ELSEVIER

Forecasting with Bayesian multivariate vintage-based VARs

Andrea Carriero [a], Michael P. Clements [b,c], Ana Beatriz Galvão [d,*]

[a] *Queen Mary University of London, UK*
[b] *ICMA Centre, University of Reading, UK*
[c] *Institute for New Economic Thinking, Oxford Martin School, University of Oxford, UK*
[d] *University of Warwick, UK*

ARTICLE INFO

ABSTRACT

We consider the forecasting of macroeconomic variables that are subject to revisions, using Bayesian vintage-based vector autoregressions. The prior incorporates the belief that, after the first few data releases, subsequent ones are likely to consist of revisions that are largely unpredictable. The Bayesian approach allows the joint modelling of the data revisions of more than one variable, while keeping the concomitant increase in parameter estimation uncertainty manageable. Our model provides markedly more accurate forecasts of post-revision values of inflation than do other models in the literature.

© 2014 The Authors. Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

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] The fact that data are subject to revisions raises issues in relation to the appropriate criterion to use for assessing forecast accuracy. A researcher may choose subsequently 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. Arguably, a question which is more important than the issue

of forecast evaluation is the impact of data revisions on model specification and estimation. At any forecast origin, the more recent observations are only lightly revised relative to the earlier, more heavily-revised data, and Clements and Galvão (2013b) and Koenig, Dolmas, and Piger (2003), among others, have shown that the traditional approach to real-time forecasting ignores this aspect, to its detriment, and have suggested alternative approaches.

The recent academic literature has proposed considering the multiple vintages of data that the forecaster has access to at any point in time explicitly, in order to model the data revisions process, see, e.g., Cunningham, Eklund, Jeffery, Kapetanios, and Labhard (2012), Garratt, Lee, Mise, and Shields (2008, 2009), Kishor and Koenig (2012) and Patterson (2003). Some of the leading multiple-vintage models have been evaluated by Clements and Galvão (2012b). For our purposes, they present two key findings. Firstly, vintage-based VARs (henceforth VB-VARs, described fully below) provide competitive forecasts of the post-revision values of data for which the forecaster has already observed an earlier release; that is, they are able to forecast revisions to *past* data better than natural comparators (such as assuming that revisions to already-released

---

data will be zero). Secondly, vintage-based VARs are not able to forecast post-revision values of *future* observations any more accurately than simple autoregressive models.

Two potential reasons for this failure to improve the forecasts of future observations are addressed in this paper. The first is that the vintage-based VARs typically used in the literature (Clements & Galvão, 2012b, 2013a; Garratt et al., 2008) are univariate: they only model the revisions to a single variable. This may be overly restrictive. For example, if we are interested in modelling inflation, the Phillips Curve suggests a role for activity variables (Stock & Watson, 2007, 2009, 2010). Restricting the analysis to a single variable avoids the issue of parameter proliferation from modelling two or more variables and their data revisions jointly, but may be costly in terms of forecasting future observations. The second reason is that, even if we restrict the number of variables to one, as is the case with the univariate multiple-vintage models that are typical in the literature, the effects of parameter-estimation uncertainty are likely to be large, and may detract seriously from the forecasting performance. To capture the nature of the Bureau of Economic Analysis revisions process, Clements and Galvão (2012b) model a vector of 14 elements, so that, even for a single-variable first-order VAR, there are 14^2 slope parameters to estimate.

In this paper, we employ a Bayesian approach which offers a solution to both of these problems, thus permitting multiple-variable multiple-vintage models. On the one hand, the coefficients are shrunk towards prior values, reducing the 'curse of dimensionality' issue that afflicts classical VARs when the number of variables increases beyond a handful. In this way, it is possible to summarize the information contained in a large number of successive releases of data revisions efficiently. On the other hand, the nature of revisions lends itself to the specification of an economically meaningful prior, which is tailored so as 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.

We design a real-time forecasting exercise to address two issues. 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 so as to allow one to disentangle the gains from the use of an extra explanatory variable in addition to those resulting from the adoption of a Bayesian approach. Second, we assess the forecasting improvements from modelling the data revision process when forecasting in real time. We compare the forecasting performance of BVB-VAR with that of a benchmark Bayesian VAR (as per Bańbura, Giannone, & Reichlin, 2010, Carriero, Clark, & Marcellino, in press, Kadiyala & Karlsson, 1997, and Sims & Zha, 1998). 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 account the nature of real-time data, following Clements and Galvão (2013b) and Koenig et al. (2003).

The plan of the rest of the paper is as follows. Section 2 describes the vintage-based VAR forecasting models. Section 3 describes the Bayesian approach we adopt in this paper, including the specification of priors and the treatment of the error-variance matrix. Next, Section 4 describes the results of the real-time empirical forecasting exercises for US output growth and inflation. Finally, Section 5 offers some concluding remarks.

2. The forecasting model

2.1. Univariate vintage-based VAR

This section describes the univariate VB-VAR of Clements and Galvão (2012b).[2] The approaches described in this section do not aim to measure the true unobserved values of a time series subject to revision, as per Jacobs and van Norden (2011) and Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2013), nor to identify news and noise revisions as per Jacobs, Sarferaz, Sturm, and van Norden (2013), but to use multiple vintages of data to improve the forecasts of revised values.

The models are based on the quarterly data vintages made available in the Real-Time Dataset for Macroeconomists (Stark & Croushore, 2002). We write the first release of the value of the variable Y in period t (the level of GDP, or the GDP deflator) as Y_t^{t+1}. The superscript on Y denotes the data vintage, and the subscript is the time period to which the observation refers. Thus, Y_t^{t+1} is the estimate of Y_t available around the 15th day of the middle month of the following quarter (quarter $t + 1$). We work in terms of growth rates defined by log-differences, so that $y_t^{t+1} = \log(Y_t^{t+1}/Y_{t-1}^{t+1})$, where we use natural logarithms. Then, $y_t^{t+2} = \log(Y_t^{t+2}/Y_{t-1}^{t+2})$ is the estimate of the quarterly growth rate (or inflation) available around the 15th day of the middle month of quarter $t + 2$, and so on. Suppose that there are revisions for the next $q - 1$ quarters after the first release, but that the observation is unrevised thereafter (i.e., $y_t^{t+q+i} = y_t^{t+q}$ for $i > 0$). Then, we can model the vintage $t + 1$ values of observations $t - q + 1$ to t as a VB-VAR:

$$\mathbf{y}^{t+1} = \mathbf{c} + \sum_{i=1}^{p} \mathbf{\Gamma}_i \mathbf{y}^{t+1-i} + \boldsymbol{\varepsilon}^{t+1}, \qquad (1)$$

where $\mathbf{y}^{t+1-i} = \left[y_{t-i}^{t+1-i}, y_{t-1-i}^{t+1-i}, \ldots, y_{t-q+1-i}^{t+1-i} \right]'$, $i = 0, 1, \ldots, p$, $\mathbf{c}$ is $q \times 1$, and $\boldsymbol{\varepsilon}^{t+1}$ is $q \times 1$. The variance–covariance matrix of the disturbances ($\mathbf{\Sigma}_\varepsilon = E(\boldsymbol{\varepsilon}^{t+1} \boldsymbol{\varepsilon}^{t+1\prime})$) captures the correlations between data of the same vintage.

The VB-VAR models the $t + 1$-vintage values of growth rates at period $t, t - 1, \ldots$ etc. Some authors have suggested modelling the vector of different vintage values for a given period. This is described by Hecq and Jacobs (2009) and labelled the observation-balanced (OB-)VAR. Suppose that we model a vector of three elements; then, for the OB-VAR, the vector would be $\left[y_{t-2}^{t+1}, y_{t-2}^t, y_{t-2}^{t-1} \right]$. From a forecasting perspective, the VB-VAR has a clear advantage. Suppose that the latest data we have are of the vintage $t + 1$. The

---

[2] As was explained by Clements and Galvão (2012b), this model is closely related to those of Garratt et al. (2008) and Hecq and Jacobs (2009).

VB-VAR makes use of data on periods t and $t-1$, whereas the most recent data used in the OB-VAR are for period $t-2$, and this is exacerbated when q is larger than three, as is the case in our empirical applications.

