | y_{t-n}^{t}) = y_{t-n}^{t} \). This is achieved by specifying \( \mathbf{\Gamma}_1 \) in (1) as:
\( \tilde{\Gamma}_1 = \begin{bmatrix} \gamma_{n \times q} \\ 0_{(q-n) \times (n-1)} & I_{(q-n) \times (q-n)} & 0_{(q-n) \times 1} \end{bmatrix}. \)  

(2)

In the empirical work, we set \( n = 2 \), so that values after the first revision are assumed to be efficient forecasts (i.e., the BEA estimate published two quarters after the period to which it refers is an efficient forecast). An unrestricted intercept is included in each equation, to accommodate non-zero mean revisions. We refer to this model as the ‘news-restricted’ vintage-based VAR, RV-VAR.

2.2 Multivariate vintage-based VAR

In (1) we model the \( t+1 \)-vintage estimates of \( y_{t-q+1} \) through \( y_t \) of the same variable (e.g., for output growth, or for inflation). Suppose we now have \( m \) variables, and define:

\[
\mathbf{x}^{t+1} = [y_{t+1}^1, \ldots, y_{t+1}^m]
\]

(3)

which is a vector of length \( mq \). The multivariate vintage-based VAR assuming \( p = 1 \) is:

\[
\mathbf{x}^{t+1} = \mathbf{c} + \mathbf{\Psi} \mathbf{x}^t + \mathbf{v}^{t+1},
\]

(4)

where \( \mathbf{c} \) and \( \mathbf{\Psi} \) are \( mq \times 1 \) and \( mq \times mq \) respectively, and \( \mathbf{v}^{t+1} = [\varepsilon_1^{t+1}, \ldots, \varepsilon_m^{t+1}] \) is a vector of length \( mq \). The disturbances have variance \( \Sigma_v = E[\mathbf{v}^{t+1} \mathbf{v}^{t+1\prime}] \), which is in general full. It measures the correlations in the disturbances to the estimates of a given variable, and to the estimates across variables, published in the same vintage.

2.3 Vintage-based Bayesian VARs

In this section we describe our vintage-based Bayesian VAR. The main difference with the approach of equation (4) is that the coefficient matrix \( \mathbf{\Psi} \) is modelled as a random variable, which in turn allows the imposition of sets of restrictions, such as those in (2). The restrictions do not have to be imposed sharply, as in the RV-VAR, but can be implemented allowing for some random noise around them. From a Bayesian perspective, this amounts to imposing a prior on the matrix \( \mathbf{\Psi} \). In our application the prior beliefs will be tailored to accommodate the joint modelling of the data revision process and the process of the revised values. In particular, we consider two alternative prior specifications.

Our baseline specification is closely related to the Minnesota prior of Doan, Litterman and Sims (1984) and Litterman (1986). Under this prior \( \mathbf{\Psi} \) has a normal prior distribution, while the
error variance matrix $\Sigma_v$ is assumed to be non-random, and is estimated in a preliminary step. Treating the matrix $\Sigma_v$ as fixed has the advantage that a closed form solution is available for the joint posterior of the model coefficients as well as for the marginal likelihood (which is needed to optimally select the degree of precision on a given prior).\footnote{Closed form solutions for the joint posteriors are also available in the case of random error variance, but only under the natural-conjugate Normal Inverse-Wishart prior. However, such prior requires a kronecker structure for the prior variance of the coefficients, and it is not implementable for non-symmetric priors (i.e. priors that are not the same - up to multiplicative constants - in each equation of the VAR). As we discuss below, the prior used in this paper does not have this characteristic.} A BVAR with a fixed $\Sigma_v$ also has a straightforward interpretation in terms of sampling theory, because it is possible to show that Bayesian estimation of this model coincides with the mixed estimator of Theil and Goldberger (1961), which is in turn equivalent to Generalized Least Squares estimation of a VAR estimated under a set of uncertain restrictions.

A case could be made that for our application the benefits of assuming a non-random $\Sigma_v$ more than offset any costs. Nonetheless, for completeness, we also estimate a version of the model with $\Sigma_v$ treated as random. In this case, we specify a prior distribution for $\Psi|\Sigma_v$ and $\Sigma_v$ which yields conditionally conjugate posteriors $\Psi|\Sigma_v, y$ and $\Sigma_v|\Psi, y$, which are then used in turn within a Gibbs-sampling algorithm to recover the marginal posteriors $\Psi|y$ and $\Sigma_v|y$.

### 2.4 Priors

In both the baseline and the alternative specifications, the element $\Psi^{(ij)}$ in row $i$ and column $j$ of the matrix of coefficients $\Psi$ appearing in equation (4) is assumed to be independent from any other element $\Psi^{(ij)}$ and to have a Gaussian distribution with prior expectation $E[\Psi^{(ij)}]$ and prior variance $\text{Var}[\Psi^{(ij)}]$.

