Forecasting with Vector Autoregressive Models of Data Vintages: US output growth and inflation.

Michael P. Clements
Department of Economics
University of Warwick
M.P.Clements@warwick.ac.uk

Ana Beatriz Galvão*
School of Economics and Finance
Queen Mary University of London
a.ferreira@qmul.ac.uk

July 21, 2011

Abstract

Vintage-based vector autoregressive models of a single macroeconomic variable are shown to be a useful vehicle for obtaining forecasts of different maturities of future and past observations, including estimates of post-revision values. The forecasting performance of models that include information on annual revisions is superior to models that only include the first two data releases. However, empirical results indicate that a model that more closely reflects the seasonal nature of data releases is not found to offer much improvement over an unrestricted vintage-based model that includes three rounds of annual revisions.

Keywords: data revisions, forecasting, data uncertainty.

JEL code C53.

*Corresponding author: Dr. Ana Beatriz Galvao. Queen Mary University of London, School of Economics and Finance, Mile End Road, E1 4NS, London, UK. Phone: ++44-20-78828825. email: a.ferreira@qmul.ac.uk. We are grateful to seminar participants at Bath, Warwick, and the IIF Workshop in Verbier, and our discussant at the IIF Workshop, Tara Sinclair, for helpful comments. We are also grateful to two anonymous referees for their comments and suggestions.

1 Introduction

The first or 'advance' estimates of national accounts data issued by statistical agencies are based on partial source data and are subject to revision. These estimates are typically revised many times. The initial revisions typically reflect the availability of more complete source data, and subsequent annual revisions incorporate new annual source data into the estimates. Finally, comprehensive or benchmark revisions make use of major periodic source data, as well as methodological and conceptual improvements.[1] From a policy perspective, these data revisions mean that there is a good deal of uncertainty about the true current (and recent past) state of the economy. In this paper, we present a real-time forecasting evaluation of data that are subject to revision. It is real time in the sense that, at each point in time, the forecasting models are specified and the parameters estimated using only data for time periods up to that point in time, *and* the vintages of data used are restricted to those which would have been available at that time.

Clements and Galvão (2011) show that a certain class of models can be used to forecast 'fully-revised' or 'post-revision' values of past and future observations, and they assess the value of those forecasts in terms of their contribution to improving real-time estimates of the output gap and trend inflation and inflation gap.[2] We undertake a more detailed investigation of the forecasting performance of these models. Our interest is not just in forecasting post-revision values, but we also consider their ability to forecast observations of different maturities (from lightly-revised to fully-revised data) published at the same date (vintage).

The class of models we consider are the vintage-based vector autoregressive (V-VAR) models of Hecq and Jacobs (2009) and Garratt, Lee, Mise and Shields (2008, 2009). The key characteristic of these models is that the vector of variables being modelled consists of estimates of recent observations on the variable of interest from the current vintage of data. We consider a number of different VAR specifications for modelling data on each variable of interest, that is, inflation and output growth. One variant seeks to better approximate the publication pattern of data releases by the statistical agency. Another specification imposes restrictions consistent with the assumption that revisions published after the first revision are efficient, in the sense that they are unpredictable based on past data vintages.

[1] In the case of US National Accounts data, the Bureau of Economic Analysis provides descriptions of the methodologies employed at: `http://www.bea.gov/methodologies/index.htm#national_meth`

[2] Orphanides (2001) and Orphanides and van Norden (2002) show that estimates of the output gap based on final data can be markedly different from those available in real time, affecting both historical evaluations of monetary policy and the effective conduct of monetary policy in real time. Clements and Galvão (2011) show that real-time estimates of trend inflation and the inflation gap (computed using Stock and Watson (2007, 2010) model) will also differ from historical estimates of these quantities.

Our models focus on the role of past vintages of data in modelling and forecasting future data vintages, and in so doing we neglect the potential usefulness of explanatory variables. Much of the recent literature looks at forecasting using large numbers of explanatory variables, as in the dynamic factor model literature (see, e.g, Stock and Watson (2011) for a recent review) but ignores the data revisions dimension by taking only the latest-available dataset at the time of the investigation. In principle, our approach is readily extended to include explanatory variables subject to revision, so that the vector of variables being modelled could include the vintage estimates for a number of variables. In practice the single-variable multiple-vintage VAR models already comprise a large number of parameters by considering a large number of estimates from each vintage, so such extensions would likely best be handled within a Bayesian framework with large numbers of parameters shrunk to prior values. We leave for future research the general question of whether explanatory variables, including past vintages of data available for such variables, could be used to improve the forecasts of the quantities we are interested in here.

In a recent review of forecasting with real-time data, Croushore (2006) finds that the results of forecasting with state-space models that incorporate data revisions are mixed, compared to simply ignoring data revisions. Examples of multiple-vintage models include Harvey, McKenzie, Blake and Desai (1983), Howrey (1984), Patterson (1995, 2003), Jacobs and van Norden (2011), Cunningham, Eklund, Jeffery, Kapetanios and Labhard (2009), Garratt *et al.* (2009, 2008) and Hecq and Jacobs (2009).

Based on the models we consider, our findings are more promising, especially for inflation. Our main contributions are as follows. For US output and inflation, we provide an extensive evaluation of vintage-based VAR model forecasts of a range of maturities of data using a variety of different 'actuals'. We distinguish between forecasting future observations and revisions to past data. We consider the performance of models that offer a 'better' characterisation of the release practices of the statistical agency, and explain with a Monte Carlo why the forecasts are no better than with the unrestricted vintage-based VAR. We also assess the information content of annual revisions to these two key US macro variables.

The plan of the remainder of the paper is as follows. Section 2 describes the basic VAR model and the alternative versions. Section 3 is a detailed study of forecast performance, where we consider forecasts of a range of data maturities. In this section we also consider the imposition of cointegrating restrictions based on levels representations relative to specifying models in growth rates, and present a Monte Carlo aimed at illuminating some aspects of the empirical findings.

Section 4 compares the V-VAR models against standard practices for forecasting both first-release[3] and latest-vintage actuals, as these are the mainstay of model forecast comparisons when data are subject to revision. Section 5 offers some concluding remarks.

2 The multiple-vintage VAR models

The models we consider are related to the vintage-based VAR (V-VAR) of Hecq and Jacobs (2009) and the models of Garratt *et al.* (2008, 2009), and are described in Clements and Galvão (2011). Here we briefly describe those models, and a number of additional variants. These models assume that the data revision process can be modelled based only on observed components. [4]

We work with growth rates, so that y_t^{t+1} is the growth rate at period t computed using data from vintage $t+1$. This corresponds to the BEA 'advance estimate', which we shall call the first estimate, to avoid confusion. The advance estimate is made available toward the end of the first month of the following quarter. We use the real-time datasets of Croushore and Stark (2003), which record the data available in the middle of the second month of the following quarter, which corresponds to the advance estimate. Our second estimate, or first revised value, is y_t^{t+2}, which corresponds to the BEA 'final' estimate. A key characteristic of the BEA data releases is that the third estimate will be unrevised, i.e., $y_t^{t+3} = y_t^{t+2}$, unless $t + 3$ is a third quarter of the year, in which case the third estimate of y_t will incorporate an annual revision, and $y_t^{t+3} \neq y_t^{t+2}$. It is sometimes assumed that revisions after the first are essentially unpredictable. For, example, Garratt *et al.* (2008) and Clark (2010) both use the BEA 'final' estimate (our y_t^{t+2}) as actual values for computing forecast errors. Their choice of target variable is based on the assumption that annual and benchmark revisions are generally unpredictable.

2.1 Unrestricted model

We begin by glossing over this institutional detail, and simply suppose that past and current vintages of data can be used to predict estimates published in future vintages. The current vintage $t + 1$ includes observations from 1 up to t. If we suppose that the true value of the observation at t is approximate by its value after $q - 1$ quarterly revisions, that is, y_t^{t+q}, then we can use a vector

[3] Some studies use the estimates available two quarters after the reference quarter as the actuals, rather than the estimates available in the following quarter.

[4] Examples of models with unobserved components include Jacobs and van Norden (2011) and Cunningham *et al.* (2009), *inter alia*. Kishor and Koenig (2011) build on earlier contributions by Howrey (1978, 1984) and Sargent (1989) and use a Kalman filtering approach to estimate post-revisions values based on the current vintage of data.

of q estimates from the current vintage to build a vintage-based VAR:

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

where $\mathbf{y}^{t+1} = \left[y_t^{t+1}, y_{t-1}^{t+1}, \ldots, y_{t-q+1}^{t+1}\right]'$, $\mathbf{y}^{t+1-i} = \left[y_{t-i}^{t+1-i}, y_{t-1-i}^{t+1-i}, \ldots, y_{t-q+1-i}^{t+1-i}\right]'$, and $\mathbf{c}$ is $q \times 1$, $\boldsymbol{\varepsilon}^{t+1}$ is $q \times 1$. Notice that y_{t-q}^{t+1} would be redundant if added as the $(q+1)^{th}$ element of $\mathbf{y}^{t+1}$, because we assume that y_{t-q}^{t+1} is identical to y_{t-q}^{t} (the last element of $\mathbf{y}^t$). The vintage-based VAR (V-VAR) models the dynamics of successive vintages of data that include both a new observation y_t^{t+1} and revised estimates of past observations $y_{t-1}^{t+1}, \ldots, y_{t-q+1}^{t+1}$. The variance-covariance matrix of the disturbances ($\Sigma_\varepsilon = E(\boldsymbol{\varepsilon}^{t+1}\boldsymbol{\varepsilon}^{t+1\prime})$) captures the correlations between data revisions published in the same vintage.

2.2 Periodic specifications

The V-VAR ignores the periodicity of the publication of data revisions. In the case of US BEA data, the elements of $\mathbf{y}^{t+1}$ other than the first two, y_t^{t+1} and y_{t-1}^{t+1}, will typically remain unchanged unless the $t+1$-vintage is an annual ($t+1 \in Q3$) or a benchmark revision.[5] In practice, a small modification to this simple seasonal pattern is required. When benchmark revisions are anticipated to be published in January, annual revisions may not be published in the previous July.[6] As a consequence, we define two dummy variables: $D_1^{t+1} = 1$ if an annual revision has been published (always in the third quarter), and $D_2^{t+1} = 1$ if either an annual or a benchmark revision has been published. The majority of cases that $D_2^{t+1} = 1$ are for vintages published in the third quarter, but there are instances for other quarters, especially the first quarter. When $s = 1$, so that only annual revisions are included, the model is known as a seasonal vintage-based VAR (SV-VAR). When $s = 2$, so that both annual and benchmark revisions are included, the model is termed a seasonal and benchmark vintage-based VAR (SBV-VAR).

The model with $p = 1$ is:

$$\mathbf{y}^{t+1} = \left[\widetilde{\mathbf{c}} + \widetilde{\mathbf{\Gamma}}_1 \mathbf{y}^t\right](1 - D_s^{t+1}) + \left[\mathbf{c} + \mathbf{\Gamma}_1 \mathbf{y}^t\right] D_s^{t+1} + \mathbf{v}^{t+1} \tag{2}$$

[5] The GNP/GDP data of the BEA are subject to three annual revisions in the July of each year: see, e.g., Fixler and Grimm (2005, 2008) and Landefeld, Seskin and Fraumeni (2008).