A closely related model is suggested by Garratt et al. (2008), following Patterson (1995). Their model is in terms of the level (of the log) of output (Y), which is taken to be integrated of order one ($I(1)$), and they assume that different vintage estimates are cointegrated such that data revisions are $I(0)$. The model is:

$$\mathbf{Z}^{t+1} = \mathbf{c} + \boldsymbol{\Phi}_1\mathbf{Z}^t + \boldsymbol{\Phi}_2\mathbf{Z}^{t-1} + \boldsymbol{\varepsilon}_t, \qquad (2)$$

where $\mathbf{Z}^{t+1-i} = (Y_{t-i}^{t+1-i} - Y_{t-1-i}^{t-i}, Y_{t-1-i}^{t+1-i} - Y_{t-1-i}^{t-i}, Y_{t-2-i}^{t+1-i} - Y_{t-2-i}^{t-i})'$ for $i = 0, 1, 2$, and where the $\boldsymbol{\Phi}_i$ are 3×3 matrices of coefficients, where the third columns consist solely of zeros. The first element of $\mathbf{Z}^{t+1}$ is a difference across vintages and observations, and subsequent terms are revisions to past data. The inclusion of two revisions reflects the view that a 'revision horizon' of two is appropriate: the vector $\mathbf{Z}^{t+1}$ can be extended to include additional revisions as necessary. Clements and Galvão (2012b) compare the VB-VAR with models such as that in Eq. (2). Based on their findings, our forecasting exercises are based on the VB-VAR.

Notice that even when $p = 1$ in Eq. (1), as per Clements and Galvão (2012b), there are still a large number of parameters to estimate when $q = 14$. A restricted VB-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^{t+n+1} is an efficient forecast, in the sense that the revision from y_t^{t+n} to y_t^{t+n+1} is unpredictable, i.e., $E\left[\left(y_t^{t+n+1} - y_t^{t+n}\right) \mid \mathbf{y}^{t+n}\right] = 0$, whereas $E\left[\left(y_t^{t+i+1} - y_t^{t+i}\right) \mid \mathbf{y}^{t+i}\right] \neq 0$ for $i < n$. We can impose this restriction on the VAR with $p = 1$, where it translates to $E\left(y_{t-n}^{t+1} \mid y_{t-n}^t\right) = y_{t-n}^t$. This is achieved by specifying $\boldsymbol{\Gamma}_1$ in Eq. (1) as:

$$\widetilde{\boldsymbol{\Gamma}}_1 = \begin{bmatrix} & \boldsymbol{\gamma}_{n\times q} & \\ \mathbf{0}_{(q-n)\times(n-1)} & \mathbf{I}_{(q-n)\times(q-n)} & \mathbf{0}_{(q-n)\times 1} \end{bmatrix}. \qquad (3)$$

In our empirical work, we set $n = 2$, so that values after the first revision are assumed to be efficient forecasts (i.e., the 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 restricted vintage-based VAR, RV-VAR.

2.2. Multivariate vintage-based VAR

In Eq. (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 that we now have m variables, and define:

$$\mathbf{x}^{t+1} = [\mathbf{y}_1^{t+1\prime}, \ldots, \mathbf{y}_m^{t+1\prime}]', \qquad (4)$$

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

$$\mathbf{x}^{t+1} = \mathbf{c} + \boldsymbol{\Psi}\mathbf{x}^t + \mathbf{v}^{t+1}, \qquad (5)$$

where $\mathbf{c}$ and $\boldsymbol{\Psi}$ are $mq \times 1$ and $mq \times mq$ respectively, and $\mathbf{v}^{t+1} = [\boldsymbol{\varepsilon}_1^{t+1\prime}, \ldots, \boldsymbol{\varepsilon}_m^{t+1\prime}]'$ is a vector of length mq. The

disturbances have variance $\boldsymbol{\Sigma}_v = E[\mathbf{v}^{t+1}\mathbf{v}^{t+1\prime}]$, which is generally 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.

3. Bayesian vintage-based VARs (BVB-VARs)

When we consider multiple-vintage, multiple-variable models, restrictions such as those in Eq. (3) reduce the overparameterization problem that would otherwise arise in such high-dimensional models. The restrictions given in Eq. (3) are economically meaningful, and are tailored to accommodate the joint modelling of the data revision process and the process for the revised values. Nonetheless, such restrictions might be too stringent, and can introduce mis-specification into the model.

One way to reduce the risk of incurring such mis-specifications is to impose the restrictions in Eq. (3) in an approximate manner, that is, subject to some random noise around them. To do this, we propose to use a Bayesian vintage-based VAR. The main difference between this and the approach of Eq. (5) is that the coefficient matrix $\boldsymbol{\Psi}$ is modelled as a random variable, for which a prior distribution consistent with Eq. (3) can be specified.

Our baseline specification is closely related to the Minnesota prior of Doan, Litterman, and Sims (1984) and Litterman (1986). The matrix $\boldsymbol{\Psi}$ has a normal prior distribution, while the error variance matrix $\boldsymbol{\Sigma}_v$ is assumed to be non-random, and is estimated in a preliminary step. Treating the matrix $\boldsymbol{\Sigma}_v$ 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 in order to select the degree of precision on a given prior optimally).[3] A BVB-VAR with a fixed $\boldsymbol{\Sigma}_v$ also has a straightforward interpretation in terms of sampling theory, because it is possible to show that the Bayesian estimation of this model coincides with the mixed estimator of Theil and Goldberger (1961), which is in turn equivalent to the Generalized Least Squares estimation of a VAR estimated under a set of uncertain restrictions.

For these reasons, treating $\boldsymbol{\Sigma}_v$ as fixed may be an attractive option. Nevertheless, we also estimate a version of the model with $\boldsymbol{\Sigma}_v$ treated as random, in order to assess the potential benefits of allowing a non-random $\boldsymbol{\Sigma}_v$.

3.1. The prior

In this subsection, we describe the prior we use for the Bayesian vintage-based VAR. The prior is centered on the economically meaningful restrictions given by Eq. (3), while allowing for noise around them.

The element $\boldsymbol{\Psi}^{(ij)}$ in row i and column j of the matrix of coefficients $\boldsymbol{\Psi}$ appearing in Eq. (5) is assumed to be independent of any other element $\boldsymbol{\Psi}^{(ij)}$ and to have a

---

[3] Closed form solutions for the joint posteriors are also available in the case of the random error variance, but only under the natural-conjugate Normal Inverse-Wishart prior. However, such a prior requires a Kronecker structure for the prior variance of the coefficients, and it cannot be implemented 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.

Gaussian distribution with prior expectation $E[\Psi^{(ij)}]$ and prior variance $Var[\Psi^{(ij)}]$. A diffuse prior with mean 0 and variance $\phi_0 = 1$ is set for the intercepts in **c**.

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 empirical applications, we have $q = 14$ and at most two variables (output growth and inflation), so that $m = 2$. For simplification, we use these dimensions to present our prior matrices, but our approach can be extended easily to generic values of m and q. The prior expectations can be collected in the following matrix:

$$E[\Psi] = \begin{bmatrix} \hat{g}_{11} & \hat{g}_{12} & \cdots & \cdots & \hat{g}_{1q} & \hat{f}_{11} & \hat{f}_{12} & \cdots & \cdots & \hat{f}_{1q} \\ 1 & \hat{g}_{22} & \hat{g}_{23} & \cdots & \hat{g}_{2q} & \hat{f}_{21} & \hat{f}_{22} & \cdots & \cdots & \hat{f}_{2q} \\ 0 & 1 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & 0 & 0 & \cdots & \ddots & \cdots \\ \hat{h}_{11} & \hat{h}_{12} & \cdots & \cdots & \hat{h}_{1q} & \hat{l}_{11} & \hat{l}_{12} & \cdots & \cdots & \hat{l}_{1q} \\ \hat{h}_{21} & \hat{h}_{22} & \cdots & \cdots & \hat{h}_{2q} & 1 & \hat{l}_{22} & \hat{l}_{23} & \cdots & \hat{l}_{2q} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}. \quad (6)$$

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

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

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

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

$$y_{2,t-1}^{t+1} - y_{2,t-1}^t = c_4 + l_{22}^0 y_{2,t-2}^t + error. \quad (10)$$

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 prior means not estimated by these regressions are set to 0. Note that Eq. (7) applies to an equation for output growth, and Eq. (8) to an equation for inflation, with the differences between the two reflecting the belief that output growth can be modelled by an AR(1) whereas inflation requires a higher order autoregression. On the coefficients with priors of zero, we set variances equal to ϕ_1/i^2 and ϕ_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 that of inflation.