We calibrate the prior means $E[\Psi^{(ij)}]$ so that the estimates are shrunk towards the RV-VAR model described in Section 2. In our application we have $q = 14$ and $m = 2$. In this case, the prior
expectations can be collected in the following matrix:

\[
E[\Psi] = \begin{bmatrix}
\hat{g}_{11} & \hat{g}_{12} & \cdots & \hat{g}_{1q} & \tilde{f}_{11} & \tilde{f}_{12} & \cdots & \tilde{f}_{1q} \\
1 & \hat{g}_{22} & \hat{g}_{23} & \cdots & \hat{g}_{21} & \tilde{f}_{21} & & 0 & \cdots & \tilde{f}_{21} \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ddots & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\
\hat{h}_{11} & \hat{h}_{12} & \cdots & \hat{h}_{1q} & \tilde{l}_{11} & \tilde{l}_{12} & \cdots & \tilde{l}_{1q} \\
\hat{h}_{21} & \hat{h}_{21} & 0 & \cdots & \hat{h}_{2q} & 1 & \tilde{l}_{22} & \tilde{l}_{23} & \cdots & \tilde{l}_{2q} \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\]  

The prior variances \( \text{Var}[\Psi^{(ij)}] \) are:

\[
\text{VAR}[\Psi] = \begin{bmatrix}
\lambda_1(2sd(\hat{g}_{11}))^2 & \lambda_1\frac{\phi_1}{\sigma} & \cdots & \lambda_1\frac{\phi_q}{\sigma} & \lambda_1(2sd(\tilde{f}_{11}))^2 & \lambda_1\frac{\phi_1}{\sigma} & \cdots & \lambda_1\frac{\phi_q}{\sigma} \\
\lambda_1\phi_3 & (2sd(\hat{g}_{21}))^2 & \cdots & (\frac{\phi_1}{\sigma})^2 & \lambda_1\phi_3 & \lambda_1\frac{\phi_1}{\sigma} & \cdots & \lambda_1\frac{\phi_q}{\sigma} \\
\lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \cdots & \lambda_1\phi_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \lambda_1\phi_3 & \cdots & \lambda_1\phi_3 \\
\lambda_2(2sd(\hat{h}_{11}))^2 & \lambda_2\frac{\phi_1}{\sigma} & \cdots & \lambda_2\frac{\phi_q}{\sigma} & \lambda_2(2sd(\tilde{l}_{11}))^2 & \lambda_2(2sd(\tilde{l}_{12}))^2 & \cdots & \lambda_2\frac{\phi_3}{\sigma} \\
\lambda_2\phi_3 & \lambda_2\frac{\phi_1}{\sigma} & \cdots & \lambda_2\frac{\phi_q}{\sigma} & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\frac{\phi_3}{\sigma} & \cdots & \lambda_2\phi_3 \\
\lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \cdots & \lambda_2\phi_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 \\
\end{bmatrix}
\]

Finally, the covariances among any element of \( \Psi \) are assumed to be zero, and a diffuse prior with mean 0 and variance \( \phi_0 = 1 \) is elicited for the intercepts.

The prior means \( \hat{g}_{ij}, \hat{h}_{ij}, \tilde{l}_{ij} \) and \( \tilde{f}_{ij} \) for \( i = 1, 2 \) and \( j = 1, \ldots, q \) are set equal to the OLS
estimates in the following models:

\[ y_{1,t+1} = c_1 + g_{11}^0 y_{1,t-1} + f_{11}^0 y_{2,t-1} + \text{error} \quad (7) \]

\[ y_{2,t+1} = c_2 + h_{11}^0 y_{1,t-1} + f_{11}^0 y_{2,t-1} + f_{12}^0 y_{2,t-2} + f_{13}^0 y_{2,t-3} + f_{14}^0 y_{2,t-4} + \text{error} \quad (8) \]

\[ y_{1,t-1} - y_{1,t-1} = c_3 + g_{22}^0 y_{1,t-2} + \text{error} \quad (9) \]

\[ y_{2,t-1} - y_{2,t-1} = c_4 + h_{22}^0 y_{2,t-2} + \text{error} \quad (10) \]

These are estimated using real-time vintage data on a pre-sample period consisting of the first 20 observations in the data set. The prior variances on these coefficients are set to twice the standard deviation estimated on the pre-sample period. The remaining prior means on \( \hat{g}_{ij} \) and \( \hat{l}_{ij} \) are set to 0. On these coefficients we set a variance equal to \( \phi_1/i^2 \) and \( \phi_2/i^2 \). This structure of prior variances resembles that of the Minnesota prior as the variance of the coefficients on lagged variables decreases with the lag with quadratic decay. We set \( \phi_1 = 0.02 \) and \( \phi_2 = 0.01 \), since the variation of output growth is twice as large as the one of inflation.

The prior means \( \hat{f}_{11} \) and \( \hat{h}_{11} \) (i.e. the prior means on the cross-dynamic coefficients that are measuring the effect of \( y_{1,t-1} \) on \( y_{2,t}^{l+1} \)) are set to the OLS estimates from the models in (7) and (8), again based on the 20-observation pre-sample period. The prior variances are set to twice the standard deviation estimated on the pre-sample period. On the remaining \( \hat{h}_{ij} \) and \( \hat{f}_{ij} \) coefficients, we set a prior mean of 0, with prior variances equal to \( \kappa \phi_1/i^2 \) and \( \kappa \phi_2/i^2 \) where \( i = 2, \ldots, q \), and \( \kappa = 0.5 \). This structure of prior variances resembles that of the Minnesota prior as the variance of the coefficients on lagged variables decreases with the lag with quadratic decay, and for those coefficients which are not multiplying some lag of the dependent variable there is an additional shrinkage as measured by the parameter \( \kappa \).

The prior on the equation of second estimates restricts \( \hat{g}_{21} = \hat{l}_{21} = 1 \), but instead of a prior mean of zero as for the remaining estimates, we use a pre-sample to obtain the values of \( \hat{g}_{22} \) and \( \hat{l}_{22} \) as indicated in equations (9) and (10). This is compatible with the support in the literature for predictability of initial revisions (Clements and Galvão, 2012a) The 0 and 1 restrictions appearing in all the remaining equations are given a very informative prior featuring variance \( \phi_3 = 0.001^2 \), or \( 2\phi_3 \) is referring to cross terms.