[6] There are 8 benchmark revisions in the data vintages from 1965:Q3 up 2010:Q1 that comprise our estimation sample. In fact there are 36 annual Q3 revisions rather than the 44 that would otherwise have occurred. There are 44 combined benchmark and annual revisions ($D_2^{t+1} = 1$) - the 8 benchmark revisions and the 36 annual revisions.

where:

$$\widetilde{\Gamma}_1 = \begin{bmatrix} & \boldsymbol{\gamma}_{2\times q} & \\ \mathbf{0}_{(q-2)\times 1} & \mathbf{I}_{(q-2)\times(q-2)} & \mathbf{0}_{(q-2)\times 1} \end{bmatrix} \tag{3}$$

and $\widetilde{\mathbf{c}}= (c_1, c_2, 0, ..., 0)'$.

Thus, when quarter $t+1$ does not incorporate an annual or benchmark revision, then $D_s^{t+1} = 0$, and hence $y_{t+1-i}^{t+1} = y_{t+1-i}^{t}$ (up to a random error term, v_{t+1-i}^{t+1}) for $i = 3, \ldots, q$. But when $D_s^{t+1} = 1$, y_{t+1-i}^{t+1} is determined by the coefficients in the i^{th} row of $\mathbf{\Gamma}_1$ multiplied into $\mathbf{y}^t$. Hence the above model captures the seasonal aspect of the BEA revisions to national accounts data because whether, say, $y_t^{t+3} - y_t^{t+2} = 0$ depends on the quarter of the year to which t belongs ($y_t^{t+3} - y_t^{t+2} = 0$ unless t falls in Q4).

In theory, we might also let the publication of an annual revision affect the first and second equations (the equations for y_t^{t+1} and y_{t-1}^{t+1}). But to keep the model relatively simple, we assume that the first two rows of $\mathbf{\Gamma}_1$ and $\widetilde{\mathbf{\Gamma}}_1$ are equal. Finally, note that the form of $\widetilde{\mathbf{c}}$ is such that intercepts are only estimated in the equations for $(y_{t-2}^{t+1}, \ldots, y_{t-q+1}^{t+1})$ when $D_s^{t+1} = 1$.

2.3 A restricted specification

We suppose that after one revision, the next estimate y_t^{t+3} is an efficient forecast, in the sense that the revision from y_t^{t+2} to y_t^{t+3} is uncorrelated with y_t^{t+2}. We can impose this restriction on the VAR, where it translates to $E\left(y_{t-2}^{t+1} \mid y_{t-2}^{t}\right) = y_{t-2}^{t}$. This is achieved by specifying Γ_1 in (1) as:

$$\widetilde{\Gamma}_1 = \begin{bmatrix} & \boldsymbol{\gamma}_{2\times q} & \\ \mathbf{0}_{(q-2)\times 1} & \mathbf{I}_{(q-2)\times(q-2)} & \mathbf{0}_{(q-2)\times 1} \end{bmatrix} \tag{4}$$

and when $p > 1$, it requires in addition that the Γ_i coefficient matrices are restricted to:

$$\widetilde{\Gamma}_i = \begin{bmatrix} \boldsymbol{\gamma}_{i,2\times q} \\ \mathbf{0} \end{bmatrix}$$

for all $i > 1$. Hence the restricted VAR (for $p = 1$ and $n = 2$) is:

$$\mathbf{y}^{t+1} = \mathbf{c} + \widetilde{\mathbf{\Gamma}}_1 \mathbf{y}^t + \mathbf{v}^{t+1} \tag{5}$$

where $\widetilde{\mathbf{\Gamma}}_1$ is given by (3). An unrestricted intercept is included in each equation, so the interpretation of the model is that revisions $y_{t+1-i}^{t+1} - y_{t+1-i}^{t}$ are uncorrelated with y_{t+1-i}^{t} for $i \geq 3$, but may be non-zero mean. For example, $E\left(y_{t+1-i}^{t+1} - y_{t+1-i}^{t}\right) = c_i$, for $i \geq 3$, where c_i is the i^{th} element of $\mathbf{c}$.

We name this model the 'news-restricted' vintage-based VAR, RV-VAR.

2.4 Additional variants

Although by working in terms of growth rates we would expect to have largely mitigated the effects of level shifts in the series due to changes of base year etc., we consider two variants of the models we have already outlined which allow for changes in intercepts. One is a seasonal model, where we take the SV-VAR and extend it to include a separate intercept dummy for each of the benchmark revisions. This model we denote the SV-VAR+BD (SV-VAR plus 'benchmark dummies'). Whereas the SBV-VAR assumes that regular and benchmark revisions have the same impact on $\mathbf{y}^{t+1}$,[7] the SV-VAR+BD model dummies out any intercept shifts from the benchmark revisions.

The other variant allows an alternative treatment of deterministic terms in the RV-VAR. Specifically, we now allow an intercept in the equations for y_{t-2}^{t+1} to y_{t-q+1}^{t+1} of the RV-VAR only when $D_2^{t+1} = 1$: this we term the RV-VAR+SBD (RV-VAR plus 'seasonal and benchmark dummies'). The RV-VAR allows intercepts in these equations for all t.

The V–VAR is an unrestricted VAR, so OLS applied equation-by-equation delivers efficient estimates. Because the elements of the vector of disturbances ε^{t+1} are likely to be correlated, OLS applied to any of the restricted specifications (RV-VAR, SV-VAR and SBV-VAR, etc.) is not efficient, and instead estimation of these models is carried out by the seemingly unrelated regression estimator (SURE), as this is equivalent to maximum likelihood (e.g., Hamilton (1994, p. 317)).

3 Forecasting data releases

In this section we assess the out-of-sample forecast performance of the vintage-based VAR models when forecasting data releases. We are interested in the forecast performance of the S(B)V-VAR and RV-VAR, and variants, relative to the basic V-VAR model, as well as whether the multiple-vintage models improve on simple single-vintage autoregressions. We begin by describing the nature of the forecasts generated by the VAR models, and we distinguish between forecasting future observations, and forecasting data revisions to past observations.

The VAR models provide forecasts of the future values of the vector of variables in $\mathbf{y}^{t+1}$, where the variables relate to future vintage values of recent and past observations. To make matters concrete, consider a forecast origin $T+1$. This means that the latest available observation is for y_T, and

[7] This is justified by the fact that we are working in growth *rates*, and that the annual and benchmark revisions are not always kept separate. As noted, annual revisions are generally not published in July if benchmark revisions are planned for the following January.

it is the first estimate, y_T^{T+1}. At this time, all the data vintages up to and including the time-$T+1$ vintage are known, i.e., $\mathbf{y}^{T+1}, \mathbf{y}^T, \ldots$ where $\mathbf{y}^{T+1-i} = (y_{T-i}^{T+1-i}, ..., y_{T-i-q+1}^{T+1-i})'$ for $i = 0, 1, 2, \ldots$. The h-step ahead forecast of the vector $\mathbf{y}^{T+h+1}$ is defined as $\mathbf{y}^{T+h+1|T+1} \equiv E\left(\mathbf{y}^{T+h+1} \mid \mathbf{y}^{T+1}, \mathbf{y}^T, \ldots\right)$. The elements of $\mathbf{y}^{T+h+1|T+1}$ are $(y_{T+h}^{T+h+1|T+1}, ..., y_{T+h-q+1}^{T+h+1|T+1})$, and thus provide forecasts of the first estimate of y_{T+h}, a forecast of the second estimate of y_{T+h-1}, and so on down to the q^{th} estimate of $y_{T+h-q+1}$. The forecasts are computed by iteration in the standard way based on the estimated models.

We can evaluate the elements of the h-step forecast vector $\mathbf{y}^{T+h+1|T+1}$ in two different ways: i) against the maturities of the data that are explicitly being forecast, and ii) using actual values drawn from the last-available vintage at the time the study is undertaken, on the grounds that these will constitute the best possible estimates of the true values. So, under i), the j^{th} element $y_{T+h+1-j}^{T+h+1|T+1}$ is compared to the actual value $y_{T+h+1-j}^{T+h+1}$, whereas under ii) the same element of the forecast vector would be evaluated against $y_{T+h+1-j}^{f}$, where f is the latest available vintage.

In some ways the first option might seem the most natural, as the forecasts are compared against maturities that are explicitly being targeted. Moreover, the first few revisions following the release by the statistical agency of the first estimate may well be predictable, but one can imagine a situation in which this might be hidden by comparing forecasts of post-revision values to final or latest-available vintage estimates. Option i) permits an investigation of the predictability of revisions at different maturities. On the other hand, in some cases it is the post-revision values of the series which are of interest,[8] especially if it is the case that maturities after a small number of revisions are largely unpredictable. In this section we focus on i), and in section 4 we present a more conventional assessment of the forecast accuracy of the models which includes consideration of ii).

In our setup, consider the forecast vector $\mathbf{y}^{T+h+1|T+1}$ of $\mathbf{y}^{T+h+1} = (y_{T+h}^{T+h+1}, ..., y_{T+h+1-q}^{T+h+1})$. If we assume that $h < q$, the first h elements of $\mathbf{y}^{T+h+1|T+1}$, i.e., $\left\{y_{T+h+1-i}^{T+h+1|T+1}\right\}$ for $i = 1, ..., h$ will constitute predictions about future observations (relative to y_T), and the remaining $q-h$ elements (for $i = h+1, ..., q$) will be forecasts of future releases of observations that have already been published in vintage T. These are the forecasts of revisions to past data. When $h > q$, all the elements of $\mathbf{y}^{T+h+1|T+1}$ relate to future observations. We will draw this distinction between forecasting future observations and revisions to past data when we consider forecast performance.

The assessment of the value of the VAR model forecasts is made relative to some simple benchmark forecasts. For forecasting future observations the benchmark model we use is the real-time-

[8] As in the real-time estimation of output and inflation 'gaps' and 'trends': see, e.g., Clements and Galvão (2011).

vintage (RTV) autoregression of Clements and Galvão (2010):[9]

$$y_t^{t+1} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i}^t + \varepsilon_t. \tag{6}$$

When $p = q$, (6) corresponds to the first equation of the V-VAR, although p will often be less than q. In some circumstances, this approach is shown by Clements and Galvão (2010) to be preferable to the traditional or 'end-of-sample' (EOS) approach that simply estimates an AR model on the single data vintage available at the forecast origin. When data revisions are non-zero mean, Clements and Galvão (2010) show that the use of the RTV autoregression (6) will not in general produce unbiased forecasts for data vintages beyond the first release. Koenig *et al.* (2003) use as the left-hand-side variable in (6) the data which have been revised a sufficient number of times to be an efficient estimate of whatever vintage it is one is trying to forecast. Clements and Galvão (2010) compare their suggestion with using the first-release data as the left-hand-side variable, as in (6), and then bias-correcting the resulting forecast. For moderate sample sizes, the latter works reasonably well, and is the approach we adopt here. We add a correction to the RTV forecasts that depends on the target vintage, so that for $h > 1$ the bias-correction for vintage $T + 1$ is given by: $bc^{T+h|T+1} = (T - h)^{-1} \sum_{t=h+1}^{T} (y_{t-(h-1)}^{t+1} - y_{t-(h-1)}^{t+1-(h-1)})$. This is the sample estimate (using data up to the forecast origin) of the mean of the revision between the first release and the estimate published $h - 1$ quarters later.

When it comes to predicting past data revisions, a natural benchmark is the 'no-change prediction' (i.e., a random walk forecast whereby $y_{t-i+h+1}^{t+h+1|t+1} = y_{t-i+h+1}^{t+1}$ for $h + 1 \leq i \leq q$).