The prior variances on the cross-dynamic coefficients $\hat{f}_{11}$ and $\hat{h}_{11}$ are set to twice the estimated standard deviation on the pre-sample period. On the remaining cross coefficients $\hat{h}_{ij}$ and $\hat{f}_{ij}$, we set a prior mean of 0, with prior variances equal to $\kappa\phi_1/i^2$ and $\kappa\phi_2/i^2$, where $i = 2, \ldots, q-1$, and $\kappa = 0.5$. Again, this is in accordance with the Minnesota prior: the variance of the coefficients on own lags

decreases with the lag by quadratic decay, and on cross lags there is additional shrinkage measured by the parameter κ.

As Eqs. (6), (9) and (10) indicate, the equations for the second estimates restrict $\hat{g}_{21} = \hat{l}_{21} = 1$, and use pre-sample estimates to obtain prior values $\hat{g}_{22}$ and $\hat{l}_{22}$. This is compatible with the support in the literature for the predictability of initial revisions (Clements & Galvão, 2012a). Subsequent elements are set to zero, $\hat{g}_{2j} = \hat{l}_{2j} = 0$ for $j = 3, \ldots, q$. These zero restrictions, and the unit restrictions appearing in all of the remaining equations, are given a very informative prior with a variance of $\phi_3 = 0.001^2$, and of $2\phi_3$ for the cross terms.

Finally, the hyperparameters λ_i, $i = 1, 2$, rescale the overall variances for the equations for each of the two variables which include intercepts. As we shall discuss below, they will be chosen optimally by maximizing the marginal data density. In this respect, our specification is more flexible than that of the traditional Minnesota prior, in which the same overall tightness is given to all variables in the system. Instead, we have two separate hyperparameters, each of which corresponds to a different block of variables (those related to output and those related to inflation). This increase in flexibility allows the use of an appropriate level of shrinkage for each variable of interest in the VAR.

Eq. (11) in Box I summarises how we set the prior variance $Var[\Psi]$.

3.2. Posterior distribution: baseline specification

In this subsection, we consider the baseline specification of our model, characterized by the prior moments in Eqs. (6) and (11) and by a fixed error variance matrix Σ_v. Consider all vintages $t = 1, \ldots, T$, and rewrite the system in Eq. (5) as:

$$\mathbf{x}^+ = \mathbf{XA} + \mathbf{v}^+, \quad (12)$$

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

$$\mathbf{x} = \Xi\alpha + \mathbf{v}, \quad (13)$$

with $x = vec(\mathbf{x}^+)$, $\Xi = \mathbf{I} \otimes \mathbf{X}$, $\alpha = vec(\mathbf{A})$, and $\mathbf{v} = vec(\mathbf{v}^+)$. Given a sample of size T, $\mathbf{x}$ and $\mathbf{v}$ are $qmT \times 1$ vectors, and Ξ is the $qmT \times (qm + 1)$ matrix of regressors. The vector $\mathbf{v}$ of disturbances of the vectorized model has a multivariate normal distribution with variance $\Omega_0 = \Sigma_v \otimes \mathbf{I}_T$, where Σ_v is the variance–covariance matrix of the disturbances of Eq. (5). Note that the variance matrix Ω_0, conditional on the knowledge of Σ_v, is not a random variable. The value Ω_0 is derived by setting Σ_v to the OLS estimates obtained from VAR(1) over the full sample.[4]

---

[4] Strictly speaking, using likelihood information to elicit the prior is not in line with a Bayesian approach. However, the approach used here for specifying priors on the error variance matrix has been used extensively in many empirical implementations of this model (more often in a variation involving a diagonal variance matrix calibrated using univariate autoregressions), including those of Bańbura et al. (2010), Doan et al. (1984), Kadiyala and Karlsson (1997), Koop (2013) and Litterman (1986).

$$
\begin{bmatrix}
\lambda_1(2sd(\hat{g}_{11}))^2 & \lambda_1\frac{\phi_1}{1^2} & \cdots & \lambda_1\frac{\phi_1}{(q-1)^2} & \lambda_1(2sd(\hat{f}_{11}))^2 & \lambda_1\kappa\frac{\phi_1}{1^2} & \cdots & \cdots & \lambda_1\kappa\frac{\phi_1}{(q-1)^2} \\
\lambda_1\phi_3 & \lambda_1(2sd(\hat{g}_{22}))^2 & \lambda_1\frac{\phi_1}{1^2} & \cdots & \frac{\phi_1}{(q-2)^2} & \lambda_1 2\phi_3 & \lambda_1\kappa\frac{\phi_1}{1^2} & \cdots & \lambda_1\kappa\frac{\phi_1}{(q-1)^2} \\
\lambda_1\phi_3 & \lambda_1\phi_3 & \cdots & \lambda_1\phi_3 & \lambda_1 2\phi_3 & \lambda_1 2\phi_3 & \cdots & \cdots & \lambda_1 2\phi_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\
\lambda_1\phi_3 & \lambda_1\phi_3 & \cdots & \lambda_1\phi_3 & \lambda_1 2\phi_3 & \lambda_1 2\phi_3 & \cdots & \cdots & \lambda_1 2\phi_3 \\
\lambda_2(2sd(\hat{h}_{11}))^2 & \lambda_2\kappa\frac{\phi_2}{1^2} & \cdots & \lambda_2\kappa\frac{\phi_2}{(q-1)^2} & \lambda_2(2sd(\hat{l}_{11}))^2 & \cdots & \lambda_2(2sd(\hat{l}_{14}))^2 & \cdots & \lambda_2\frac{\phi_2}{(q-4)^2} \\
\lambda_2 2\phi_3 & \lambda_2\kappa\frac{\phi_2}{1^2} & \cdots & \lambda_2\kappa\frac{\phi_2}{(q-1)^2} & \lambda_2\phi_3 & \lambda_2(2sd(\hat{l}_{22}))^2 & \lambda_2\frac{\phi_2}{1^2} & \cdots & \lambda_2\frac{\phi_2}{(q-2)^2} \\
\lambda_2 2\phi_3 & \lambda_2 2\phi_3 & \cdots & \lambda_2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \cdots & \cdots & \lambda_2\phi_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\
\lambda_2 2\phi_3 & \lambda_2 2\phi_3 & \cdots & \lambda_2 2\phi_3 & \lambda_2\phi_3 & \lambda_2\phi_3 & \cdots & \cdots & \lambda_2\phi_3
\end{bmatrix}. \quad (11)
$$

Box I.

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

$$\boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\alpha_0}(\lambda_i)), \quad (14)$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\Sigma}_{\alpha_0}$ are such that Eqs. (6) and (11) are satisfied. In particular, $\boldsymbol{\alpha}_0$ is a vector containing the expectations in Eq. (6), while $\boldsymbol{\Sigma}_{\alpha_0}$ is a diagonal matrix and depends on the hyperparameters λ_i with $i = 1, \ldots, m$.[5] The parameters λ_i measure the overall degree of precision of the prior. As $\lambda_i \to 0$, the variances of a specific block in Eq. (11) would be multiplied by an increasingly small scalar, meaning that the expectations in Eq. (6) would be imposed on the system with less and less noise, and would eventually 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 Eq. (6) will become very loose, and eventually not influence the estimates at all.