Finally, the hyperparameters \( \lambda_1 \) and \( \lambda_2 \) rescale the overall variance for the equations related in a given variable in the VAR. As we shall discuss below, they will be chosen optimally by maximising the marginal likelihood of the system.
2.5 Baseline specification

In this subsection we consider the baseline specification of our model, characterized by the prior moments (5) and (6) and by a fixed error variance matrix $\Sigma_v$. Consider all vintages $t = 1, \ldots, T$ and rewrite the system in (4) as follows by stacking observations by column:

$$x^+ = XA + v^+$$

where $x^+ = [x^T \ldots x^{T+1}]'$ and $v^+ = [v^T \ldots v^{T+1}]'$, while $X = [i x^T]$ contains an intercept ($i$ is a $T$-dimensional vector of ones) as well as the lags of $x^+$. $A = [c \Psi']$, that is, a $(mq + 1) \times mq$ matrix. Then we vectorize the system as follows:

$$x = \Xi \alpha + v$$

(11)

with $x = vec(x^+)$, $\Xi = I \otimes X$, $\alpha = vec(A)$, and $v = vec(v^+)$. Given a sample size $T$, $x$ and $v$ are $qmT \times 1$ vectors, and $\Xi$ is the $qmT \times (qm + 1)$ matrix of regressors. Defining $\Sigma_v$ as the variance matrix of the disturbances in equation (11), the vector $v$ of disturbances of the vectorized model has a multivariate normal distribution with variance $\Omega_0 = \Sigma_v \otimes I_T$. Note that the variance matrix $\Omega_0$, conditional on the knowledge of $\Sigma_v$, is not a random variable. The value $\Omega_0$ is derived by setting $\Sigma_v$ to the OLS estimates obtained from VAR(1) over the full sample.3

The vector of coefficients $\alpha$ has the following distribution:

$$\alpha \sim N(\alpha_0, \Sigma_{\alpha_0}(\lambda_i))$$

(12)

where $\alpha_0$ and $\Sigma_{\alpha_0}$ are such that (5) and (6) are satisfied. In particular, $\alpha_0$ is a vector containing the expectations in (5) while $\Sigma_{\alpha_0}$ is a diagonal matrix and depends on the hyperparameters $\lambda_i$ with $i = 1, \ldots, m$.4 The parameters $\lambda_i$ measure the overall degree of precision of the prior. As $\lambda_i \to 0$, the variances of a specific block in (6) would be multiplied by an increasingly smaller scalar, which

3Using likelihood information to elicit the prior is strictly speaking not in line with a Bayesian approach. However the approach used here for specifying priors on the error variance matrix has been extensively used in many empirical implementations of this model (more often in a variation involving a diagonal variance matrix calibrated using univariate autoregressions), including Doan et al. (1984), Litterman (1986), Kadiyala and Karlsson (1997), Bańbura et al. (2010), Koop (2011), Carriero, Clark and Marcellino (2012).

4Note that the vectorization of the matrices of moments in (5) and (6) is such that e.g. the first row of (5) and (6) are the first $Nq$ elements in $\alpha_0$ and on the diagonal of $\Sigma_{\alpha_0}$, representing the priors on the first equation. The vector $\alpha_0$ and the matrix $\Sigma_{\alpha_0}$ also contain, in the appropriate positions, the prior mean and variances of the intercept. $\Sigma_{\alpha_0}$ is diagonal due to the assumption of prior independence among the coefficients.
means that the expectations in (5) would be imposed on the system with increasingly less noise, and eventually will become a set of sharp restrictions. On the other hand, as \( \lambda_i \to \infty \), the overall variance of the prior will increase, and the prior beliefs expressed in (5) would become very loose and eventually not influence the estimates at all. [Need a justification for the tightness varying with \( m \)]

Under the prior in (12) it is possible to show that the posterior distribution of \( \alpha \) is given by:

\[
\alpha | data \sim N(\bar{\alpha}, \Sigma_\alpha),
\]

with:

\[
\Sigma_\alpha = (\Sigma_{a0}^{-1}(\lambda_i) + \Xi'\Omega_0^{-1}\Xi)^{-1}
\]

\[
\bar{\alpha} = \Sigma_\alpha (\Sigma_{a0}^{-1}(\lambda_i)\alpha_0 + \Xi'\Omega_0^{-1}x).
\]

Note that by recognizing that \( \Xi'\Omega_0^{-1}\Xi = \Sigma_v^{-1} \otimes \mathbf{X}'\mathbf{X} \) and \( \Xi'\Omega_0^{-1}x = \Sigma_v^{-1} \otimes \mathbf{X}'\hat{\alpha} \), where \( \hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'x \) is the OLS estimator, the posterior mean can be written as follows:

\[
\bar{\alpha} = (\Sigma_{a0}^{-1}(\lambda_i) + \Sigma_v^{-1} \otimes \mathbf{X}'\mathbf{X})^{-1} (\Sigma_{a0}^{-1}(\lambda_i)\alpha_0 + \Sigma_v^{-1} \otimes \mathbf{X}'\hat{\alpha}).
\]

Equation (16) shows that the posterior mean is a weighted average of the prior mean (\( \alpha_0 \)) and the mean implied by the likelihood (i.e. the OLS estimator \( \hat{\alpha} \)), each rescaled with weights proportional to the inverse of the respective variances (\( \Sigma_{a0}^{-1}(\lambda_i) \) and \( \Sigma_v^{-1} \otimes \mathbf{X}'\mathbf{X} \)). The parameters \( \lambda_i \) are measuring the overall tightness of the prior, and we discuss below how it can be selected optimally by means of a data driven procedure. As is clear in (16), when \( \lambda_i \to \infty \), the posterior mean \( \bar{\alpha} \) tends to the OLS estimate, \( \bar{\alpha} \to \hat{\alpha} \), while with \( \lambda_i \to 0 \) the posterior mean tends to the prior mean, i.e. \( \bar{\alpha} \to \alpha_0 \).