By comparing the forecast accuracy of the benchmark with vintage-based VARs, we can assess the predictability of annual and benchmark revisions. Such revisions are generally the only source of deviation between y_{t+1-i}^{t+h+1} and y_{t+1-i}^{t+1} for $i \geq 2$. That is, after an observation has been revised once, in the absence of annual revisions we would expect to find $y_{t-1}^{t+h+1} = y_{t-1}^{t+1}$ and $y_{t-2}^{t+h} = y_{T-2}^{t}$, etc., for $h \geq 1$. An indication that annual and benchmark revisions are in part predictable would be a finding in favour of the vintage-based VAR forecasts $\hat{y}_{t-1}^{t+h+1|t+1}$, $\hat{y}_{t-2}^{t+h+1|t+1}$,. . . , compared to the benchmark forecasts y_{t-1}^{t+1}, y_{t-2}^{t+1},.... Note that the RV-VAR forecasts of $\hat{y}_{t+1-n}^{t+h+1|t+1}$ for $n \geq 2$ (i.e., the forecasts for data revised at least once) are not equal to the benchmark forecasts: the RV-VAR forecasts are given by $\hat{y}_{t+1-n}^{t+h+1|t+1} = y_{T+1-n}^{t+1}$ (the benchmark forecast) *plus* an intercept term, for $n \geq 2$, because the RV-VAR allows a non-zero intercept for all q equations.

[9]See also Koenig, Dolmas and Piger (2003) on single-equation approaches for formulating forecasting models when there are data revisions.

Section 3.1 compares the forecasts to actuals which are of the same maturities as the data explicitly being forecast. Section 3.2 investigates 'levels' specifications of the VAR models and the role of cointegration, and section 3.3 is a Monte Carlo aimed at illuminating the findings of the empirical forecast comparisons.

3.1 Evaluation using matching maturities

Forecast errors are computed as $\mathbf{e}^{t+h} = \mathbf{y}^{t+h} - \mathbf{y}^{t+h|t}$ for $t = T+1, .., T+N$, where N is the number of the observations in the out-of-sample period. Using the forecast errors, we compute individual measures of forecast accuracy as well as the trace of the forecast-error covariance matrix, where the forecast-error covariance matrix is defined in population as the square matrix of order q, $\Sigma_h = E\left[\mathbf{e}^{t+1}\mathbf{e}^{t+1\prime}\right]$. The individual MSFEs (or RMSFEs) are the diagonal elements of Σ_h. So, the individual MSFEs are calculated as:

$$\frac{1}{N}\sum_{t=T+1}^{T+N}\left(y_{t+h-i}^{t+h} - y_{t+h-i}^{t+h|t}\right)^2$$

where the summation is over t for a given h, and where $i = 1, 2, \ldots 14$. The sum of these MSFEs (over i), i.e., the trace measure, $trace(\Sigma_h)$, provides a useful summary in some circumstances compared to simply considering the diagonal elements separately. As summary measures, we also sum the the MSFEs that correspond to revisions to past data, and those that correspond to future observations, to emphasize any differences in forecast performance between forecasting 'future observations' and forecasting 'data revisions'.

The models are estimated on real-time data on real output (GNP/GDP) and the deflator published by the Philadelphia Fed (see Croushore and Stark (2001)). We make use of data vintages from 1965:Q3 up 2010:Q1 (that is, 178 vintages). The out-of-sample period covers vintages from 1995:Q3 up to 2007:Q1 ($N = 47$). The models are estimated using increasing windows of data starting with the 1965:Q4 vintage.[10] Recursive estimation is employed because both the SV-VAR and SBV-VAR models need a long enough history to estimate the dynamics of annual revisions. The vintage-based VARs are estimated with $p = 1$ at each forecasting origin. We set $q = 14$, implying that we consider maturities that have only suffered initial revisions and up to three rounds of annual revisions.

Tables 1 and 2 present MSFEs of the benchmark model for $h = 1, 4, 8, 12$, as well as ratios of the

[10] The total number of vintages is $T + N = 167$ while remaining $h = 12$ vintages are employed to compute forecast errors. Note also that even in the case of ealier vintages, we required q observations, so we use observations from 1960:Q1.

MSFEs for all the vintage-based VAR models to the benchmark. We include all the models discussed in section 2 in the forecast exercise: the unrestricted VAR (V-VAR); the news-restricted (RV-VAR); the periodic VAR with annual revisions (SV-VAR); the periodic VAR with annual and benchmark revisions (SBV-VAR); the SV-VAR with benchmark dummies (SV-VAR+BD); and RV-VAR with seasonal and benchmark dummies (RV-VAR+SBD). Brief descriptions of the key characteristics of these models are provided in table A for convenience. When computing forecasts for the SV-VAR and the SBV-VAR, we assume that we know the future values of $D_s^{t+2|t+1}, ..., D_s^{t+h+1|t+1}$. This means that the forecasts are conditional on the dates of the annual and benchmark revisions being assumed known: at least for the former this seems unobjectionable. To aid interpretation and cross model comparisons, in both tables 1 and 2 we bolden the entries corresponding to the most accurate model.

Consider Table 1 for output growth. When forecasting one-step ahead (i.e., the next vintage values, $h = 1$), none of the multiple-vintage VAR models are able to improve on the accuracy of the benchmark forecasts judged on the multivariate measure, although improvements to forecasts of specific data releases are apparent (e.g., the forecasts $\hat{y}_{t-11}^{t+2|t+1}$ from the SV-VAR+BD are 16% more accurate than from the benchmark model). In terms of the trace measure, the SV- and SBV models are worse than both the V-VAR and RV-VAR. At the longer horizons, the V-VAR and RV-VAR models remain the best VAR forecasting models. The gains to the V-VAR and RV-VAR relative to the benchmark are small, and are in terms of forecasting 'future observations' rather than 'data revisions': the last two rows of each panel of the table report the trace measure (relative to the benchmark) separately for each of the two sets of forecasts.

The similarity in the forecast performance of the V-VAR and the RV-VAR suggests that the restriction that revisions after the first are unpredictable (apart from possibly having a non-zero mean) does not harm overall forecast accuracy. However, there is some limited evidence that data vintage-estimates beyond the first revision (i.e., y_t^{t+2+i}, $i = 1, 2, \ldots$) are predictable since the V-VAR performs better than the RV-VAR and the benchmark for a few mature observations when $h = 1, 4$.

There are no generalised forecasting gains from fully specifying a model that incorporates the periodicity of the publication of data revisions (SBV-VAR versus V-VAR). The two variant models SV-VAR+BD and RV-VAR+SBD do generally not improve forecast accuracy.

The results for inflation (Table 2) are similar to the extent that the SV-VAR and SBV-VAR are again the worst performing models overall (on the trace measure), but the V-VAR and RV-VAR outperform the benchmark for predicting future observations at $h = 1$, and clearly dominate the benchmarks at longer horizons, especially for predicting data revisions to past observations.

There is thus an interesting difference between output growth and inflation in terms of the relative ranking of the VAR models and the benchmark in terms of predicting future observations versus revisions. The improvement in forecast accuracy of the V-VAR for output growth when predicting vintage $t + 13|t + 1$ (relative to the benchmark, and true of the other VAR models as well, although to a lesser extent) is largely in terms of predicting future observations (see the rows $t+6$ down to $t+1$). But in the case of inflation, the largest gains are realized from the predictability of data revisions using the V-VAR.

This indicates greater predictability of the revisions to estimates of inflation than of revisions to output growth estimates. Thus for inflation there is clear evidence that vintage-estimates beyond the first revision are predictable (relative to the 'no-change' benchmark).

3.2 Models in growth rates versus assuming stationarity of data revisions

Following Patterson (1995), Garratt *et al.* (2008, 2009) work in terms of the level (of the log) of output, so that consideration of issues to do with whether different vintages are cointegrated arise. They assume the revisions are stationary. Instead, we work with growth rates, and revisions in growth rates. On the face of it, ignoring long-run information constitutes a form of model mis-specification. That said, there is now a body of literature that suggests that error-correcting models will perform poorly when there are (unmodelled) shifts (see, e.g., Clements and Hendry (2006, 2011)) as in the case of shifts caused by occasional changes in the base year or other methodological changes in the definition of the series. Models specified solely in terms of growth rates might be more robust to the effects of changes in levels due to re-basing and other benchmark revisions. We compare the V-VAR model's forecast performance against that of our implementation of Garratt *et al.* (2008), to assess the impact of neglecting the long-run relationships that may exist between the different vintage estimates of the (log) levels. We briefly describe our implementation of Garratt *et al.* (2008) (henceforth the GLMS-VAR) model, and then turn to the forecast comparison exercise. An appendix outlines why models specified solely in terms of growth rates are more likely to be robust to the effects of changes in levels due to changes of base year and other benchmark revisions, in the spirit of Clements and Hendry (2006, 2011).

Letting Y denote the log-level, the GLMS-VAR can be written as a VAR for the vector $\mathbf{Z}^{t+1}$:

$$\mathbf{Z}^{t+1} = \left(\begin{bmatrix} Y_t^{t+1} \\ Y_{t-1}^{t+1} \\ Y_{t-2}^{t+1} \\ Y_{t-3}^{t+1} \end{bmatrix} - \begin{bmatrix} Y_{t-1}^{t} \\ Y_{t-1}^{t} \\ Y_{t-2}^{t} \\ Y_{t-3}^{t} \end{bmatrix} \right)$$

The first term $Y_t^{t+1} - Y_{t-1}^t$ a difference across vintage and observation, and is used to predict future observations, while the remaining terms are data revisions to past data. The autoregressive order is set to 2, and lag operator is applied to both vintages and observations, i.e.:,

$$\mathbf{Z}^{t+1-i} = \left(\begin{bmatrix} Y_{t-i}^{t+1-i} \\ Y_{t-1-i}^{t+1-i} \\ Y_{t-2-i}^{t+1-i} \\ Y_{t-3-i}^{t+1-i} \end{bmatrix} - \begin{bmatrix} Y_{t-1-i}^{t-i} \\ Y_{t-1-i}^{t-i} \\ Y_{t-2-i}^{t-i} \\ Y_{t-3-i}^{t-i} \end{bmatrix} \right)$$

for $i = 1, 2$. It is not clear how Garratt *et al.* (2008) treat the effects of changes in the base year. Our implementation includes a dummy variable for each benchmark revision. These dummies are found to capture significant level changes, as expected. From this model we obtain forecasts of the (log) levels, namely, $(\hat{Y}_{t+h}^{t+h+1|t+1}, \hat{Y}_{t+h-1}^{t+h+1|t+1}, \hat{Y}_{t+h-2}^{t+h+1|t+1}, \hat{Y}_{t+h-3}^{t+h+1|t+1})$, from which we calculate forecasts of same-vintage growth rates:

$$\hat{\mathbf{y}}^{t+h+1|t+1} = \begin{bmatrix} 400(\hat{Y}_{t+h}^{t+h+1|t+1} - \hat{Y}_{t+h-1}^{t+h+1|t+1}) \\ 400(\hat{Y}_{t+h-1}^{t+h+1|t+1} - \hat{Y}_{t+h-2}^{t+h+1|t+1}) \\ 400(\hat{Y}_{t+h-2}^{t+h+1|t+1} - \hat{Y}_{t+h-3}^{t+h+1|t+1}) \end{bmatrix}.$$

These are comparable to the forecasts summarized in tables 1 and 2. But note that the GLMS-VAR only provides forecasts of a subset of the variables reported in those tables. Table 3 reports the forecast comparisons of the V-VAR and GLMS-VAR for the common variables, expressed relative to the benchmark. For output growth, the GLMS offers some improvement over the V-VAR for this small subset of maturities for 2 of the 4 forecast horizons ($h = 1, 8$) on the trace measure, but is much worse for inflation.[11]

In summary, it does not appear that ignoring level terms has a serious impact on forecast accuracy. Even if the data generating process is a cointegrated VAR in the levels, the costs of

[11] An autoregressive order of 4 is used for forecasting inflation, because of the greater persistence of inflation compared to output growth. Note that Garratt *et al.* (2008) do not apply their model to the GDP deflator.

omitting cointegrating combinations are likely to be tempered by having a model that is more robust to changes in the base year.