In the following discussion, we denote the complete set of data on which posterior inference is based by $\mathbf{y}$. Under the prior in Eq. (14), it is possible to show (Geweke, 2005) that the posterior distribution of $\boldsymbol{\alpha}$ is given by:

$$\boldsymbol{\alpha}|\mathbf{y} \sim N(\bar{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\bar{\alpha}}), \quad (15)$$

with:

$$\boldsymbol{\Sigma}_{\bar{\alpha}} = \left(\boldsymbol{\Sigma}_{\alpha_0}^{-1}(\lambda_i) + \boldsymbol{\Xi}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Xi}\right)^{-1} \quad (16)$$

$$\bar{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\bar{\alpha}}\left(\boldsymbol{\Sigma}_{\alpha_0}^{-1}(\lambda_i)\boldsymbol{\alpha}_0 + \boldsymbol{\Xi}'\boldsymbol{\Omega}_0^{-1}\mathbf{x}\right). \quad (17)$$

Note that, by recognizing that $\boldsymbol{\Xi}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Xi} = \boldsymbol{\Sigma}_v^{-1} \otimes \mathbf{X}'\mathbf{X}$ and $\boldsymbol{\Xi}'\boldsymbol{\Omega}_0^{-1}\mathbf{x} = \boldsymbol{\Sigma}_v \otimes \mathbf{X}'\mathbf{x} = \boldsymbol{\Sigma}_v^{-1} \otimes \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}$ is the OLS estimator, the posterior mean can be written as follows:

$$\bar{\boldsymbol{\alpha}} = \left(\boldsymbol{\Sigma}_{\alpha_0}^{-1}(\lambda_i) + \boldsymbol{\Sigma}_v^{-1} \otimes \mathbf{X}'\mathbf{X}\right)^{-1}$$
$$\times \left(\boldsymbol{\Sigma}_{\alpha_0}^{-1}(\lambda_i)\boldsymbol{\alpha}_0 + \boldsymbol{\Sigma}_v^{-1} \otimes \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}\right). \quad (18)$$

Eq. (18) shows that the posterior mean is a weighted average of the prior mean ($\boldsymbol{\alpha}_0$) and the mean implied by the likelihood (i.e., the OLS estimator $\hat{\boldsymbol{\alpha}}$), each rescaled with weights that are proportional to the inverse of the respective variances ($\boldsymbol{\Sigma}_{\alpha_0}^{-1}(\lambda_i)$ and $\boldsymbol{\Sigma}_v^{-1} \otimes \mathbf{X}'\mathbf{X}$). The parameters λ_i measure the overall tightness of the prior, and we discuss below how they can be selected optimally by means of a data-driven procedure. As is clear from Eq. (18), when $\lambda_i \to \infty$, the posterior mean $\bar{\boldsymbol{\alpha}}$ tends to the OLS estimate, $\bar{\boldsymbol{\alpha}} \to \hat{\boldsymbol{\alpha}}$, while with $\lambda_i \to 0$, the posterior mean tends to the prior mean, i.e., $\bar{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}_0$.

Another interesting feature of the prior distribution in Eq. (14) is that it can be interpreted as a set of uncertain restrictions on the system in Eq. (13). If we define the variable $\mathbf{v}^+$ as $\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}$, which under Eq. (14) will be distributed as:

$$\mathbf{v}^+ \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\alpha_0}(\lambda_i)), \quad (19)$$

and stack it at the end of the VAR in Eq. (13), we have:

$$\mathbf{x}^* = \boldsymbol{\Xi}^*\boldsymbol{\alpha} + \mathbf{v}^* \quad (20)$$

$$\mathbf{x}^* = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\alpha}_0 \end{bmatrix}, \qquad \boldsymbol{\Xi}^* = \begin{bmatrix} \boldsymbol{\Xi} \\ \mathbf{I} \end{bmatrix}, \qquad v^* = \begin{bmatrix} \mathbf{v} \\ \mathbf{v}^+ \end{bmatrix}. \quad (21)$$

This new system has an error term $v^* = (v', v^{+\prime})'$, which has a covariance matrix $\Omega^* = E\left(v^*v^{*\prime}\right)$ given by:

$$\boldsymbol{\Omega}^* = \begin{bmatrix} \boldsymbol{\Omega}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\alpha_0}(\lambda_i) \end{bmatrix}. \quad (22)$$

Therefore, the coefficient vector α can be estimated by Generalized Least Squares:

$$\boldsymbol{\alpha}_{GLS} = \left(\boldsymbol{\Xi}^*\boldsymbol{\Omega}^{*-1}\boldsymbol{\Xi}^*\right)^{-1}\boldsymbol{\Xi}^*\boldsymbol{\Omega}^{*-1}\mathbf{x}^*. \quad (23)$$

It is possible to show algebraically that Eq. (23) is equivalent to Eq. (17).

3.3. Choice of the tightness

Before employing the prior described in Section 3.1, we need to set the values of the hyperparameters λ_i. To do so, we select the hyperparameters to maximize the marginal

---

[5] Note that the vectorization of the matrices of moments in Eqs. (6) and (11) is such that, for example, the first row of Eqs. (6) and (11) consists of the first Nq elements in α_0 and on the diagonal of Σ_{α_0}, representing the priors on the first equation. The vector α_0 and the matrix Σ_{α_0} also contain, in the appropriate positions, the prior mean and variances of the intercept. Σ_{α_0} is diagonal due to the assumption of prior independence among the coefficients.

data density of the model:[6]

$$\{\lambda_i^*\}_{i=1}^m = \arg\max_{\lambda_i} p(\mathbf{y}|\lambda_i). \tag{24}$$

Let $\mathbf{\Theta}$ denote all of the coefficients of the model; then, the marginal data density can be obtained by integrating $\mathbf{\Theta}$ out of the likelihood function:

$$p(\mathbf{y}|\lambda_i) = \int p(\mathbf{y}|\mathbf{\Theta}, \lambda_i)p(\mathbf{\Theta}|\lambda_i)d\mathbf{\Theta}. \tag{25}$$

The density $p(\mathbf{y}|\lambda_i)$, being a function of λ_i, can be viewed as the marginal likelihood of λ_i, and λ_i^* is the value maximizing it.

In our baseline specification, the integral in Eq. (25) can be obtained easily in closed form (see for example Geweke, 2005):

$$p(\mathbf{y}|\lambda_i) = (2\pi)^{-T} |\mathbf{\Omega}|^{-1/2} |\mathbf{\Sigma}_{\bar{\alpha}}|^{1/2}$$
$$\times \left|\mathbf{\Sigma}_{\alpha_0}\right|^{-1/2} \exp\left(-\mathbf{Q}/2\right), \tag{26}$$

where $\mathbf{Q} = \mathbf{x}'\mathbf{\Omega}^{-1}\mathbf{x} + \boldsymbol{\alpha}_0'\mathbf{\Sigma}_{\alpha_0}^{-1}\boldsymbol{\alpha}_0 - \bar{\boldsymbol{\alpha}}'\mathbf{\Sigma}_{\bar{\alpha}}^{-1}\bar{\boldsymbol{\alpha}}$.

3.4. Posterior distribution: random error variance

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

$$\mathbf{\Sigma}_v^* \sim IW(\mathbf{\Sigma}_v, \mathbf{v}_0), \tag{27}$$

where $v_0 = mq + 2$, and therefore, by construction, $E[\mathbf{\Sigma}_v^*] = \mathbf{\Sigma}_v$. The prior for α is specified as before, see Eq. (14). In this case, one can derive the following conditional posterior distributions (see for example Geweke, 2005):

$$\boldsymbol{\alpha}|\mathbf{\Sigma}_v^*, \mathbf{y} \sim N(\bar{\boldsymbol{\alpha}}, \mathbf{\Sigma}_{\bar{\alpha}}) \tag{28}$$

$$\mathbf{\Sigma}_v^*|\boldsymbol{\alpha}, y \sim IW(\bar{\mathbf{\Sigma}}_v^*, \bar{\mathbf{v}}) \tag{29}$$

with:

$$\mathbf{\Sigma}_{\bar{\alpha}} = \left(\mathbf{\Sigma}_{\alpha_0}^{-1}(\lambda_i) + \mathbf{\Sigma}_v^{*-1} \otimes \mathbf{X}'\mathbf{X}\right)^{-1} \tag{30}$$

$$\bar{\boldsymbol{\alpha}} = \mathbf{\Sigma}_{\bar{\alpha}} \left(\mathbf{\Sigma}_{\alpha_0}^{-1}(\lambda_i)\boldsymbol{\alpha}_0 + \mathbf{\Sigma}_v^{*-1} \otimes \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}\right) \tag{31}$$

$$\bar{\mathbf{\Sigma}}_v^* = \mathbf{\Sigma}_v + (\mathbf{x}^+ - \mathbf{X}\mathbf{A})'(\mathbf{x}^+ - \mathbf{X}\mathbf{A}) \tag{32}$$

$$\bar{\mathbf{v}} = \mathbf{v}_0 + T. \tag{33}$$

---

[6] This approach has been used by Carriero (2007), Carriero, Kapetanios, and Marcellino (2012) and Carriero et al. (in press), and corresponds to choosing the model with the highest posterior odds (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 which features a closed-form solution for the marginal data density. A similar approach was employed by Schorfheide and Song (2013) using an approximation for the marginal data density of a Mixed Frequency VAR model.