Another interesting feature of the prior distribution in (12) is that it can be interpreted as a set of uncertain restriction on the system in (11). Defining the variable \( v^+ = \alpha_0 - \alpha \), which under (12) will be distributed as:

\[
v^+ \sim N(0, \Sigma_{a0}(\lambda_i)),
\]

and stacking it at the end of the VAR in (11), we have:

\[
x^* = \Xi^*\alpha + v^* \quad \text{(18)}
\]

\[
x^* = \begin{bmatrix} x \\ \alpha_0 \end{bmatrix}; \Xi^* = \begin{bmatrix} \Xi \\ I \end{bmatrix}; v^* = \begin{bmatrix} v \\ v^+ \end{bmatrix}. \quad \text{(19)}
\]
This new system has an error term \(v^* = (v', v^+)'\) which has a covariance matrix \(\Omega^* = E(v^*v'^*)\) given by:

\[
\Omega^* = \begin{bmatrix} \Omega_0 & 0 \\ 0 & \Sigma_{a_0}(\lambda_i) \end{bmatrix},
\]

and which can be estimated with Generalized Least Squares:

\[
\alpha_{GLS} = (\Xi^*\Omega^*^{-1}\Xi^*)^{-1}\Xi^*\Omega^*^{-1}x^*.
\]

It is possible to show with simple algebra that (21) is equivalent to (15).

### 2.5.1 Random error variance

A shortcoming of the baseline specification is that the variance matrix of the errors is assumed to be fixed to \(\Sigma_v\). Therefore we also consider an alternative prior specification in which the variance matrix of the errors is treated as a random. In particular we label the random variance matrix \(\Sigma^*_v\) and assume that it has a priori an Inverse Wishart distribution with scale and shape parameter calibrated in a way that its expectation is the fixed diagonal matrix \(\Sigma_v\) considered in the baseline specification:

\[
\Sigma^*_v \sim IW(\Sigma_v, v_0),
\]

where we set \(v_0\) to the smallest value guaranteeing that \(E[\Sigma^*_v] = \Sigma_v\), i.e. \(v_0 = qN + 2\). The prior for \(\alpha\) is specified as before, see equation (12). In this case, one can only derive conditional posterior distributions:

\[
\alpha|\Sigma^*_v, data \sim N(\tilde{\alpha}, \Sigma_{\tilde{\alpha}})
\]

\[
\Sigma^*_v|\alpha, data \sim IW(\Sigma^*_v, \tilde{v})
\]

with:

\[
\Sigma_{\tilde{\alpha}} = (\Sigma_{a_{0}}^{-1}(\lambda_i) + \Sigma_v^{-1} \otimes X'X)^{-1}
\]

\[
\tilde{\alpha} = \Sigma_{\tilde{\alpha}} (\Sigma_{a_{0}}^{-1}(\lambda_i) \alpha_0 + \Sigma_v^{-1} \otimes X'X\tilde{\alpha})
\]

\[
\tilde{\Sigma}^*_v = \Sigma_v + (x^+ - X\tilde{\alpha})(x^+ - X\tilde{\alpha})^T
\]

\[
\tilde{v} = v_0 + T
\]

Note that, recalling that \(\Xi'\Omega_0^{-1}\Xi = \Sigma_v^{-1} \otimes X'X\) and \(\Xi'\Omega_0^{-1}x = \Sigma_v^{-1} \otimes X'X\tilde{\alpha}\), (25) and (26) closely resemble (14) and (15), the only difference being that (25) and (26) involve the matrix \(\Sigma^*_v\), which
is a random variable with mean $E[\Sigma_v] = \Sigma_v$, while in (14) and (15) this matrix was fixed to the value $\Sigma_v^* = \Sigma_v$.

The prior defined by (12) and (22), known as the independent Normal-Inverse Wishart prior, allows the error variance matrix to be modelled as random, but has the inconvenience that while the conditional posterior distributions can be derived the joint posterior distribution needs to be simulated via Gibbs sampling.\footnote{To avoid simulation one could use the natural conjugate N-IW prior which features a matrix-variate $t$ joint posterior distribution. However doing so require that each equation is treated symmetrically in term of the prior, i.e., the prior variance of each coefficient in $\mathbf{\Gamma}$ can only differ by a scale factor from one equation to another. See Kadiyala and Karlsson (1997) for details.} In particular draws from the joint distribution $p(\alpha, \Sigma^*_v | Y)$ can be obtained by drawing in turn from the conditionals (24) and (23). The marginal likelihood can only be evaluated numerically: we do so Geweke’s harmonic mean estimator.