3.3 A Monte Carlo investigation of the forecasting performance of the SV-VAR

Our empirical forecast comparisons show that modelling the seasonal nature of revisions does not result in more accurate forecasts, even though the resulting models more closely reflect the practice of the US government statistics agency, the BEA. There are a number of reasons why this might be the case.

Firstly, consider the comparison of the SV-VAR to the V-VAR. The SV-VAR imposes the restriction that $E\left(y_{t-2}^{t+1} \mid \mathbf{y}^t, \mathbf{y}^{t-1} \ldots\right) = y_{t-2}^t$ when $t+1 \notin Q3$. Suppose this is true. However, the behaviour of y_{t-2}^{t+1} may be similar irrespective of the quarter $t+1$ falls in (so, for example, $E\left(y_{t-2}^{t+1} \mid \mathbf{y}^t, \mathbf{y}^{t-1} \ldots\right) \simeq y_{t-2}^t$ for $t+1 \in Q3$ as well). The SV-VAR estimates the equation for y_{t-2}^{t+1} on only one quarter of the sample (those t for which $t+1 \in Q3$). This may worsen forecasts because of high parameter estimation uncertainty relative to estimating the equation for y_{t-2}^{t+1} on the whole sample (as in the V-VAR). Relatedly, 'shrinkage' has often been found to improve VAR forecasts (e.g., the Bayesian shrinkage approach of Doan, Litterman and Sims (1984)). Estimating the equation for y_{t-2}^{t+1} of the V-VAR using all observations - those for which y_{t-2}^{t+1} is equal to y_{t-2}^t, as well as those for which the coefficient on y_{t-2}^t is close to but not equal to unity ($t+1 \in Q3$), amounts to shrinking this coefficient to one, which may pay dividends in terms of forecast accuracy.

One can also think of this in terms of imposing parameters at values which are incorrect but nevertheless close to the true values, rather than freely estimating the parameters. For example, the RV-VAR imposes coefficients for the equations y_{t-i}^{t+1}, $i \geq 2$, but these might be sufficiently close to the true values that more accurate forecasts result using squared-error loss criteria (such as MSFE) once an allowance is made for parameter estimation uncertainty (see, e.g., Clements and Hendry (1998) and Giacomini and White (2006)).

The aim of the Monte Carlo is to investigate the relative importance of these factors for different sample sizes, especially the number of vintages of the in-sample period. The data generating process in the Monte Carlo is the SV-VAR. This is estimated on the full sample of data vintages for output growth (Table 4.A) and inflation (Table 4.B).[12] Data are simulated assuming multivariate normal distribution for the disturbances. We present results for two sample sizes ($T+N$): 120 and 500. The number of data-vintages in the out-of-sample period is $N = 47$, matching the empirical exercise.

[12] The model estimated with full sample uses D_2^{t+1}, that is, $D_2^{t+1} = 1$ if either an annual or a benchmark revision is published. Data is generated assuming regular revisions published every year in Q3.

On each replication we calculate the $trace(\Sigma_h)$ measure for $h = 1, 4, 8, 12$. We generate forecasts from the benchmark, RV-VAR, V-VAR and SV-VAR, with $q = 14$, and run a separate exercise with only three maturities so that results for the GLMS-VAR model can also be included.

The results (Tables 4.A and 4.B) indicate that the correctly-specified model, which is the SV-VAR, is only marginally the best forecasting model on the shorter sample (120). This indicates that even when the SV-VAR is a 'perfect' description of the data (apart from the model's parameters being unknown) the forecast gains relative to the rival VAR models and the benchmark are modest, unless the estimation sample is large. A comparison of the results with 120 and 500 vintages of data serves to bring out the impact of estimation uncertainty.

In addition, because only 25% of the vintages in the out-of-sample period include benchmark or annual revisions, the benchmark model and the RV-VAR will closely mimic the SV-VAR for 75% of the vintages at least for $h = 1$ (both for forecasting future observations, and for forecasting data revisions). To assess the force of this, Table 4.C compares separately the one-step ahead relative forecast performance of the VAR models (against the benchmark) for those observations subject to annual revisions ($D_1^{t+1} = 1$). Hence evaluations are over two groups of forecast errors, those for which $D_1^{t+1} = 1$, and those for which $D_1^{t+1} = 0$. The Monte Carlo results with a small sample show that when there are no annual revisions, the benchmark and RV-VAR are on a par, and that the V-VAR forecasts are less accurate. When there are annual revisions, the deficiencies of the benchmark and RV-VAR forecasts are highlighted - for output growth and inflation these models produce forecasts which are around 15 and 30% less accurate than the SV-VAR (when $T+N = 120$). Recall that the benchmark forecasts for $h = 1$ are of 'no-change' for all the 13 forecasts produced of data revisions, and a forecast from an RTV model of the one future observation. The RV-VAR will produce forecasts of no-change (apart from an intercept) for the 3rd to 14th elements of the vector of forecasts. As is apparent, the similarity in forecast accuracy of the V-VAR and RV-VAR forecasts of both output growth and inflation across all observations (both around 10% worse than the SV-VAR for inflation when $T + N = 120$) covers a wide disparity in the accuracy of the two models across the two sets of observations.

The large sample results show that the SV-VAR is superior to the V-VAR for predicting observations affected by annual revisions ($D_1^{t+1} = 1$), whereas when $T + N = 120$, the two are roughly equal. For the short sample, the 'shrinkage' caused by estimating the VAR model across data which are revised and data which are unrevised[13] tends to offset the estimation uncertainty that comes

[13]By which we mean, observations for which y_{t-i}^t is equal to y_{t-i}^{t-1}, $i = 2, 3, \ldots$, as well as those for which $y_{t-i}^t \neq y_{t-i}^{t-1}$ when $t \in Q3$.

from having a short sample. On the large sample, the parameters are more precisely estimated, and this form of averaging over different types of observations instead results in model mis-specification which inflates the MSFE of the V-VAR relative to the SV-VAR.

In summary, our Monte Carlo results suggest that the SV-VAR is unlikely to clearly dominate the other models even if it were a good approximation to the data generating process unless the number of vintages employed in the estimation is large. In practice, of course, the 'regularity' of data revisions imposed in the Monte Carlo by simulating data from an estimated SV-VAR model may only hold to varying degrees, which will further disadvantage the SV-VAR model compared to 'no-change' forecasts.

4 Forecasting recent past and future observations: Using final and early-vintage actuals

The vintage published at $T+h+1$ will have estimates that are yet to be revised (the first element of $\mathbf{y}^{T+h+1}$ is y_{T+h}^{T+h+1}) as well as only lightly-revised data, moving to heavily-revised and post-revision data. The $t+h+1$ data vintage was the yardstick for assessing the accuracy of forecasts made at time $t+1$ in section 3.1. This meant that for a particular observation, we matched the data maturity explicitly being forecast to the same maturity actual value. Frequently, though, forecasts are evaluated against post-revision estimates, which typically means data from the last available vintage. In fact, forecasters may be interested in the accuracy of only a subset of the maturities considered in the previous forecasting exercise. Conditional on the current vintage $t+1$, we suppose forecasters are interested in the next four values of y_t, namely, y_{t+1} to y_{t+4}. In a separate exercise, the focus shifts to the four most recent past observations, namely y_{t-3} to y_t. We allow that the forecaster may wish to forecast either an early-release or the latest-available vintage data in the case of predicting future observations, and either post-revision data (after $q-1$ revisions) or the latest-available vintage in the case of predicting past data. As argued earlier, the vintage-based VAR model forecast $\hat{y}_{T+h-q+1}^{T+1+h|T+1}$ is designed to forecast the value of $y_{T+h-q+1}$ in the q^{th} release, but it might also provide competitive forecasts of $y_{T+h-q+1}^{f}$ (where f is the latest available vintage) if revisions after the first few are essentially unpredictable. Because the vintage-based VAR specifications use $q=14$, we set the post-revision maturity to y_t^{t+14}. The 2010:Q1 vintage is the latest-available vintage. We compute forecasts for observations up to 2006:Q2, because the remaining observations are still subject to regular revisions. Differences between y_t^{t+14} and $y_t^{f=2010:Q1}$ are mainly due to benchmark revisions. Lee, Olekalns and Shields (2010) use y_t^{t+2} as their post-revision value for evaluating forecasts of the output level from a multivariate version of the GLMS-VAR model. To facilitate comparison with their results, we also report results using the second release to compute forecast errors.

The forecasting exercise in this section makes use of rolling windows of data to estimate the models. We use a window of 120 observations, for an out-of-sample period of 1995:Q3 up to 2006:Q3 ($N = 45$). The use of a rolling estimation scheme facilitates testing of whether the differences in forecast performance between the models that we observe are statistically significant. Because the vintage-based VAR specifications nest the benchmark forecasting models, for both future observations (eqn. 6) and for data releases (random walk), care is required in obtaining tests which are correctly sized. Clark and West (2007) propose a modification to the Diebold and Mariano (1995) test to take into account the parameter estimation uncertainty of the nesting model exceeding that of the smaller 'null' model, while keeping the test distribution approximately normal. However, for testing equal accuracy of multi-step forecasts ($h > 1$) using a recursive scheme (estimation with increasing windows of data) as in section 3, it is not clear that the normal distribution is a good approximation to the distribution of the test statistic under the null (see, e.g., Clark and McCracken (2005)). As a consequence, in this section we use a rolling scheme. For the benchmark forecast denoted by $y_{b,t}^{t+f|t-h+f}$, and the vintage-based model forecast of $y_{b,t}^{t+f|t-h+f}$, we compute the following (adjusted) loss function:

$$d_{t,h} = (y_t^{t+f} - y_{b,t}^{t+f|t-h+f})^2 - [(y_t^{t+f} - y_{v,t}^{t+f|t-h+f})^2 - (y_{b,t}^{t+f|t-h+f} - y_{v,t}^{t+f|t-h+f})^2], \qquad (7)$$

for $t = T, ..., T+N-1$. f defines the forecasting target. $f = 1$ implies that the target is first-released data, and $f = 14$ refers to data after 13 revisions. We also use $f =$ 2010:Q1 – the latest available vintage, as actual values to evaluate forecasts – using $f = 14$ to compute forecasts. The t-statistics are calculated using the sample mean loss $\overline{d}_h$ and the standard deviation $std(d_{t,h})$ (computed using the Newey-West estimator) of the adjusted loss function (eqn. 7). As we are use rolling estimation windows, we compute p-values for the test using the normal distribution.