Note that, recalling that $\mathbf{\Xi}'\mathbf{\Omega}_0^{-1}\mathbf{\Xi} = \mathbf{\Sigma}_v^{*-1} \otimes \mathbf{X}'\mathbf{X}$ and $\mathbf{\Xi}'\mathbf{\Omega}_0^{-1}\mathbf{x} = \mathbf{\Sigma}_v^{*-1} \otimes \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}$, Eqs. (30) and (31) resemble Eqs. (16) and (17) closely, with the only difference being that Eqs. (30) and (31) involve the matrix $\mathbf{\Sigma}_v^*$, which is a random variable with mean $E[\mathbf{\Sigma}_v^*] = \mathbf{\Sigma}_v$, while this matrix was fixed to the value $\mathbf{\Sigma}_v^* = \mathbf{\Sigma}_v$ in Eqs. (16) and (17).

The prior defined by Eqs. (14) and (27), known as the independent Normal-Inverse Wishart prior, allows the error variance matrix to be modelled as random, but has the drawback of requiring Gibbs sampling in order to compute the joint posterior distribution.[7] Draws from the joint distribution $p(\boldsymbol{\alpha}, \mathbf{\Sigma}_v^*|\mathbf{Y})$ are obtained by drawing in turn from the conditional distributions in Eqs. (28) and (29).

The evaluation of the integral for computing the marginal distribution as in Eq. (25) poses problems, because a closed-form solution is not available. For models of limited dimension, this problem is typically solved using the Geweke modified harmonic mean estimator (see Geweke, 2005, for details) or the Chib (1995) estimator. However, these methods encountered numerical problems and generated unstable estimates across different chains when applied to our large VAR of $mq = 28$ variables. For this reason, rather than relying on an unstable estimate for λ_i^*, we chose to use the same optimal tightness as was selected for the fixed variance case. As we will see, the forecasts produced with either the fixed or random error variance (based on the optimal tightness of the fixed variance case) are very similar, so that any benefit from using a random error variance would appear to be outweighed by the extra complexity.

4. Empirical forecasting exercise

The purpose of this section is to evaluate the relative forecasting performances of the BVB-VAR models described in Section 3. We assume that at each forecast origin $t + 1$, the aim is to forecast the fully-revised values of data. These will comprise both future time-periods for which no estimates have been published, and past periods for which estimates have been published by the statistical agency. We require forecasts of the fully-revised values of the future observations $\{y_{t+h}\}$ for $h = 1, \ldots, 8$, that is, $\{y_{t+1}, \ldots, y_{t+8}\}$. These are future observations in the sense that, at $t + 1$, the most recent point for which a first estimate is available is y_t. Relative to a forecast origin of $t + 1$, the past periods for which we have fully-revised data are $\{y_{t+h}\}$ with $h = -13, -14, \ldots$, and we will be generating forecasts of $\{y_{t-12}, \ldots, y_t\}$. These are past observations in the sense that, at $t + 1$, an earlier vintage estimate will be available for these observations. For $h = 0$, for example, we forecast the fully-revised value of y_t after having observed the first estimate, y_t^{t+1}. At the other end of the spectrum, $h = -12$, a forecast of the fully-revised value requires a forecast of the 13th (last) revision after having observed the 13th estimate, y_{t-12}^{t+1}.

---

[7] Simulation could be avoided by using a natural conjugate N-IW prior, but this requires a symmetric treatment of the prior of each VAR equation. The prior described in Section 3.1 does not have this symmetry. See Kadiyala and Karlsson (1997) for details.

Forecasts for the VB-VAR models are computed by iteration, where $\mathbf{y}^{t+1+\tau|t+1}$ denotes the τ-step-ahead forecast of the $t + 1 + \tau$-vintage of data. The forecast of the post-revision value of the next quarter's observation $(y_{t+1}^{t+1+q|t+1})$ is the last element of the forecast vector $\mathbf{y}^{t+1+\tau|t+1}$ 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 qth element of the vector $\mathbf{y}^{t+1+\tau|t+1}$ (in the case of the bivariate VAR, assuming that data on output are the first block).

We use quarterly data vintages of 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^{t+1}), so that $y_t^{t+2} \neq y_t^{t+1}$, then the data are 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 is belongs. In addition, there are also periodic benchmark revisions reflecting base year changes, among other things, although working in growth rates lessens their impact: in-depth analyses are provided by Fixler and Grimm (2005, 2008), *inter alia*.

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

4.1. Comparing the BVB-VAR with the univariate VB-VAR

The first question we address is the forecasting accuracy of the BVB-VAR relative to the other non-Bayesian vintage-based VAR models—the unrestricted VB-VAR and the RV-VAR described in Section 2. We use both univariate and bivariate specifications to forecast quarterly growth rates at annual rates of real GDP and the GDP deflator (where the suffixes '_un' and '_bi' in the tables indicate the univariate and bivariate models respectively). 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 accuracies in Table 1. The first column (VB-VAR_un) entries are RMSFEs, and the results in the remaining columns are RMSFE ratios to the univariate VB-VAR benchmark.

Consider first the rows of the table that pertain to forecasting past observations (the rows $h = -12, \ldots, 0$). For both output growth (Table 1, panel A) and inflation (Table 1, panel B), none of the models improve on the benchmark: virtually none of the entries are less than one. This is true of both the univariate and bivariate Bayesian vintage-based VARs estimated with a fixed error-variance matrix (columns headed BVB-VAR_un and BVB-VAR_bi), and also when a random error variance matrix is allowed

(the last column, with the suffix '_G'). In fact, the differences between the fixed and random error-variance matrix Bayesian specifications for the same values of prior tightness are generally small, and we do not comment on these further.

Secondly, the rows reporting RMSFEs for future observations show that the BVB-VAR models produce better forecasts than the benchmark at all horizons for inflation, but not for output growth. For inflation, there are RMSFE reductions of up to 30% at long horizons. This is in sharp contrast to the results obtained when restrictions are imposed (the RV-VAR model), which are generally worse than the benchmark results. The bivariate BVB-VAR registers small improvements over the univariate for inflation.

In summary, the Bayesian vintage-based VAR clearly provides more accurate forecasts than the univariate vintage-based VAR model for forecasting future observations of inflation, but not for output growth.[8] This benchmark was, by and large, the most successful multiple-vintage model evaluated by Clements and Galvão (2012b).

4.2. Comparing the BVB-VAR with other forecasting models

The second question we address is the benefit of modelling the data revision process when forecasting future values of output growth and inflation. To address this question, we use an autoregressive model estimated with real-time-vintage data as a benchmark, as suggested by Clements and Galvão (2013b). This model recognizes that more recent observations are revised lightly relative to earlier, more heavily revised 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 (2013b) for details.

We also consider two alternative VAR models of output growth and inflation, estimated using 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), Carriero et al. (in press), Carriero, Kapetanios, and Marcellino (2009), Carriero et al. (2012), Kadiyala and Karlsson (1997) and Sims and Zha (1998). It is a Bayesian VAR of output and the price level of order 4, and is estimated by imposing both the Minnesota prior and the sum of coefficients prior. The VAR is:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{k=1}^{p} \mathbf{\Psi}_k \mathbf{Y}_{t-k} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{Y}_t$ is a $m \times 1$ vector of macroeconomic variables in log(levels), where all vectors of observations are from the

---

[8] The forecasting performance of the BVB-VAR model is robust to our choice of priors because we select the prior tightness optimally based on the data, as was described in Section 3.2. Still, we have also checked the sensitivity of our results to the prior by rescaling the prior variances of some of the coefficients. In particular, we have rescaled the prior standard deviations of the first two equations of each block (inflation and output) to represent two alternative priors with the overall standard deviation doubled or halved. These changes have no effect on the forecasting performance of our model.