2.5.2 Choice of the tightness

To make the prior operational, one needs to choose the value of the hyperparameter $\Lambda$, which controls the overall tightness of the prior. To do so we choose $\Lambda$ by maximizing the marginal likelihood of the model:\footnote{This approach has been used by Carriero, Kapetanios and Marcellino (2012) and Carriero, Clark and Marcellino (2012), and corresponds to choosing the model with the highest posterior odds ratio (under a flat prior over a discrete set of values for the tightness parameters). Giannone, Lenza and Primiceri (2012) propose a similar strategy which assumes a proper (albeit uninformative) prior on a continuum of values for the tightness parameter, and then integrates out the hyperparameters by means of a Metropolis step. However the latter approach is only applicable to a system featuring a closed-form solution for the marginal likelihood.}

$$\{\lambda^*_i\}_{i=1}^m = \arg \max_{\lambda} p(y)$$

where $p(Y)$ can be obtained by integrating the coefficients of the model out of the likelihood function:

$$p(Y) = \int p(Y|\Theta)p(\Theta)d\Theta.$$  \hspace{1cm} (30)

where $\Theta$ is the set of all the coefficients of the model.

In our baseline specification, the integration in (30) can be obtained in closed form and is given by:

$$p(y) = (2\pi)^{-T} |\Omega|^{-1/2} |\Sigma_\alpha|^{1/2} |\Sigma_\alpha^0|^{-1/2} \exp \left(-\frac{Q}{2}\right).$$

where $Q = x^T \Omega^{-1} x + \alpha_0^T \Sigma_\alpha^{-1} \alpha_0 - \bar{\alpha}^T \Sigma_\alpha^{-1} \bar{\alpha}$. In the specification with random variance, the integral (30) can be approximated using Geweke modified harmonic mean estimator (see Geweke and...
Whiteman, 2006) for details). In particular, considering $M$ simulated posterior draws from $\Theta$, $\{\theta_j\}_{j=1}^M$ the estimator is:

$$\hat{p}(Y) = \left[\frac{1}{M} \sum_{j=1}^M \frac{f(\theta_j)}{p(Y|\theta_j)p(\theta_j)}\right]^{-1},$$

where $f(\cdot)$ is a truncated multivariate normal distribution calibrated using the moments of the simulated posterior draws $\{\theta_j\}_{j=1}^M$. As the dimension of our model is large in the number of parameters, the evaluation of $f(\theta_j)$ and $p(\theta_j)$ is numerically unfeasible. In order to have a well-behaved numerical behaviour we focus only on the parameters of the rows 1, 2, $q + 1$, and $q + 2$ of (6). We partition the parameter space into $\Theta = \{\Theta_1, \Theta_2\}$ where $\Theta_1$ contains all the parameters in rows 1, 2, $q + 1$, and $q + 2$ of (6) and we choose $f(\theta_j) = f(\theta_1^j)f(\theta_2^j) = f(\theta_1^j)p(\theta_2^j)$, so that (31) becomes:

$$\hat{p}(Y) = \left[\frac{1}{M} \sum_{j=1}^M \frac{f(\theta_1^j)}{p(Y|\theta_1)p(\theta_1)}\right]^{-1},$$

which can be computed easily without encountering numerical problems. The estimate in (32) might not be stable, however, as it may not satisfy a Gaussian central limit theorem because the random variable $f(\theta_1)/(p(Y|\theta)p(\theta_1))$ can feature infinite variance. We check stability of the estimate by running independent chains and comparing the results across them. In our experiments the estimated marginal likelihoods look stable after about 3000 simulations and are almost identical across multiple simulated chains.

### 2.6 Competitor BVAR benchmark

We include in our forecasting comparison the BVAR with conjugate N-IW prior, as considered in studies such as Kadiyala and Karlsson (1997), Sims and Zha (1998), Bańbura et al. (2010), Carriero, Kapetanios and Marcellino (2009), Carriero, Kapetanios and Marcellino (2012), and Carriero, Clark and Marcellino (2012). This model is estimated with the latest-available vintage at each point in time, that is, disregarding the data revision process, but constitutes a genuine real-time out-of-sample exercise as only data available at each forecast origin are used in model estimation.

The VAR is:

$$y_t = c + \sum_{k=1}^p \Psi_k y_{t-k} + \varepsilon_t,$$

where $y_t$ is a $m \times 1$ vector of macroeconomic variables in log(levels), where all vectors of observations are from the forecast-origin data vintage. For example, for forecasting at $T + 1$, the VAR:

$$y_t^{T+1} = c + \sum_{k=1}^p \Psi_k y_{t-k}^{T+1} + \varepsilon_t,$$
is estimated using \( t = 1, \ldots, T \). In particular, we consider the following prior moments:

\[
E[\Psi_{kk}^{ij}] = \begin{cases} 
\Psi^* & \text{if } i = j, k = 1 \\
0 & \text{otherwise}
\end{cases}, \quad \text{Var}[\Psi_{kk}^{ij}] = \begin{cases} 
\omega_1^2 \sigma_i^2 \frac{1}{\sigma_j^2}, & k = 1, \ldots, p
\end{cases},
\]

(33)

where \( \Psi_{kk}^{ij} \) denotes the element in position \((i, j)\) in the matrix \( \Psi_k \). We estimate this specification in levels, and the prior mean \( \Psi^* \) is set to 1. The shrinkage parameter \( \omega_1 \) measures the overall tightness of the prior, and is chosen via optimization of the marginal likelihood, while \( \kappa \) imposes additional shrinkage on higher order lags. To set each scale parameter \( \sigma_i \) we follow common practice (Litterman, 1986; Sims and Zha, 1998) and set \( \sigma_i \) equal to the standard deviation of the residuals from a univariate autoregressive model. For the intercept we assume an uninformative prior with mean 0. In addition we add the ‘sum of coefficients’ prior, which expresses the belief that when the average of the lagged values of a variable is at some level \( \bar{y}_{0i} \), that same value \( \bar{y}_{0i} \) is likely to be a good forecast of future observations.\(^7\) Finally, we add a N-IW prior on the error variance such that the expectation of the error variance matrix coincides with the traditional Minnesota prior of fixed, diagonal error variance matrix (obtained by OLS estimation of univariate AR models for each variable in the VAR, based on the full sample). Therefore our benchmark model is the same as that of Banbura et al. (2010).