Tables 5 and 6 record the results for future observations and the past data, respectively. Consider future observations. Table 5 contains three panels for each of output growth and inflation, corresponding to forecasting the first, second or final release. The forecasts of first-release data are obtained from the first equation of the VAR. That is, conditional on the vintage $t + 1$ (and for $t = T, ..., T + N - 1$), we compute $\hat{y}_{t+h}^{t+h+1|t+1}$, for $h = 1, 2, 3, 4$. This generates 1 to 4-step ahead forecasts of, respectively, the first release values of y_{t+1} to y_{t+4}. Forecasts of second releases come from the second equation of the VAR: $\hat{y}_{t+h-1}^{t+h+1|t+1}$, for $h = 2, 3, 4, 5$. The forecasts of the latest-available vintage are obtained using the last equation of the VAR, that is, we compute $\hat{y}_{T+h-13}^{t+h+1|t+1}$, for $h = 14, 15, 16, 17$. Table 5 reports results for all the vintage-based VAR specifications described in section 2, and the GLMS-VAR described in section 3.2. Forecasts from the GLMS-VAR are computed in a similar fashion: for the first-release, as $400(\hat{Y}_{t+h}^{t+h+1|t+1} - \hat{Y}_{t+h-1}^{t+h+1|t+1})$ for $h = 1, .., 4$, and for the second release and the final as $400(\hat{Y}_{t+h-1}^{t+h+1|t+1} - \hat{Y}_{t+h-2}^{t+h+1|t+1})$ for $h = 2, .., 5$.

The set of models we consider permits an evaluation of the predictive content of data of releases beyond the BEA 'final' (i.e., y_t^{t+2} in our notation) for forecasting future observations. For instance, the V-VAR and the SV-VAR include information on annual revisions which is excluded from the RV-VAR and the GLMS-VAR. As before, the benchmark model is an autoregression estimated with real-time vintage (RTV) with bias-correction depending on the forecasting target as in (6). We also report results for an AR estimated on the single vintage of data at the forecast origin - EOS. This is the traditional way of forecasting in real-time, as the information set is confined to data vintages that would have been available at that time. That said, it uses only one vintage of data at each forecast origin: the latest-available at that forecast origin.

The first column (headed 'RTV') in Table 5 presents root mean squared forecast errors (RMSFE) for the benchmark model (6), and the values in the remaining columns are ratios of RMSFEs to the benchmark RMSFE. For output growth, the vintage-based VAR models do not improve upon the benchmark irrespective of the vintage we forecast. For forecasting future inflation, we observe statistically significant improvements in accuracy from the vintage VAR models at the four step ahead horizon for forecasting both first and second-release data, and at 1-step for the second-release data. Contrary to the findings of Clements and Galvão (2010) for a longer sample period, the use of RTV data to estimate autoregressive models is not found to improve on the use of EOS data. For this shorter out-of-sample period, the impact of the change from fixed-weighting to chain-weighting in the computation of real output in 1996 may explain why the multiple-vintage models and also the use of RTV data generates less accurate forecasts than the use of a single-vintage (EOS) to estimate autoregressive models.

Lee *et al.* (2010) show that their multivariate version of the GLMS-VAR model provides more accurate forecasts of first-release and post-revision (i.e., second release) output levels, compared to a VAR model estimated using only vintage $T+1$ data (traditional, end-of-sample approach - EOS). However, our results indicate that the GLMS-VAR is generally less accurate than the V-VAR for forecasting inflation. This provides support for the contention that annual revisions help forecast future observations (because the GLMS-VAR only includes information on the last two revisions), but note that there are other differences between the models (e.g., the GLMS-VAR model imposes cointegration across maturities, while the V-VAR is in terms of growth rates: see the appendix) which caution against drawing strong conclusions.

Table 6 presents RMSFE measures of forecasts of the last four observations in vintage $t+1$, i.e., y_{t-3}^{t+1} to y_t^{t+1}; where $t = T, .., T+N-1$. Only y_t^{t+1} is a first-release value, all the other observations have undergone at least one revision. Our forecasts are of the post-revision values of these observations, and so are taken to be $\hat{y}_{t+h-13}^{t+h+1|t+1}$, for $h = 10, 11, 12, 13$, corresponding to the forecast from the last equation of the VAR. To compute forecasts from the GLMS-VAR, we use $400(\hat{Y}_{t+h-2}^{t+h+1|t+1} - \hat{Y}_{t+h-3}^{t+h+1|t+1})$ with $h = 1$ as the prediction of y_{t-1}, and $400(\hat{Y}_{t+h-1}^{t+h+1|t+1} - \hat{Y}_{t+h-2}^{t+h+1|t+1})$ with $h = 2$ as the prediction of y_T. As in section 3, the benchmark model is a random walk, that is, $\hat{y}_{t-i+1} = y_{t-i+1}^{t+1}$ for $i = 1, ..., 4$.

When predicting the post-revision value of output growth, the vintage-based VAR models are not better than the benchmark forecasts, while the GLMS-VAR shows a small gain for observations that have been revised once (the rows labelled '$t-1$'). When predicting the latest available data, the V-VAR yields statistically significant improvements in accuracy on RMSFE for all maturities other than the first. These results suggest there is no information in past vintages for forecasting subsequent revisions to the first release value of output growth. The results for inflation indicate a different pattern: the V-VAR and RV-VAR models are able to predict data revisions to inflation of all maturities. The comparison between the RV-VAR and the V-VAR clearly shows the value of data revisions to mature data for predicting data revisions to 'lightly-revised' observations, especially when the actuals are taken to be latest-vintage values.

The results in this section confirm the findings in section 3 that models that incorporate the seasonal nature of data releases generally deliver worse forecasts that unrestricted models such as the V-VAR.

5 Conclusions

We have undertaken an in-depth study of the predictability of data vintages and post-revision values to the two headline US macro variables, output growth and inflation. The VAR models of data vintages on a single variable can claim some success in characterizing and forecasting the data revisions process for US inflation, but are less useful for US output growth. There is some evidence that vintage-based VAR models provide more accurate forecasts of future observations of output growth, but clear evidence that revisions to past inflation data are predictable.

Our seasonal VAR models were designed to more closely reflect the way in which the statistical agency revises data. However, they were generally no better than the unrestricted VAR in terms of out-of-sample forecasting of data vintages and post-revision values of output growth and inflation. Our Monte Carlo study offered some support for our conjectures as to why this might be the case, and suggested that datasets containing more vintage estimates would be necessary for this model to be competitive with the simpler models.

Finally, we found that modelling growth rates rather than levels incurred little loss from a forecasting perspective.

6 Appendix A

To show that the Garratt *et al.* (2008) formulation is susceptible to structural breaks, and this comes about because it includes levels of (log) output. Our models are immune to changes in levels which leave growth rates unchanged.

Their VECM representation, when $p = 1$, is given by (their equation (A1)):

$$\begin{bmatrix} y_t^{t+1} - y_{t-1}^t \\ y_{t-1}^{t+1} - y_{t-2}^t \end{bmatrix} = a + \Gamma \begin{bmatrix} y_{t-1}^t \\ y_{t-2}^t \end{bmatrix} + \varepsilon_t \tag{8}$$

where $\Gamma = \alpha\beta'$, $\beta' = [1 \ -1]$.

Suppose this is the true process for $t = 1, \ldots, \tau - 1$.

At τ the data are published after a change of the base year. This means $y_{\tau+i}^{\tau+1+i,*} = d + y_{\tau+i}^{\tau+1+i}$, $i = 0, 1, \ldots$, so the log of output is changed by a fixed amount. The same is true of $y_{\tau-1+i}^{\tau+1+i,*}$ etc.

This means that in period τ the DGP will experience a level shift:

$$\begin{bmatrix} y_\tau^{\tau+1,*} - y_{\tau-1}^\tau \\ y_{\tau-1}^{\tau+1,*} - y_{\tau-2}^\tau \end{bmatrix} = I_2 d + a + \Gamma \begin{bmatrix} y_{\tau-1}^\tau \\ y_{\tau-2}^\tau \end{bmatrix} + \varepsilon^{\tau+1}$$

Next period the intercept reverts to its original value:

$$\begin{bmatrix} y_{\tau+1}^{\tau+2,*} - y_\tau^{\tau+1,*} \\ y_\tau^{\tau+2,*} - y_{\tau-1}^{\tau=1,*} \end{bmatrix} = a + \Gamma \begin{bmatrix} y_\tau^{\tau+1,*} \\ y_{\tau-1}^{\tau+1,*} \end{bmatrix} + \varepsilon^{\tau+2}$$

because $y_{\tau+1}^{\tau+2,*} - y_\tau^{\tau+1,*} = y_{\tau+1}^{\tau+2} - y_\tau^{\tau+1}$, and note that $\beta' y^t = \beta' y^{t,*}$ for all t.

So if we try to forecast next vintage with the vintage $\tau + 1$ (containing observations through τ), there will be an expected forecast error of $I_2 d$.

If instead we estimate the model on the data up to vintage $\tau + 1$ (so on observations through τ), the estimate of the intercept will be (in population) $a + \frac{d}{T}$, where T is the number of observations.

This is not a result of defining growth rates as the change in first estimates (and second estimates) of adjacent observations, rather than (as we do) defining growth rates on the same vintage (e.g., $y_t^{t+1} - y_{t-1}^{t+1}$).

To see this, note that we can re-write (8) as:

$$\begin{bmatrix} y_t^{t+1} \\ y_{t-1}^{t+1} \end{bmatrix} = a + (\Gamma + I_2) \begin{bmatrix} y_{t-1}^t \\ y_{t-2}^t \end{bmatrix} + \varepsilon^{t+1}$$

and then pre-multiplying by the non-singular matrix

$$P = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

gives:

$$\begin{aligned} \begin{bmatrix} y_t^{t+1} - y_{t-1}^{t=1} \\ y_{t-1}^{t+1} \end{bmatrix} &= Pa + \left(P\alpha\beta' + P\right) \begin{bmatrix} y_{t-1}^t \\ y_{t-2}^t \end{bmatrix} + \varepsilon^t \\ &= Pa + \begin{bmatrix} \left(\beta'\alpha + 1\right) & 0 \\ \alpha_2 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1}^t - y_{t-2}^t \\ y_{t-2}^t \end{bmatrix} + \varepsilon_t. \end{aligned}$$

In this representation, the first equation, which corresponds to our 'same-vintage' growth rates is immune to re-basing, whereas the second equation will be non-constant.

References

Clark, T. E. (2010). Real-time density forecasting from BVARs with stochastic volatility. *Journal of Business and Economic Statistics*. Forthcoming.

Clark, T. E., and McCracken, M. W. (2005). Evaluating direct multi-step forecasts. *Econometric Reviews*, **24**, 369–404.

Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, **138**, 291–311.

Clements, M. P., and Galvão, A. B. (2010). Real-time forecasting of inflation and output growth in the presence of data revisions. Warwick economics research paper, no. 953, Department of Economics, University of Warwick.

Clements, M. P., and Galvão, A. B. (2011). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. Discussion paper, Department of Economics, University of Warwick.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

Clements, M. P., and Hendry, D. F. (2011). Forecasting from misspecified models in the presence of unanticipated location shifts. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting, Chapter 10*, pp. 271–314. Oxford: Oxford University Press.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 961–982: Elsevier, Horth-Holland.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105**, 111–130.