Table 1
Bayesian VB-VAR versus VB-VAR. Vintages at forecast origin: 1989:Q2–2009:Q1.

y_{t+h}	VB-VAR_un	VB-VAR_bi	RV-VAR_un	RV-VAR_bi	BVB-VAR_un	BVB-VAR_bi	BVB-VAR_bi_G
Panel A: Forecasting output growth							
−12	1.213	1.00	1.00	1.00	1.00	1.00	1.00
−11	1.263	0.97	1.03	1.03	1.03	1.03	1.03
−10	1.224	0.98	1.04	1.04	1.04	1.04	1.04
−9	1.227	1.00	1.05	1.05	1.05	1.05	1.05
−8	1.181	1.02	1.04	1.04	1.04	1.04	1.04
−7	1.167	1.04	1.04	1.04	1.04	1.04	1.05
−6	1.129	1.07	1.06	1.06	1.06	1.06	1.06
−5	1.088	1.05	1.11	1.11	1.11	1.11	1.11
−4	1.267	1.02	1.07	1.07	1.07	1.07	1.07
−3	1.320	1.03	1.04	1.04	1.04	1.04	1.04
−2	1.362	1.02	1.04	1.04	1.03	1.03	1.03
−1	1.382	1.02	1.04	1.04	1.04	1.04	1.04
0	1.568	1.02	1.01	1.00	1.00	1.00	1.00
1	2.480	1.13	1.00	1.15	0.99	1.00	1.00
2	2.786	1.11	1.01	1.15	0.98	0.99	0.99
3	2.846	1.09	1.01	1.15	0.98	0.98	0.97
4	2.845	1.12	1.00	1.15	0.99	0.99	0.99
5	2.879	1.08	1.01	1.12	0.99	0.99	0.99
6	2.827	1.06	1.01	1.11	1.01	1.00	1.00
7	2.813	1.06	1.01	1.12	1.01	1.00	1.00
8	2.780	1.05	1.01	1.11	1.02	1.01	1.01
Panel B: Forecasting inflation							
−12	0.556	0.99	1.01	1.01	1.01	1.01	1.01
−11	0.567	1.00	1.05	1.05	1.05	1.05	1.05
−10	0.539	1.02	1.07	1.07	1.07	1.07	1.08
−9	0.510	1.04	1.11	1.11	1.11	1.11	1.11
−8	0.491	1.07	1.10	1.10	1.11	1.11	1.11
−7	0.504	1.09	1.14	1.14	1.14	1.14	1.14
−6	0.544	1.09	1.11	1.11	1.11	1.11	1.11
−5	0.514	1.07	1.15	1.15	1.15	1.15	1.15
−4	0.530	1.03	1.12	1.12	1.11	1.11	1.11
−3	0.559	1.09	1.13	1.13	1.13	1.13	1.12
−2	0.525	1.07	1.13	1.13	1.13	1.13	1.12
−1	0.546	1.06	1.11	1.11	1.11	1.11	1.11
0	0.584	1.08	1.13	1.13	1.13	1.13	1.13
1	0.944	0.98	1.04	1.00	0.96	0.93	0.93
2	1.092	0.97	1.04	1.01	0.82	0.80	0.79
3	1.203	0.96	1.03	0.98	0.79	0.77	0.76
4	1.282	0.94	1.05	0.95	0.78	0.74	0.73
5	1.316	1.03	1.06	1.06	0.78	0.76	0.75
6	1.438	1.00	1.05	1.04	0.73	0.72	0.70
7	1.543	0.97	1.06	1.01	0.71	0.71	0.69
8	1.586	0.93	1.07	0.96	0.67	0.68	0.66

Notes: The first column reports RMSFEs for the VB-VAR_un, and subsequent columns are RMSFE ratios relative to the VB-VAR_un. The rows headed y_{t+h} with $h < 0$ indicate forecasts of past data, in the sense that the forecasts are of the final vintage values of data for which an earlier estimate has already been observed. $h = 0$ corresponds to data for which the first estimate is available. $h > 0$ denotes a forecast of a future observation, in the sense that no estimates are available at the time when the forecast is made. The models are estimated using expanding windows of data. BVB-VAR_bi_G is computed using the values of lambda chosen using BVB-VAR_bi (fixed sigma matrix), but using Gibbs sampling (so that the sigma matrix (disturbances variance) is not fixed).

forecast-origin data vintage. For example, for forecasting at $T + 1$, the VAR:

$$\mathbf{Y}_t^{T+1} = \mathbf{c} + \sum_{k=1}^{p} \mathbf{\Psi}_k \mathbf{Y}_{t-k}^{T+1} + \boldsymbol{\varepsilon}_t, \tag{34}$$

is estimated using $t = 1, \ldots, T$. In particular, we consider the following prior moments:

$$E[\mathbf{\Psi}_k^{(ij)}] = \begin{cases} \Psi^* & \text{if } i = j, k = 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$Var[\mathbf{\Psi}_k^{(ij)}] = \begin{cases} \dfrac{\omega_1^2}{\kappa^2} \dfrac{\sigma_i^2}{\sigma_j^2}, & k = 1, \ldots, p, \end{cases} \tag{35}$$

where $\mathbf{\Psi}_k^{(ij)}$ denotes the element in position (i, j) in the matrix $\mathbf{\Psi}_k$. We estimate this specification in levels, and the prior mean Ψ^* is set to 1. To compute forecasts of growth rates, we use $y_{t+h} = 400(Y_{t+h} - Y_{t+h-1})$, where the forecast Y_{t+h} is computed by iteration. The shrinkage parameter ω_1 measures the overall tightness of the prior, and is chosen via optimization of the marginal likelihood, which is available for this case (see Carriero et al., 2012, Carriero et al., in press, and Giannone et al., 2012) because this prior is symmetric across equations. The parameter κ imposes additional shrinkage on higher order lags. To set each scale parameter σ_i, we follow common practice (Litterman, 1986; Sims & Zha, 1998) and set σ_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 $\overline{Y}_{0i}$, that same value $\overline{Y}_{0i}$ is likely to be a good forecast of future observations.[9] Finally, we add a N-IW prior to the error variance, such that the expectation of the error variance matrix coincides with the traditional Minnesota prior of a 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 Bańbura et al. (2010).

Table 2 compares the forecasting performances of these models. The out-of-sample period is the same as in Table 1, and the models are all estimated again on increasing windows of data. For output growth, none of the VAR models offer any improvements over the benchmark, whether they are Bayesian or not. For inflation, the BVB-VAR models provide sizeable improvements relative to both the benchmark and the VAR(4) and BVAR(4) models when $h > 4$. Neglecting data revisions (as in the VAR(4) and BVAR(4) models) leads to notably inferior inflation forecasts at the longer horizons. Any differences in forecast accuracy between the univariate and bivariate models are small.

Table 3 presents the results of the same forecasting exercise as in Table 2, but with a rolling rather than a recursive forecasting scheme. The use of rolling windows of data has little effect on the forecast comparisons for output growth, whether in absolute or relative terms. For inflation, the use of rolling windows tends to improve the benchmark relative to the VAR models, although the rankings are unchanged and the BVB-VAR_bi continues to provide the best inflation forecasts.

4.2.1. Assessing bias and variance gains

So far, our results suggest large improvements in forecasting inflation from using the BVB-VAR specification. Both the VAR and the BVAR use real-time data only for the current vintage; that is, they disregard the past data revision process. As a consequence, the results based on RMSFE comparisons indicate that the revisions process has a role if the aim is to forecast post-revision values of inflation. One caveat for using RMSFEs for measuring the forecasting accuracy is that forecast gains may arise from reducing either the bias, which could be particularly important if data revisions have a non-zero mean, or forecast-error variance.

Table 4 shows the variances of the forecast errors computed for the benchmark model, and the ratios to these values for the VAR, BVAR, BVB-VAR and VB-VAR models. If $e_{t+h,j}$ is the forecast error for model j for predicting the target variable at $t+h$, then we compute the forecast-error variance using the sample counterparts of:

$$var(e_{t+h,j}) = E\left(e_{t+h,j}^2\right) - \left[E(e_{t+h,j})\right]^2,$$

where the first term on the right-hand-side is the expected squared error and the second term is the squared bias. By comparing Table 4 with Table 2, we can assess the relative importances of the bias and variance in the forecast accuracy improvements. When forecasting output growth, there are only small differences across models, in terms of both the RMSFE and variance. When forecasting inflation, the improvements from using VAR and BVAR models rather than the AR benchmark are mainly in terms of bias.