3 Empirical Forecasting Exercise

We suppose that, at each forecast origin \( t + 1 \), the aim is to forecast the post-revision (or fully-revised) values of the future observations \( \{y_{t+h}\} \), where \( h = 1, \ldots, 8 \), that is, \( \{y_{t+1}, \ldots, y_{t+8}\} \). These are future observations in the sense that at \( t + 1 \) the most recent data for which a first estimate is available is for \( y_t \). Forecasts are computed by iteration using the VAR models, where \( \hat{y}_{t+1+\tau}^{t+\tau} \) denotes the \( \tau \)-step ahead forecast of the \( t + 1 + \tau \)-vintage of data. The forecast of the post-revision value of next quarter’s observation \( \hat{y}_{t+1+q}^{t+1+\tau} \) is the last element of the forecast vector \( \hat{y}_{t+1+\tau}^{t+1+\tau} \) with \( \tau = q \). To forecast the fully-revised value of output growth one year ahead, say, we set \( h = 4 \) and \( \tau = q - 1 + h \), and use the \( q^{th} \) element of the vector \( \hat{y}_{t+1+\tau}^{t+1+\tau} \) (in the case of the bivariate VAR, assuming that data on output are the first block).

\(^7\)Note that here the sum of coefficients prior is set by using the average of the variables over the full sample, as in Banbura et al. (2010), not to the average of the first \( p \) observations in the sample, as in other implementations of this prior, e.g. Sims and Zha (1998)
We use quarterly data vintages on US real GDP and the GDP deflator taken from the Philadelphia Fed real-time dataset. At the time of writing, the latest vintage is 2012:Q2. In our empirical analysis we use \( q = 14 \), that is, we model 14 estimates (or maturities) of each observation. The data revision process of national accounts data published by the US statistical agency, the Bureau of Economic Analysis (BEA), is such that data are typically revised one quarter after the first estimate (\( y_{t+1} \)), so that \( y_{t+2} \neq y_{t+1} \), and that the data are then only revised in the third-quarter data releases for each of the next three years. Setting \( q = 14 \) means that we capture the three rounds of annual revisions for each observation irrespective of the quarter of the year to which it belongs. In addition there are periodic benchmark revisions reflecting base year changes amongst other things, although working in growth rates lessens their impact: in-depth analyses are provided by Fixler and Grimm (2005, 2008) \textit{inter alia}.

We consider a 20 year out-of-sample period comprising vintages from 1989:Q2 up to 2009:Q1. We use data from the latest-available vintage, 2012:Q2, as actual values to compute forecast errors, and the end of the out-of-sample period is set such that all the actuals have been revised at least 13 times.

### 3.1 Comparing the BVB-VAR with the Univariate V-VAR

The first question we address is the forecasting accuracy of the BVB-VAR versus the unrestricted VB-VAR and the RV-VAR. We use both univariate and bivariate specifications to forecast quarterly growth rates at annual rates of real GDP and the GDP deflator. The BVB-VAR models in this section were estimated with a fixed \( \Sigma_e \). At each forecast origin during the out-of-sample period, we add one more vintage to the estimation, that is, we use a recursive forecasting scheme. We use root mean squared forecast errors (RMSFE) to measure forecast accuracy for each variable at each horizon in Table 1. The first column (univariate VB-VAR) entries are RMSFEs, and the results in the remaining columns are RMSFE ratios to the univariate VB-VAR benchmark.

The results in Table 1 show that the BVB-VAR models improve forecasts at all horizons in comparison with the benchmark, with RMSFE reductions up to 40% at long horizons. The bivariate BVB-VAR registers small improvements over the univariate for GDP growth, but larger gains of around 5% for inflation at \( h > 4 \). In summary, the Bayesian vintage-based VAR clearly provides more accurate forecasts compared to the univariate vintage-based VAR model, which was by and large the most successful multiple-vintage model evaluated by Clements and Galvão (2012c).
3.2 Comparing the BVB-VAR with other Forecasting Models

The second question we address is the benefit of modelling the data revision process within a forecasting model of output growth and inflation. To address this question, we use as a benchmark an autoregressive model estimated with real-time-vintage data, as suggested by Clements and Galvão (2012d). This model recognises that more recent observations are lightly revised relative to more heavily revised earlier data, and that mixing the two may not be desirable, and also allows for the possibility that data revisions may have a non-zero mean: see Clements and Galvão (2012d) for details.

We also consider two alternative VAR models of output growth and inflation, estimated with the latest-available vintage at each forecast origin, which is the traditional approach to real-time forecasting. The first is a VAR model estimated with growth rates of order 4. The second follows Bańbura et al. (2010) and it is a Bayesian VAR of output and the price level of order 4, estimated by imposing both the Minnesota prior and the sum of coefficients prior. The model is estimated in log-levels, but we compute forecasts of growth rates for comparability to the other models. Bańbura et al. (2010) contend that Bayesian VARs are suitable models for forecasting macroeconomic variables when there are a large number of relevant series.