Croushore, D., and Stark, T. (2003). A real-time data set for macroeconomists: Does the data vintage matter?. *The Review of Economics and Statistics*, **85**, 605–617.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G., and Labhard, V. (2009). A state space approach to extracting the signal from uncertain data. *Journal of Business & Economic Statistics*. ahead of print. doi:10.1198/jbes.2009.08171.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

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

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

Fixler, D. J., and 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., and Shields, K. (2008). Real time representations of the output gap. *Review of Economics and Statistics*, **90**, 792–804.

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

Giacomini, R., and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74**, 1545 – 1578.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Harvey, A. C., McKenzie, C. R., Blake, D. P. C., and Desai, M. J. (1983). Irregular data revisions. In Zellner, A. (ed.), *Applied Time Series Analysis of Economic Data*, pp. 329–347: US Department of Commerce, Washington D.C., Economic Research Report ER-5.

Hecq, A., and Jacobs, J. P. A. M. (2009). On the VAR-VECM representation of real time data. Discussion paper, mimeo, University of Maastricht, Department of Quantitative Economics.

Howrey, E. P. (1978). The use of preliminary data in economic forecasting. *The Review of Economics and Statistics*, **60**, 193–201.

Howrey, E. P. (1984). Data revisions, reconstruction and prediction: an application to inventory investment. *The Review of Economics and Statistics*, **66**, 386–393.

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

Kishor, N. K., and Koenig, E. F. (2011). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*. Forthcoming.

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

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

Lee, K., Olekalns, N., and Shields, K. (2010). Nowcasting, Business Cycle Dating and the Interpretation of New Information when Real-Time Data are Available. Discussion paper.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, **91(4)**, 964–985.

Orphanides, A., and van Norden, S. (2002). The unreliability of output gap estimates in real time. *The Review of Economics and Statistics*, **84**, 569–583.

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.

Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy*, **97**, 251–287.

Stock, J. H., and 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., and Watson, M. W. (2010). Modelling Inflation after the Crisis. *NBER Working Paper Series*, **16488**.

Stock, J. H., and Watson, M. W. (2011). Dynamic factor models. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting, Chapter 2*, pp. 35–60. Oxford: Oxford University Press.

Table A: Glossary. Description of multiple-vintage models

Acronyms	Description
V-VAR	Vintage-based VAR; unrestricted model. See eq. (1) in main text.
RV-VAR	Restricted Vintage-based VAR; restricted such that releases after the second are unpredictable, except perhaps by a constant amount. See eqs. (4) and (5).
SV-VAR	Seasonal Vintage-based VAR; unrestricted for vintages in which annual revisions are published, and restricted as the RV-VAR for vintages when annual revisions are not published. See eqs. (2) and (3).
SBV-VAR	Seasonal Benchmark Vintage-based VAR; unrestricted for vintages in which either annual revisions or benchmark revisions are published, and restricted as the RV-VAR for vintages when neither type of revisions is published. See eqs. (2) and (3).
RV-VAR + SBD	Restricted Vintage-based VAR + seasonal and benchmark dummy intercepts; restricted as the RV-VAR, except that the intercept is estimated only when either annual or benchmark revisions are published.
SV-VAR + BD	Seasonal Vintage-based VAR + benchmark dummies; unrestricted for vintages in which either annual revisions or benchmark revisions are published, and restricted as the RV-VAR for vintages when neither type of revision is published. Separate intercept dummies are included for each benchmark revision.
GLMS-VAR	Garratt, Lee, Mise and Shields (2008) VAR model. See section 3.2.

Table 1. Comparing forecasts of vintages of output growth at h=1,4,8,12 (vintages at origin: 1995:Q3-2007:Q1; N=47).

	Forecasting vintage t+2|t+1						
	Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD
t+1	**3.124**	1.031	1.036	1.025	1.032	1.039	1.025
t	**0.454**	1.086	1.076	1.096	1.087	1.076	1.091
t-1	**0.106**	1.012	1.008	2.230	1.603	1.043	2.208
t-2	0.195	0.979	1.015	**0.924**	1.617	1.017	0.971
t-3	**0.220**	1.000	1.008	1.341	1.360	1.009	1.265
t-4	0.231	1.044	0.995	1.332	0.941	**0.940**	1.323
t-5	**0.164**	1.077	1.003	1.776	1.842	1.094	1.794
t-6	**0.114**	1.394	1.113	1.627	2.300	1.370	1.624
t-7	0.115	1.026	**0.983**	1.514	1.130	1.040	1.514
t-8	0.150	1.067	1.043	1.053	**0.881**	1.179	1.048
t-9	**0.133**	1.022	1.009	1.944	1.350	1.062	1.940
t-10	**0.125**	1.009	1.010	1.296	1.297	1.035	1.358
t-11	0.171	0.916	1.002	0.842	0.851	0.978	**0.839**
t-12	**0.081**	1.135	1.031	1.234	1.199	1.062	1.231
$tr(\hat{\Sigma}_{h=1})$: all	**5.381**	1.040	1.032	1.150	1.139	1.048	1.149
$tr(\hat{\Sigma}_{h=1})$: obs.	**3.124**	1.031	1.036	1.025	1.032	1.039	1.025
$tr(\hat{\Sigma}_{h=1})$: rev.	**2.258**	1.052	1.028	1.324	1.287	1.059	1.321

	Forecasting vintage t+5|t+1						
	Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD
t+4	2.938	1.124	1.097	1.069	1.120	1.114	1.091
t+3	3.566	1.024	0.993	**0.972**	1.035	1.006	0.992
t+2	3.603	1.014	**0.997**	1.037	1.070	1.007	1.052
t+1	3.867	**0.933**	0.965	1.018	0.988	0.970	1.027
t	**0.913**	1.108	1.067	1.329	1.475	1.080	1.328
t-1	0.708	**0.871**	1.009	1.370	1.362	0.995	1.341
t-2	0.847	**0.923**	1.033	1.113	1.191	1.020	1.102
t-3	**0.760**	1.039	1.070	1.374	1.277	1.076	1.354
t-4	0.629	**0.986**	1.031	1.460	1.266	1.116	1.459
t-5	**0.528**	1.055	1.069	1.329	1.313	1.221	1.331
t-6	**0.527**	1.091	1.079	1.425	1.336	1.230	1.421
t-7	0.482	**0.974**	1.011	1.594	1.205	1.087	1.604
t-8	0.606	**0.963**	1.003	1.297	1.028	1.033	1.311
t-9	0.528	**0.967**	1.009	1.281	1.097	1.026	1.301
$tr(\hat{\Sigma}_{h=4})$: all	**20.502**	1.011	1.018	1.124	1.118	1.039	1.134
$tr(\hat{\Sigma}_{h=4})$: obs.	**13.974**	1.017	1.008	1.022	1.049	1.019	1.038
$tr(\hat{\Sigma}_{h=4})$: rev.	**6.528**	0.998	1.039	1.344	1.266	1.081	1.340

	Forecasting vintage t+9|t+1						
	Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD
t+8	**3.857**	1.075	1.057	1.061	1.065	1.062	1.075
t+7	3.884	1.007	0.998	**0.972**	1.030	1.008	0.989
t+6	**3.797**	1.122	1.077	1.083	1.188	1.094	1.102
t+5	**3.643**	1.093	1.064	1.131	1.174	1.083	1.155
t+4	**4.179**	1.084	1.057	1.095	1.170	1.077	1.115
t+3	**3.910**	1.025	1.019	1.046	1.151	1.036	1.062
t+2	4.118	**0.977**	0.989	1.019	1.100	0.998	1.031
t+1	4.044	**0.937**	0.960	1.026	1.067	0.969	1.029
t	**1.471**	1.170	1.161	1.470	1.519	1.210	1.450
t-1	**1.055**	1.048	1.083	1.629	1.465	1.191	1.592
t-2	**1.107**	1.029	1.110	1.500	1.375	1.224	1.480
t-3	**1.078**	1.041	1.068	1.688	1.470	1.162	1.670
t-4	**1.136**	0.984	1.051	1.454	1.346	1.131	1.446
t-5	**1.082**	0.981	1.049	1.428	1.335	1.138	1.423
$tr(\hat{\Sigma}_{h=8})$: all	**38.360**	1.041	1.039	1.138	1.173	1.065	1.148
$tr(\hat{\Sigma}_{h=8})$: obs.	**31.431**	1.039	1.027	1.053	1.118	1.040	1.069
$tr(\hat{\Sigma}_{h=8})$: rev.	**6.930**	1.049	1.091	1.524	1.423	1.178	1.506

	Forecasting vintage t+13|t+1						
	Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD
t+12	5.968	1.013	1.014	1.007	**0.998**	1.014	1.012
t+11	7.740	0.998	0.999	**0.978**	0.984	0.999	0.983
t+10	7.475	1.002	0.993	**0.978**	0.995	0.997	0.985
t+9	7.414	0.997	0.989	**0.979**	0.988	0.991	0.984
t+8	**5.952**	1.091	1.065	1.084	1.086	1.070	1.094
t+7	4.776	0.997	**0.985**	1.019	1.028	0.998	1.032
t+6	4.207	0.995	**0.972**	0.993	1.084	0.988	1.009
t+5	4.042	0.952	**0.948**	1.001	1.026	0.961	1.019
t+4	3.872	**0.928**	0.940	0.957	1.045	0.965	0.973
t+3	3.814	**0.928**	0.929	0.942	1.027	0.947	0.960
t+2	3.725	**0.916**	0.943	0.988	1.059	0.970	1.003
t+1	3.969	**0.892**	0.928	1.042	1.038	0.933	1.056
t	**1.506**	1.180	1.170	1.588	1.460	1.247	1.581
t-1	**1.368**	1.040	1.085	1.687	1.447	1.181	1.672
$tr(\hat{\Sigma}_{h=12})$: all	65.828	0.991	**0.989**	1.026	1.042	1.001	1.035
$tr(\hat{\Sigma}_{h=12})$: obs.	62.954	0.985	**0.983**	0.998	1.024	0.992	1.008
$tr(\hat{\Sigma}_{h=12})$: rev.	**2.874**	1.113	1.130	1.635	1.454	1.215	1.624

Note: Computed for t=T+1,...,T+N, where T=120 and N=47. Models are estimated with increasing windows of data in the out-of-sample period and with q=14. The entries in the first column of each panel are the mean squared error and the successive entries are ratios to the benchmark values. Benchmark is an AR(1) with RTV data for predicting observations, and a random walk for predicting data revisions. The last three rows present multivariate measures of forecasting accuracy: $tr(\Sigma_h)$. Bold values indicate the most accurate in each row.

Table 2. Comparing forecasts of vintages of inflation at h=1,4,8,12 (vintages at origin; 1995:Q3-2007:Q1; N=47).