The MSFE accuracy improvements from using BVB-VAR models are due largely to reductions in the forecast-error variance, as might have been anticipated. The forecast error variance depends on the parameter estimation uncertainty. We would expect that imposing restrictions would reduce the variance, possibly at the cost of an increase in bias. Compared to frequentist-estimated VB-VAR models, Bayesian shrinkage yields variance reductions of around 25% for predicting inflation one year ahead. Using reasonable assumptions about the nature of data revisions as a prior in a Bayesian framework reduces the potentially adverse impact on the forecast accuracy of the high dimensionality of the VAR.
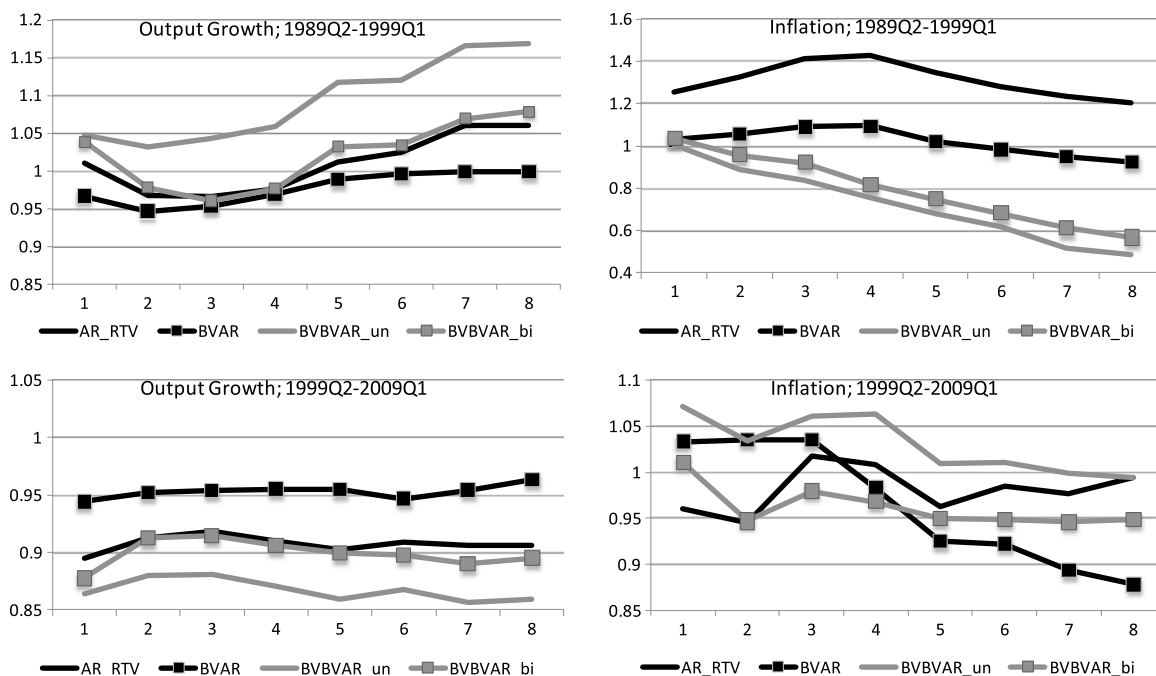
4.3. Sub-sample results

The recent literature suggests that the predictive ability of models of output growth and inflation may change over time (Stock & Watson, 2003, 2007). To investigate this, we split the out-of-sample period into two subsamples, each containing 40 observations, with forecast-origin vintages of 1989:Q2–1999:Q1 and 1999:Q2–2009:Q1, respectively. As a benchmark for comparison, we take the VAR(4), which is also the benchmark used by Bańbura et al. (2010), and it is estimated using real-time data in the traditional way. That is, when the latest vintage of data at the time of forecasting is the $t + 1$-vintage, we estimate the model and condition the forecast on data from that vintage. Next quarter we have the $t + 2$-vintage, and the model is estimated and the forecasts are conditioned on the $t + 2$-vintage values. The model we estimate is of the sort given by Eq. (34), where all of the observations come from the same (the latest) data vintage. This is in contrast to the approach embodied in the VB-VAR, which uses multiple vintages of data to estimate the model. Fig. 1 presents the RMSFE ratios to the VAR(4) benchmark for $h = 1, \ldots, 8$ for both sub-periods and variables.

The choice of the out-of-sample period is shown to affect the ranking between the performances of the bivariate and univariate BVB-VARs. In the earlier period (1989–99), the bivariate model improves the forecasts of output growth, but it does not do so in the later period. In contrast, the bivariate model yields superior inflation forecasts in the later period (1999–2009), but not when the forecast period consists of the 1990s. We also consider whether the modelling of data revisions has the same effect across the two periods; see Fig. 1. For output growth, the BVB-VAR clearly provides more accurate forecasts than the traditional BVAR for the second subsample. For inflation, the BVB-VAR is superior at all forecast horizons in the first subsample, and similarly in the second subsample, except at the longer horizons. In summary, we find some evidence of instability in the relative forecasting performances of the BVB-VAR specifications.

---

[9] Note that the sum of coefficients prior is set here by using the average of the variables over the full sample, as per Bańbura et al. (2010), not the average of the first p observations in the sample, as in other implementations of this prior, such as that of Sims and Zha (1998).

Table 2
Bayesian VB-VAR versus other forecasting models: RMSFE, recursive forecasting scheme. Forecasting vintages: 1989:Q2–2009:Q1.

y_{t+h}	AR_RTV	VAR(4)	BVAR(4)	BVB-VAR_un	BVB-VAR_bi	VB-VAR_un	VB-VAR_bi
	Forecasting output growth						
1	2.482	1.07	1.02	0.99	1.00	1.00	1.13
2	2.742	1.08	1.03	0.99	1.00	1.02	1.13
3	2.795	1.07	1.02	0.99	1.00	1.02	1.11
4	2.825	1.08	1.03	1.00	1.00	1.01	1.12
5	2.837	1.07	1.03	1.00	1.00	1.01	1.09
6	2.842	1.06	1.02	1.00	0.99	0.99	1.06
7	2.838	1.05	1.02	1.00	0.99	0.99	1.05
8	2.818	1.05	1.02	1.01	1.00	0.99	1.04
	Forecasting inflation						
1	0.921	0.94	0.97	0.98	0.95	1.02	1.00
2	1.023	0.90	0.94	0.88	0.85	1.07	1.04
3	1.143	0.85	0.89	0.83	0.81	1.05	1.01
4	1.243	0.84	0.86	0.80	0.77	1.03	0.97
5	1.321	0.88	0.85	0.77	0.76	1.00	1.03
6	1.427	0.88	0.84	0.74	0.73	1.01	1.01
7	1.537	0.90	0.83	0.71	0.71	1.00	0.98
8	1.599	0.89	0.81	0.67	0.68	0.99	0.92

Notes: The first column reports RMSFEs for the benchmark model, the AR_RTV. Subsequent columns give RMSFEs as ratios of the benchmark, where a value less than 1 indicates that the model in question has a smaller RMSFE than the benchmark. The BVAR(4) model is estimated in levels of the variables (all other models are estimated in growth rates) with the 'sum of coefficients' prior. The models are estimated using expanding windows of data.

Notes. The graphs depict RMSFE ratios relative to the VAR(4), for horizons $h = 1,...,8$, separately for the periods 1982Q2–1999Q1 and 1999Q2–2009Q1.

Fig. 1. Consistency of relative forecasting performances.