Table 2 compares the forecasting performance of these models. The out-of-sample period is the same as in Table 1, and all the models are again estimated on increasing windows of data. For output growth the VAR models offer no improvement on the benchmark, and the BVAR(4) and BVB_VAR_bi have a similar forecasting performance. For inflation, the BVB-VAR_bi provides sizeable improvements relative to both the benchmark and the BVAR(4) model when $h > 2$. Neglecting data revisions (as in the VAR(4) and BVAR(4)) leads to notably inferior inflation forecasts at the longer horizons.

Table 3 presents the results of the same forecasting exercise as in table 2, but with a rolling as opposed to recursive forecasting scheme. The use of rolling windows of data has little effect on the forecast comparisons for output growth, either in absolute terms or in relative terms, except that the bivariate BVB-VAR shows some modest improvement over the benchmark at long horizons ($h = 7, 8$). For inflation there are large improvements in the benchmark which are not matched one-for-one by the Bayesian VARs, although the rankings are unchanged and the BVB-VAR_bi provides the best inflation forecasts except at the shortest horizons.
3.3 Sub-sample results

The recent literature suggests that the predictability of models of output growth and inflation may change over time (Stock and Watson, 2003; Stock and Watson, 2007). To investigate this, we split the out-of-sample period into two subsamples, each of 40 observations, with forecast-origin vintages of 1989:Q2-1999:Q1, and of 1999:Q2-2009:Q1. We take as a benchmark for comparison a VAR(4), as in Bańbura et al. (2010). This model is estimated using only data from the end-of-sample vintage (unlike the real-time-vintage autoregression).\footnote{The use of a benchmark that makes no allowance for data revisions more clearly brings out the value of modelling multiple vintages of data. As explained by Koenig et al. (2003) and Clements and Galvão (2012d), the real-time-vintage autoregression is one way of dealing with data revisions.} Figure 1 presents the RMSFE ratios to the VAR(4) benchmark for $h = 1, \ldots, 8$ for both sub-periods and variables.

As expected, the BVAR(4) is more accurate than the VAR(4) for all variables and periods. However, the univariate real-time-vintage model (RTV in the figure) provides even more accurate forecasts of output growth, in particular during the 1999-2009 period. For inflation, there is a clear change in forecasting performance of the BVAR(4) between periods: it beats the VAR(4) in the 1989-1999 period, but is inferior in the most recent one. This is in tune with the pseudo-real-time findings of Stock and Watson (2007) that activity variables are not useful for predicting inflation in recent times.

The results in figure 1 confirm the results of table 2: the BVB-VAR _bi provides markedly more accurate forecasts of inflation for $h > 2$, and this is true across sub-periods. For output growth we find that the BVB-VAR models (univariate and binary) are markedly more effective in the second period than in the first.

4 Conclusions

In this paper we propose modelling and forecasting macroeconomic variables that are subject to revision using a Bayesian vintage-based VAR. The vintage-based VAR models in the literature are typically univariate - they model the relationships between different maturities of data of a single variable. While such approaches show promise in forecasting revisions to data for which initial estimates have been published, they are less successful at forecasting post-revision values of future observations. The use of a Bayesian approach allows us to build multivariate multiple-vintage models. In our empirical work, these models are estimated for output growth and inflation, and
are shown to provide competitive forecasts of the post-revision (or fully-revised) values of these variables.

We show that the nature of data revisions suggests a prior for the Bayesian approach. Specifically, the prior captures the fact that data releases after the first few are likely to be efficient, in the sense of being largely unpredictable. This information is incorporated in a way that \textit{de facto} results in more accurate forecasts than imposing sharp zero restrictions, as in the RV-VAR (compare the RV-VAR and BVB-VAR results in table 1). This enhances the benefits of including other variables without large penalisation from increasing parameter uncertainty.

We find that it is the VB-VAR in conjunction with Bayesian estimation that delivers the sizeable improvements, especially for inflation. We show that the Bayesian approach of Bărbura et al. (2010), which allows large datasets, but does not model multiple vintages, is far less successful for forecasting inflation.
References


<table>
<thead>
<tr>
<th>$y_{t+h}$</th>
<th>VB-VAR_un</th>
<th>VB-VAR_bi</th>
<th>RV-VAR_un</th>
<th>RV-VAR_bi</th>
<th>BVB-VAR_un</th>
<th>BVB-VAR_bi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecasting Output Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.523</td>
<td>0.99</td>
<td>0.99</td>
<td>1.03</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>2.646</td>
<td>1.01</td>
<td>1.00</td>
<td>1.08</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>3</td>
<td>2.590</td>
<td>1.04</td>
<td>0.99</td>
<td>1.09</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>2.666</td>
<td>1.06</td>
<td>0.98</td>
<td>1.10</td>
<td>0.82</td>
<td>0.79</td>
</tr>
<tr>
<td>5</td>
<td>2.672</td>
<td>1.07</td>
<td>0.98</td>
<td>1.12</td>
<td>0.81</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>2.730</td>
<td>1.05</td>
<td>0.99</td>
<td>1.11</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>7</td>
<td>2.907</td>
<td>1.05</td>
<td>1.02</td>
<td>1.12</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>3.045</td>
<td>1.04</td>
<td>1.02</td>
<td>1.09</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Forecasting Inflation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.014</td>
<td>1.11</td>
<td>1.09</td>
<td>1.22</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>1.054</td>
<td>1.16</td>
<td>1.09</td>
<td>1.28</td>
<td>0.75</td>
<td>0.74</td>
</tr>
<tr>
<td>3</td>
<td>1.052</td>
<td>1.22</td>
<td>1.09</td>
<td>1.31</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>1.044</td>
<td>1.20</td>
<td>1.13</td>
<td>1.28</td>
<td>0.78</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>0.952</td>
<td>1.24</td>
<td>1.15</td>
<td>1.30</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>6</td>
<td>0.995</td>
<td>1.19</td>
<td>1.15</td>
<td>1.28</td>
<td>0.58</td>
<td>0.53</td>
</tr>
<tr>
<td>7</td>
<td>1.225</td>
<td>0.99</td>
<td>1.10</td>
<td>1.07</td>
<td>0.69</td>
<td>0.63</td>
</tr>
<tr>
<td>8</td>
<td>1.323</td>
<td>0.87</td>
<td>1.11</td>
<td>0.95</td>
<td>0.68</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Notes: The table reports RMSFEs for the VB-VAR_un, subsequent columns are RMSFE ratios to that of the VB-VAR_un. The models are estimated using expanding windows of data.
Table 2: Bayesian VB-VAR versus other forecasting models. Forecasting vintages 1989:Q2-2009:Q1.