	Forecasting vintage t+2|t+1								Forecasting vintage t+5|t+1						
	Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD		Bench.	V-VAR	RV-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD
t+1	0.658	0.975	0.969	**0.963**	0.983	0.978	**0.963**	t+4	1.150	**0.820**	0.846	0.836	0.856	0.835	0.837
t	0.138	0.983	0.982	0.997	0.996	**0.978**	1.033	t+3	**0.880**	1.012	1.037	1.033	1.078	1.014	1.034
t-1	**0.031**	1.248	1.007	1.169	2.068	1.026	1.192	t+2	**0.879**	1.044	1.038	1.000	1.053	1.023	1.017
t-2	**0.057**	1.172	1.011	1.583	1.946	1.014	1.520	t+1	**0.670**	1.054	1.018	1.115	1.211	1.016	1.128
t-3	0.042	1.175	1.010	1.560	1.170	**0.998**	1.536	t	**0.179**	1.061	1.032	1.250	1.435	1.055	1.239
t-4	0.045	1.131	0.865	0.798	0.636	**0.592**	0.796	t-1	0.150	1.146	0.944	1.274	1.446	**0.804**	1.251
t-5	0.037	1.027	1.005	1.828	1.557	**0.937**	1.795	t-2	0.158	1.177	0.888	1.179	1.148	**0.789**	1.160
t-6	**0.034**	1.027	1.009	1.941	1.363	1.032	1.941	t-3	0.142	1.002	0.892	1.311	1.053	**0.842**	1.292
t-7	0.120	**0.958**	0.988	1.067	0.669	1.091	1.055	t-4	0.219	0.912	**0.910**	1.237	0.941	0.964	1.215
t-8	0.038	**0.948**	0.997	1.176	0.950	1.031	1.181	t-5	0.206	**0.844**	1.006	1.367	1.007	1.061	1.354
t-9	0.034	1.020	0.980	1.271	1.370	**0.945**	1.294	t-6	0.222	**0.907**	0.992	1.470	0.901	1.016	1.457
t-10	**0.014**	1.095	1.000	1.708	2.007	1.007	1.438	t-7	0.204	0.899	0.999	1.347	**0.858**	1.017	1.340
t-11	0.012	1.252	0.984	1.163	1.886	**0.976**	1.163	t-8	0.078	**0.856**	1.009	1.482	1.490	0.956	1.464
t-12	**0.020**	1.197	1.015	1.212	1.530	1.025	1.268	t-9	0.062	1.037	0.992	1.394	1.935	**0.937**	1.388
$tr(\hat{\Sigma}_{h=1})$: all	1.280	1.012	**0.977**	1.102	1.082	0.980	1.099	$tr(\hat{\Sigma}_{h=4})$: all	5.197	0.969	0.970	1.085	1.058	**0.958**	1.085
$tr(\hat{\Sigma}_{h=1})$: obs.	0.658	0.975	**0.969**	0.963	0.983	0.978	0.963	$tr(\hat{\Sigma}_{h=4})$: obs.	3.578	0.966	0.972	0.977	1.025	**0.959**	0.984
$tr(\hat{\Sigma}_{h=1})$: rev.	0.622	1.052	0.986	1.248	1.186	**0.983**	1.243	$tr(\hat{\Sigma}_{h=4})$: rev.	1.618	0.975	0.966	1.323	1.129	**0.956**	1.308
	Forecasting vintage t+9|t+1								Forecasting vintage t+13|t+1						
t+8	2.031	**0.744**	0.840	0.795	0.746	0.826	0.802	t+12	2.603	**0.685**	0.792	0.773	0.697	0.772	0.799
t+7	1.731	**0.902**	1.016	0.953	0.904	0.996	0.928	t+11	2.397	**0.763**	0.906	0.855	0.772	0.884	0.871
t+6	1.601	**0.891**	0.940	0.922	**0.891**	0.912	0.897	t+10	2.459	**0.795**	0.908	0.871	0.798	0.889	0.878
t+5	1.305	0.893	0.909	0.894	0.951	0.881	**0.857**	t+9	2.170	**0.813**	0.907	0.877	0.850	0.879	0.874
t+4	1.175	0.943	0.958	0.958	1.051	0.950	**0.918**	t+8	1.969	**0.825**	0.917	0.907	0.900	0.902	0.883
t+3	1.043	1.006	1.003	1.056	1.114	**0.943**	1.056	t+7	1.850	**0.901**	0.986	0.961	0.957	0.951	0.924
t+2	0.972	1.101	1.052	1.056	1.026	**0.992**	1.083	t+6	1.617	0.902	0.924	0.895	0.912	0.871	**0.857**
t+1	0.832	0.994	0.982	0.992	**0.923**	0.932	0.999	t+5	1.354	0.857	0.870	0.854	0.881	0.819	**0.810**
t	0.350	**0.725**	0.780	0.990	0.882	0.863	0.981	t+4	0.946	**0.893**	0.959	0.942	0.945	0.958	0.901
t-1	0.268	0.951	**0.921**	1.289	1.267	0.943	1.268	t+3	0.792	**0.988**	1.048	1.156	1.076	1.004	1.131
t-2	0.345	**0.881**	0.900	1.203	0.957	0.923	1.189	t+2	**0.815**	1.149	1.124	1.229	1.146	1.080	1.229
t-3	0.343	**0.845**	0.904	1.367	1.025	0.942	1.365	t+1	**0.697**	1.082	1.049	1.198	1.146	1.048	1.185
t-4	0.293	**0.894**	0.919	1.292	0.996	0.980	1.278	t	0.414	**0.767**	0.779	1.108	0.987	0.866	1.101
t-5	0.279	0.764	0.995	1.340	1.016	1.031	1.333	t-1	0.330	**0.897**	0.902	1.447	1.449	0.950	1.426
$tr(\hat{\Sigma}_{h=8})$: all	12.567	**0.899**	0.943	0.980	0.940	0.925	0.968	$tr(\hat{\Sigma}_{h=12})$: all	20.412	**0.841**	0.918	0.924	0.884	0.894	0.914
$tr(\hat{\Sigma}_{h=8})$: obs.	10.690	**0.909**	0.951	0.934	0.927	0.921	0.922	$tr(\hat{\Sigma}_{h=12})$: obs.	19.668	**0.842**	0.921	0.911	0.873	0.894	0.901
$tr(\hat{\Sigma}_{h=8})$: rev.	1.877	**0.840**	0.899	1.240	1.014	0.943	1.229	$tr(\hat{\Sigma}_{h=12})$: rev.	0.744	**0.824**	0.833	1.259	1.192	0.903	1.245

Note: See the notes to Table 3. Benchmark is an AR(4) with RTV data for predicting observations, and a random walk for predicting data revisions.

Table 3. Comparing forecasts of vintages focusing only on the three most recent observations at h=1,4,8,12 (vintages at origin: 1995:Q3-2007:Q1; N=47).

Table 3A. Forecasting output growth

	Forecasting Vintage t+2|t+1				Forecasting Vintage t+5|t+1		
	Bench.	V-VAR	GLMS-VAR		Bench.	V-VAR	GLMS-VAR
t+1	3.124	1.031	0.985	t+4	2.938	1.124	1.083
t	0.454	1.086	1.039	t+3	3.566	1.024	1.069
t-1	0.106	1.012	1.075	t+2	3.603	1.014	1.062
$tr(\hat{\Sigma}_{h=1})$	3.683	1.037	0.994	$tr(\hat{\Sigma}_{h=4})$	10.107	1.049	1.070
	For. Vintage t+9|t+1				For. Vintage t+13|t+1		
t+8	3.857	1.075	1.028	t+12	5.968	1.013	1.001
t+7	3.884	1.007	1.049	t+11	7.740	0.998	1.015
t+6	3.797	1.122	1.059	t+10	7.475	1.002	1.021
$tr(\hat{\Sigma}_{h=8})$	11.538	1.068	1.045	$tr(\hat{\Sigma}_{h=12})$	21.183	1.004	1.013

Table 3B. Forecasting inflation

	For. Vintage t+2|t+1				For. Vintage t+5|t+1		
	Bench.	V-VAR	GLMS-VAR		Bench.	V-VAR	GLMS-VAR
t+1	0.658	0.975	1.181	t+4	1.150	0.820	1.105
t	0.138	0.983	1.165	t+3	0.880	1.012	1.190
t-1	0.031	1.248	1.081	t+2	0.879	1.044	1.181
$tr(\hat{\Sigma}_{h=1})$	0.827	0.986	1.175	$tr(\hat{\Sigma}_{h=4})$	2.908	0.946	1.154
	For. Vintage t+9|t+1				For. Vintage t+13|t+1		
t+8	2.031	0.744	1.098	t+12	2.603	0.685	1.188
t+7	1.731	0.902	1.131	t+11	2.397	0.763	1.233
t+6	1.601	0.891	1.093	t+10	2.459	0.795	1.184
$tr(\hat{\Sigma}_{h=8})$	5.363	0.839	1.107	$tr(\hat{\Sigma}_{h=12})$	7.459	0.746	1.201

Note: See the notes to Tables 1 and 2. The V-VAR model is specified with q=14 and autoregressive order p=1. The autoregressive order of the GLMS-VAR model is 2 for output growth and 4 for inflation.

Table 4. A Monte Carlo evaluation of relative forecasting performance with the SV-VAR as the data generating process (500 replications).

Table 4A. SV-VAR estimated with output growth data

	h=1		h = 4		h = 8		h = 12	
	$tr(\hat{\Sigma}_{h=1})$	Ratio	$tr(\hat{\Sigma}_{h=4})$	Ratio	$tr(\hat{\Sigma}_{h=8})$	Ratio	$tr(\hat{\Sigma}_{h=12})$	Ratio
	T+N=120, N=47							
bench	11.11	1.02	54.99	1.03	110.31	1.03	160.21	0.99
V-VAR	11.39	1.05	54.21	1.02	107.80	1.01	160.70	0.99
RV-VAR	11.44	1.05	54.80	1.03	109.46	1.02	163.59	1.01
SV-VAR	10.89	--	53.30	--	106.89	--	161.75	--
	T+N=500, N=47							
bench	10.91	1.25	53.65	1.23	106.77	1.19	154.31	1.12
V-VAR	9.75	1.12	46.65	1.07	94.12	1.05	141.57	1.03
RV-VAR	10.04	1.15	47.74	1.09	95.81	1.07	142.99	1.04
SV-VAR	8.73	--	43.66	--	89.60	--	137.74	--
Forecasting only the last three observations of each vintage								
	T+N=120, N=47							
bench	8.64	0.97	35.70	1.01	37.01	1.04	36.38	0.94
V-VAR	8.86	0.99	34.83	0.99	35.26	0.99	37.79	0.98
RV-VAR	8.97	1.01	35.25	1.00	35.69	1.00	38.58	1.00
GLMS-VAR	9.31	1.04	36.70	1.04	37.09	1.04	36.23	0.94
SV-VAR	8.91	--	35.24	-	35.73	--	38.53	--
	T+N=500, N=47							
bench	8.48	1.13	34.58	1.16	35.48	1.14	35.21	1.04
V-VAR	7.58	1.01	29.84	1.00	31.23	1.00	34.03	1.00
RV-VAR	7.62	1.01	29.96	1.01	31.37	1.01	34.13	1.01
GLMS-VAR	8.42	1.12	34.58	1.16	35.36	1.14	35.07	1.03
SV-VAR	7.52	--	29.76	--	31.09	--	33.93	--