5. Conclusions

In this paper we propose modelling and forecasting macroeconomic variables that are subject to revisions using a Bayesian vintage-based VAR. The vintage-based VAR models in the literature are typically univariate; that is, they model the relationships between different maturities of data of a single variable. While such approaches have shown promise in forecasting revisions of 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

Table 3

Bayesian VB-VAR versus other forecasting models, rolling forecasting scheme. Vintages at forecast origin: 1989:Q2–2009:Q1.

y_{t+h}	AR_RTV	VAR(4)	BVAR(4)	BVB-VAR_un	BVB-VAR_bi
	Forecasting output growth				
1	2.458	1.08	1.03	1.00	1.01
2	2.724	1.08	1.03	1.00	1.00
3	2.782	1.10	1.03	1.00	0.99
4	2.814	1.10	1.04	1.00	0.99
5	2.825	1.08	1.04	1.00	1.00
6	2.830	1.06	1.02	1.01	0.99
7	2.832	1.04	1.02	1.01	0.99
8	2.812	1.05	1.02	1.01	0.99
	Forecasting inflation				
1	0.861	0.99	1.01	1.04	1.00
2	0.956	0.96	0.99	0.93	0.89
3	1.049	0.90	0.96	0.90	0.86
4	1.120	0.90	0.93	0.89	0.83
5	1.167	0.93	0.93	0.88	0.83
6	1.253	0.93	0.92	0.85	0.81
7	1.332	0.95	0.92	0.84	0.80
8	1.368	0.94	0.91	0.81	0.76

Notes: See the notes to Table 2, except that the estimation uses rolling windows of data.

Table 4

Bayesian VB-VAR versus other forecasting models: variance, recursive forecasting scheme. Forecasting vintages 1989:Q2–2009:Q1.

y_{t+h}	AR_RTV	VAR(4)	BVAR(4)	BVB-VAR_un	BVB-VAR_bi	VB-VAR_un	VB-VAR_bi
	Forecasting output growth						
1	2.456	1.02	0.99	1.00	1.00	1.01	1.13
2	2.707	1.01	0.98	1.01	1.01	1.02	1.11
3	2.757	0.99	0.97	1.01	1.00	1.03	1.09
4	2.787	1.00	0.98	1.01	1.00	1.02	1.10
5	2.806	0.99	0.99	1.01	1.01	1.02	1.07
6	2.810	0.99	0.97	1.01	1.00	1.00	1.04
7	2.810	0.98	0.98	1.01	1.00	1.00	1.03
8	2.795	0.99	0.99	1.01	1.00	0.99	1.03
	Forecasting inflation						
1	0.872	0.99	1.02	1.04	1.01	1.06	1.05
2	0.925	0.99	1.03	0.97	0.94	1.14	1.13
3	1.003	0.95	1.00	0.94	0.91	1.12	1.13
4	1.048	0.97	1.00	0.95	0.89	1.11	1.12
5	1.074	1.02	1.00	0.95	0.92	1.07	1.23
6	1.124	1.02	1.00	0.93	0.89	1.09	1.22
7	1.180	1.04	1.00	0.92	0.89	1.10	1.19
8	1.184	1.03	1.01	0.90	0.87	1.10	1.13

Notes: Entries in the first column are the square root of the forecast error variance, computed as the square root of the difference between the MSFE and the squared average bias (which is the square of the average of the difference between actuals and forecasts). Subsequent columns are forecast error standard deviations as ratios of the benchmark, where a value less than 1 indicates that the model in question has a smaller variance than the benchmark. See also the notes to Table 2.

and inflation, and are shown to provide competitive forecasts of the post-revision (or fully-revised) values of future inflation.

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 *de facto* results in more accurate forecasts than simply 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, but without a large penalty from increasing the parameter uncertainty.

We find that the VB-VAR, in conjunction with Bayesian estimation, delivers the most sizeable improvements for inflation forecasting. For the full 20-year out-of-sample period, the bivariate model is no better than the univariate model. However, for the more recent period (1999–2009), the bivariate VB-VAR is markedly better than the unvariate VB-VAR. We find that our implementation[10] of the Bayesian approach of Bańbura et al. (2010), which does not model multiple vintages, is far less successful for forecasting inflation than the VB-VAR.

---

[10] Note that our implementation of their model is based only on output growth and inflation, whereas their framework allows for a large number of variables.

Acknowledgments

References

Aruoba, B., Diebold, F. X., Nalewaik, J., Schorfheide, F., & Song, D. (2013). *Improving GDP measurement: A measurement-error perspective*. Mimeo: University of Pennsylvannia.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 25(1), 71–92.

Carriero, A. (2007). *A Bayesian framework for the expectations hypothesis: How to extract additional information from the term structure of interest rates*. Working Papers 591, Queen Mary, University of London, School of Economics and Finance. http://ideas.repec.org/p/qmw/qmwecw/wp591.html.

Carriero, A., Clark, T., & Marcellino, M. (2013). Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics*, in press.

Carriero, A., Kapetanios, G., & Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25(2), 400–417.

Carriero, A., Kapetanios, G., & Marcellino, M. (2012). Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking and Finance*, 36(7), 2026–2047.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432), 1313–1321.

Clements, M. P., & Galvão, A. B. (2012a). *Anticipating early data revisions to US output growth*. Mimeo: University of Warwick and Queen Mary University of London.

Clements, M. P., & Galvão, A. B. (2012b). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business and Economic Statistics*, 30(4), 554–562.

Clements, M. P., & Galvão, A. B. (2013a). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, 29(4), 698–714.

Clements, M. P., & Galvão, A. B. (2013b). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28, 458–477.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G., & Labhard, V. (2012). A state space approach to extracting the signal from uncertain data. *Journal of Business and Economic Statistics*, 30, 173–180.

Doan, T., Litterman, R., & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3, 1–100.

Fixler, D. J., & Grimm, B. T. (2005). Reliability of the NIPA estimates of U.S. economic activity. *Survey of Current Business*, 85, 9–19.

Fixler, D. J., & Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, 88, 16–32.

Garratt, A., Lee, K., Mise, E., & Shields, K. (2008). Real time representations of the output gap. *Review of Economics and Statistics*, 90, 792–804.

Garratt, A., Lee, K., Mise, E., & Shields, K. (2009). Real time representations of the UK output gap in the presence of model uncertainty. *International Journal of Forecasting*, 25, 81–102.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*. New Jersey: John Wiley & Sons, Wiley.

Giannone, D., Lenza, M., & Primiceri, G. E. (2012). *Prior selection for vector autoregressions*. CEPR Discussion Papers 8755, C.E.P.R. Discussion Papers.

Hecq, A., & Jacobs, J.P.A.M. (2009). *On the VAR-VECM representation of real time data*. Technical report. Mimeo: University of Maastricht, Department of Quantitative Economics.

Jacobs, J. P. A. M., Sarferaz, S., Sturm, J., & van Norden, S. (2013). *Modelling multivariate data revisions*. CIRANO—Scientific Publication N. 2013/44.

Jacobs, J. P. A. M., & van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of 'true' values. *Journal of Econometrics*, 161, 101–109.

Kadiyala, K. R., & Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12(2), 99–132.

Kishor, N. K., & Koenig, E. F. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*, 30, 181–190.

Koenig, E. F., Dolmas, S., & Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, 85(3), 618–628.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177–203.

Landefeld, J. S., Seskin, E. P., & Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal of Economic Perspectives*, 22, 193–216.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics*, 4, 25–38.

Patterson, K. D. (1995). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure. *Economic Journal*, 105, 54–76.

Patterson, K. D. (2003). Exploiting information in vintages of time-series data. *International Journal of Forecasting*, 19, 177–197.

Schorfheide, F., & Song, D. (2013). *Real-time forecasting with a mixed-frequency VAR*. NBER working paper no. 19712.

Sims, C. A., & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4), 949–968.

Stark, T., & Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24, 507–531.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41, 788–829.

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *Supplement to Vol. 39*, 3–33.

Stock, J. H., & Watson, M. W. (2009). Phillips curve inflation forecasts, understanding inflation and implications for monetary policy. Federal Reserve Bank of Boston.

Stock, J. H., & Watson, M. W. (2010). *Modelling inflation after the crisis*, NBER Working Paper Series 16488.

Theil, H., & Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics. *International Economic Review*, 2(1), 65–78.

Andrea Carriero is Professor of Economics at the School of Economics and Finance, Queen Mary University of London.

Michael P. Clements is Professor of Econometrics at ICMA Centre, University of Reading, and associate member of the Institute for New Economic Thinking, Oxford Martin School, University of Oxford.

Ana Beatriz Galvão is Associate Professor of Economic Modelling and Forecasting at Warwick Business School, University of Warwick.