<table>
<thead>
<tr>
<th>$y_{t+h}$</th>
<th>AR_RTIV</th>
<th>VAR(4)</th>
<th>BVAR(4)</th>
<th>BVB-VAR_un</th>
<th>BVB-VAR_bi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Forecasting Output Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.046</td>
<td>1.03</td>
<td>1.01</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>2</td>
<td>2.127</td>
<td>1.06</td>
<td>1.01</td>
<td>1.05</td>
<td>1.03</td>
</tr>
<tr>
<td>3</td>
<td>1.973</td>
<td>1.11</td>
<td>1.04</td>
<td>1.05</td>
<td>1.02</td>
</tr>
<tr>
<td>4</td>
<td>2.050</td>
<td>1.11</td>
<td>1.03</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>5</td>
<td>2.049</td>
<td>1.12</td>
<td>1.03</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>6</td>
<td>2.179</td>
<td>1.10</td>
<td>1.01</td>
<td>1.05</td>
<td>1.02</td>
</tr>
<tr>
<td>7</td>
<td>2.578</td>
<td>1.06</td>
<td>1.01</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>8</td>
<td>2.800</td>
<td>1.06</td>
<td>1.02</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forecasting Inflation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.881</td>
<td>1.10</td>
<td>1.02</td>
<td>1.10</td>
<td>1.09</td>
</tr>
<tr>
<td>2</td>
<td>0.831</td>
<td>1.07</td>
<td>1.01</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>3</td>
<td>0.869</td>
<td>1.02</td>
<td>0.96</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>4</td>
<td>0.923</td>
<td>0.97</td>
<td>0.92</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>5</td>
<td>0.911</td>
<td>0.91</td>
<td>0.85</td>
<td>0.78</td>
<td>0.73</td>
</tr>
<tr>
<td>6</td>
<td>0.953</td>
<td>0.81</td>
<td>0.80</td>
<td>0.61</td>
<td>0.55</td>
</tr>
<tr>
<td>7</td>
<td>1.184</td>
<td>0.83</td>
<td>0.82</td>
<td>0.71</td>
<td>0.65</td>
</tr>
<tr>
<td>8</td>
<td>1.291</td>
<td>0.86</td>
<td>0.81</td>
<td>0.70</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Notes: As for table 1. In addition, the BVAR(4) is estimated in the levels of the variables (all other models are estimated in growth rates) with the ‘sum of coefficients’ prior.

<table>
<thead>
<tr>
<th></th>
<th>(y_{t+h})</th>
<th>AR_RTV</th>
<th>VAR(4)</th>
<th>BVAR(4)</th>
<th>BVB-VAR_un</th>
<th>BVB-VAR_bi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forecasting Output Growth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.030</td>
<td>1.05</td>
<td>1.02</td>
<td>1.05</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.105</td>
<td>1.06</td>
<td>1.01</td>
<td>1.06</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.972</td>
<td>1.06</td>
<td>1.00</td>
<td>1.05</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.044</td>
<td>1.07</td>
<td>1.00</td>
<td>1.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.034</td>
<td>1.07</td>
<td>1.01</td>
<td>1.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.167</td>
<td>1.06</td>
<td>1.00</td>
<td>1.03</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.573</td>
<td>1.04</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.792</td>
<td>1.04</td>
<td>1.01</td>
<td>0.99</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td><strong>Forecasting Inflation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.836</td>
<td>1.11</td>
<td>1.01</td>
<td>1.16</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.773</td>
<td>1.09</td>
<td>1.02</td>
<td>1.05</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.809</td>
<td>1.05</td>
<td>0.98</td>
<td>0.98</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.838</td>
<td>0.99</td>
<td>0.92</td>
<td>0.99</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.782</td>
<td>0.94</td>
<td>0.90</td>
<td>0.95</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.796</td>
<td>0.88</td>
<td>0.87</td>
<td>0.83</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.949</td>
<td>0.88</td>
<td>0.89</td>
<td>0.95</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.032</td>
<td>0.89</td>
<td>0.89</td>
<td>0.92</td>
<td>0.80</td>
<td></td>
</tr>
</tbody>
</table>

Notes: As for table 2, except that estimation uses rolling windows of data.
Figure 1: Forecasting Performance against the VAR(4) (RMSFEs Ratios) for h=1,...,8, split-sample results.