Table 4B. SV-VAR estimated with inflation data

	h=1		h=4		h=8		h=12	
	$tr(\hat{\Sigma}_{\mathrm{h=1}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=4}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=4}})$	Ratio
	T+N=120, N=47							
bench	3.03	1.14	13.60	1.08	31.89	0.96	51.02	0.89
V-VAR	2.92	1.10	13.32	1.06	33.21	1.00	55.27	0.96
RV-VAR	2.98	1.13	13.78	1.09	34.56	1.04	57.57	1.00
SV-VAR	2.64	--	12.62	--	33.15	--	57.45	--
	T+N=500, N=47							
bench	2.98	1.45	13.51	1.33	32.02	1.21	51.07	1.14
V-VAR	2.53	1.24	11.59	1.15	28.30	1.07	46.35	1.03
RV-VAR	2.73	1.33	12.28	1.21	29.19	1.10	47.02	1.05
SV-VAR	2.05	--	10.12	--	26.54	--	44.96	--
Forecasting only the last three observations of each vintage								
	T+N=120, N=47							
bench	2.05	1.03	8.10	0.99	13.59	0.88	15.92	0.86
V-VAR	1.98	1.00	8.15	0.99	15.18	0.98	18.15	0.98
RV-VAR	2.01	1.01	8.34	1.01	15.65	1.01	18.63	1.00
GLMS-VAR	2.37	1.19	8.71	1.06	14.36	0.93	16.69	0.90
SV-VAR	1.99	-	8.22	-	15.43	--	18.56	--
	T+N=500, N=47							
bench	2.00	1.17	8.02	1.15	13.78	1.10	15.89	1.07
V-VAR	1.73	1.01	7.04	1.01	12.62	1.01	14.85	1.00
RV-VAR	1.75	1.02	7.10	1.02	12.68	1.01	14.88	1.01
GLMS-VAR	1.99	1.17	7.89	1.13	13.69	1.10	15.90	1.08
SV-VAR	1.71	--	6.95	--	12.50	--	14.79	--

Table 4C. Forecasting evaluation conditional on whether the next vintage is subject to an annual revision.

	h=1: SV-VAR estimated with output growth data						h=1: SV-VAR estimated with inflation data					
	$D_1^{t+1} = 1\ or = 0$		$D_1^{t+1} = 0$		$D_1^{t+1} = 1$		$D_1^{t+1} = 1\ or = 0$		$D_1^{t+1} = 0$		$D_1^{t+1} = 1$	
	$tr(\hat{\Sigma}_{\mathrm{h=1}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1,D_1^{T+1}}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1,D_1^{T+1}}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1,D_1^{T+1}}})$	Ratio	$tr(\hat{\Sigma}_{\mathrm{h=1,D_1^{T+1}}})$	Ratio
	T+N=120, N=47											
bench	11.11	1.02	9.71	0.97	15.18	1.12	3.03	1.14	2.32	1.02	5.08	1.37
V-VAR	11.39	1.05	10.68	1.07	13.46	0.99	2.92	1.10	2.61	1.14	3.79	1.03
RV-VAR	11.44	1.05	10.05	1.01	15.48	1.14	2.98	1.13	2.32	1.01	4.93	1.33
SV-VAR	10.89	--	9.97	--	13.57	--	2.64	--	2.28	--	3.7	-
	T+N=500, N=47											
bench	10.91	1.25	9.52	1.10	14.98	1.68	2.98	1.45	2.23	1.09	5.04	2.45
V-VAR	9.75	1.12	9.78	1.13	11.71	1.31	2.53	1.24	2.25	1.10	3.35	1.63
RV-VAR	10.04	1.15	8.72	1.01	13.89	1.55	2.73	1.33	2.07	1.01	4.65	2.26
SV-VAR	8.73	--	8.66	--	8.94	--	2.05	--	2.05	-	2.06	--

Note: The vintage-based VARs are specified with q=14. The GLMS-VAR model uses four maturities and is framed in levels, with an autoregressive order of 2 when fitted to data generated from the SV-VAR estimated on output growth data, and 4 when the SV-VAR is fitted to inflation. In Table 4.C, $D_1^{t+1} = 1$ in 12 out of 47 data vintages in the out-of-sample period (for t=T+1,...,T+N)

Table 5. Comparing forecasts of future observations using first and second releases and the latest available vintage (2010:Q1) to compute forecast errors. (Data vintages at origin: 1995:Q3-2006:Q3, N=45).

Forecasting Future Output Growth									
	RTV	EOS	V-VAR	R-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD	GLMS-VAR
Forecasting First-Release Data (y_{t+h}^{t+h+1})									
t+4	1.814	0.966	1.101	1.081	1.099	1.103	1.096	1.111	0.999
t+3	1.785	0.974	1.102	1.084	1.098	1.105	1.098	1.108	1.031
t+2	1.762	0.961	1.107	1.092	1.116	1.112	1.105	1.125	1.027
t+1	1.780	0.952*	1.050	1.046	1.055	1.054	1.052	1.057	1.002
Forecasting Second-Release Data (y_{t+h}^{t+h+2})									
t+4	1.980	0.983	1.062	1.044	1.054	1.065	1.059	1.064	0.997
t+3	1.954	0.989	1.060	1.044	1.053	1.067	1.058	1.061	1.019
t+2	1.927	0.979	1.053	1.042	1.055	1.059	1.053	1.062	0.998
t+1	1.899	0.974	1.011	1.002	1.009	1.012	1.008	1.011	0.991
Forecasting the Latest Available Vintage ($y_{t+h}^{2010:Q1}$)									
t+4	2.152	1.020	1.024	1.021	1.146	1.151	1.055	1.159	1.022
t+3	2.119	1.025	1.022	1.028	1.168	1.151	1.062	1.182	1.022
t+2	2.113	1.015	1.012	1.024	1.133	1.130	1.061	1.142	1.052
t+1	2.080	1.000	0.987*	0.996	1.127	1.105	1.030	1.131	1.046
Forecasting Future Inflation									
	RTV	EOS	V-VAR	R-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD	GLMS-VAR
Forecasting First-Release Data (y_{t+h}^{t+h+1})									
t+4	1.019	1.009	0.951**	0.954**	0.954**	0.970**	0.950**	0.941**	1.110
t+3	0.953	1.004	1.009	1.010	1.003	1.022	0.999	0.995	1.113
t+2	0.891	1.011	1.039	1.034	1.034	1.040	1.026	1.032	1.099
t+1	0.773	1.032	1.002	0.997*	1.001	1.013	1.000	0.997	1.098
Forecasting Second-Release Data (y_{t+h}^{t+h+2})									
t+4	1.041	0.967*	0.953***	0.965**	0.965**	0.979**	0.959**	0.933**	1.082
t+3	0.956	0.967*	1.005	1.016	1.003	1.027	1.001	0.980	1.100
t+2	0.888	0.977	1.035	1.036	1.024	1.034	1.024	1.020	1.109
t+1	0.780	1.004	0.989*	0.987*	0.986*	0.995*	0.985*	0.987*	1.159
Forecasting the Latest Available Vintage ($y_{t+h}^{2010:Q1}$)									
t+4	1.064	0.991	1.000	1.019	1.204	1.164	1.045	1.210	1.044
t+3	0.976	0.986	1.029	1.050	1.124	1.143	1.049	1.134	1.055
t+2	0.926	1.009	1.057	1.056	1.096	1.127	1.053	1.100	1.100
t+1	0.791	1.036	1.037	1.032	1.243	1.217	1.051	1.255	1.147

Note: All models are estimated with rolling windows of 120 observations, and out-of-sample period has N=45 observations. RTV is the benchmark model with RMSFEs entries in the first column. RTV forecasts are bias-corrected and the correction depends on the target vintage (with no correction for predicting first-release data). All other columns are ratios to the benchmark RMSFE. With vintage-based VAR models, first-release forecasts are computed using the first equation: $\hat{y}_{t+h}^{t+h+1t+1}$ for h=1,2,3,4. Forecasts of second releases are $\hat{y}_{t+h-1}^{t+h+1|t+1}$ for h=2,...,5. Forecasts of the latest available vintage are computed using the last equation: $\hat{y}_{t+h-13}^{t+h+1|t+1}$ for h=14,15,16 and 17. Forecasts from GLSM-VAR model are for first-releases are computed using $400(\hat{Y}_{t+h}^{t+h+1|t+1}$-$\hat{Y}_{t+h-1}^{t+h+1|t+1})$ for h=1,2,3,4. For second releases and the latest available vintage, they are computed using $400(\hat{Y}_{t+h-1}^{t+h+1|t+1}$-$\hat{Y}_{t+h-2}^{t+h+1|t+1})$ for h=2,...,5. For MSFE ratios that favour the multiple-vintage model relative to the benchmark model, we indicate the significance level of the rejection of the null of equal forecast accuracy using the Clark and West (2007) test (one-sided) by * for 10%, ** for 5%, and *** for the 1% level.

Table 6. Comparing forecasts of data revisions using post-revision data (after 13 revisions) and the latest available vintage (2010:Q1) to compute forecast errors. (Data vintages at origin: 1995:Q3-2007:Q1, N=45).

Forecasting Data Releases of Output Growth								
	RW	V-VAR	R-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD	GLMS-VAR
Forecasting post-revision data (after 13 revisions)								
t	1.352	1.070	1.070	1.404	1.321	1.129	1.398	1.013
t-1	1.152	1.003	1.035	1.417	1.357	1.104	1.408	0.990
t-2	1.156	1.008	1.040	1.389	1.278	1.109	1.387	
t-3	1.185	1.027	1.030	1.445	1.346	1.089	1.443	
Forecasting the Latest Available Vintage								
t	1.613	1.026	1.027	1.314	1.224	1.092	1.314	1.012
t-1	1.461	0.963***	1.010	1.297	1.196	1.082	1.297	1.001
t-2	1.467	0.966***	1.011	1.277	1.147	1.085	1.280	
t-3	1.449	0.981*	1.004	1.329	1.184	1.077	1.332	
Forecasting Data Releases of Inflation								
	RW	V-VAR	R-VAR	SV-VAR	SBV-VAR	RV-VAR+SBD	SV-VAR+BD	GLMS-VAR
Forecasting Post-revision data (after 13 revisions)								
t	0.643	0.939*	0.912***	1.143	1.211	0.979**	1.168	1.060
t-1	0.559	1.008	0.973***	1.346	1.400	1.010	1.356	1.032
t-2	0.601	0.944	0.950**	1.232	1.233	0.986	1.231	
t-3	0.622	0.922*	0.954***	1.219	1.099	0.981*	1.209	
Forecasting the Latest Available Vintage								
t	0.614	0.915**	0.932***	1.174	1.221	1.031	1.205	1.052
t-1	0.545	0.940*	0.980*	1.353	1.318	1.054	1.379	1.034
t-2	0.568	0.900**	0.957***	1.246	1.248	1.035	1.254	
t-3	0.587	0.878**	0.959**	1.268	1.158	1.012	1.255	

Note: All models are estimated with rolling windows of 120 observations. Benchmark is a random walk with RMSFE described in the first column. The remaining columns are ratios to the benchmark. Forecasts from the models are the same for both targets (after 13 revisions or latest available vintage). With V-VAR models, forecasts are computed using the last equation: $\hat{y}_{t+h-13}^{t+h+1|t+1}$ for *h = 10, 11, 12, 13*. With GSLM-VAR models, forecasts are computed using $400(\hat{Y}_{t+h-2}^{t+h+1|t+1}$-$\hat{Y}_{t+h-3}^{t+h+1|t+1})$ where *h = 1* provides the prediction of *t-1* observation, and $400(\hat{Y}_{t+h-1}^{t+h+1|t+1}$-$\hat{Y}_{t+h-2}^{t+h+1|t+1})$ with *h = 2* provides the prediction of period t. (Forecasts for t-2 and t-3 from this model are not defined). For MSFE ratios that favour the multiple-vintage model relative to the benchmark model, we indicate the significance level of the rejection of the null of equal forecast accuracy using the Clark and West (2007) test (one-sided) by * for 10%, ** for 5%, and *** for the 1% level.