

# *Data revisions and real-time forecasting*

Article

Accepted Version

Clements, M. P. ORCID: <https://orcid.org/0000-0001-6329-1341> and Galvão, A. B. (2019) Data revisions and real-time forecasting. Oxford Research Encyclopedia of Economics and Finance. doi:  
<https://doi.org/10.1093/acrefore/9780190625979.013.248>  
Available at <https://centaur.reading.ac.uk/78512/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI:

<http://dx.doi.org/10.1093/acrefore/9780190625979.013.248>

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Data Revisions and Real-Time Forecasting

Michael P. Clements\*

Ana Beatriz Galvão

ICMA Centre

Warwick Business School

Henley Business School

University of Warwick

University of Reading

ana.galvao@wbs.ac.uk

m.p.clements@reading.ac.uk

August 7, 2018

## Summary

At a given point in time, a forecaster will have access to data on macroeconomic variables that have been subject to different numbers of rounds of revisions, leading to varying degrees of data maturity. Observations referring to the very recent past will be first-release data, or data which has as yet been revised only a few times. Observations referring to a decade ago will typically have been subject to many rounds of revisions. How should the forecaster use the data to generate forecasts of the future? The conventional approach would be to estimate the forecasting model using the latest vintage of data available at that time, implicitly ignoring the differences in data maturity across observations. In this article we draw on recent research on real-time forecasting to consider whether there are better ways of forecasting.

---

\*Prepared for the Oxford Research Encyclopedia of Economics and Finance

The conventional approach for real-time forecasting treats the data as given, that is, it ignores the fact that it will be revised. In some cases, the costs of this approach are point predictions and assessments of forecasting uncertainty that are less accurate than a number of approaches to forecasting which explicitly allow for data revisions. There are a number of ways to ‘allow for data revisions’, including modelling the data revisions explicitly, an agnostic or reduced-form approach, and using only largely unrevised data. We survey these methods, and provide guidance on choosing one, depending on whether the aim is to forecast an earlier release or the fully-revised values.

JEL code: C53.

Keywords: data revisions, news, noise, unobserved components, vector autoregressions, point predictions, density forecasts.

# 1 Introduction

The majority of the macroeconomic time series of interest to policy makers and forecasters are subject to data revisions. Monetary policy decisions require accurate measurement of the current state of the economy, and also the likely future paths for inflation and economic activity. Fiscal policy decisions also rely on current and predicted values for economic activity to enable the calculation of the size of future government deficits. Data revisions to key economic activity aggregates, such as GDP, add an extra layer of uncertainty to the measurement of current and future economic conditions, and to the conduct of economic policy.

Both forecasting and policy analysis have to be undertaken in real time. We look at different approaches to forecasting in real time, when the available information set is restricted to only the vintages of data that were available at each historical period. With the advent of readily-available real-time data sets of key macroeconomic indicators (as, for example, Croushore and Stark (2001)), we are able to look back and gauge how well different approaches would have fared over the past 50 years or so, to serve as a guide to what might be expected to work for forecasters today. As an example, by forecasting data revisions we are able to improve the reliability of output gap estimates in real time (see, e.g., Garratt, Lee, Mise and Shields (2009, 2008) and Clements and Galvão (2012)).

This article considers the implications of the nature of the estimates of macroeconomic data published by national statistics agencies, and the subsequent revisions to these estimates, for forecasting the future values of those variables. Excellent general surveys on data revisions and real-time analysis, that go beyond our emphasis on forecasting, are provided by Croushore (2006, 2011b, 2011a).

We begin in section 2 by explaining why statistical offices revise macroeconomic data,

including details of the nature and timing of their updates to initial estimates (i.e., data revisions) for the US, and for other OECD countries.

One approach to forecasting when data are subject to revision is to attempt to model the behaviour of the statistics office (henceforth, SO), and incorporate this into the forecasting model. Section 3 discusses some simple models of the behaviour of the SO in terms of its processing of source data to generate initial estimates of macroeconomic variables, and how these models lead to published data inheriting the properties of news or noise revisions. Initially, we assume a very simple revisions process, namely that the second estimate (equivalently, first revision) reveals the truth. Later on, we describe the practical implementation of the approach suggested by Kishor and Koenig (2012) for data subject to multiple revisions. This approach requires a model for the true process and for data revisions, and the application of the Kalman Filter. Section 4 considers data revisions in the context of forecasting with Dynamic Stochastic General Equilibrium (DSGE) models. DSGE models are typically the Central Banker's model of choice, especially for policy analysis, but increasingly for forecasting too.

The data revisions modelling approaches in sections 3 and 4 rely on state-space models and a set of unobserved components. In Section 5 we survey approaches for modelling data revisions and forecasting that do not require the estimation of unobserved components. Section 6 considers simple single-equation approaches to forecasting, and includes the 'traditional approach', which amounts to effectively ignoring revisions. The scope of our survey is also widened to consider density forecasts in addition to point forecasts.

With the aim of providing guidance as to which modelling approach to choose when forecasting with data subject to revision, section 7 reviews some of the empirical evidence on the impact of data revisions on forecasting performance, and on the relative performance of some of the approaches discussed in this article. Section 8 offers some

concluding remarks.

## 2 Why data revisions?

Why are macroeconomic data revised? This can perhaps be best understood with reference to the US Bureau of Economic Analysis (BEA) timetable for data releases.<sup>1</sup> The BEA publishes its first or ‘advance’ estimates of quarterly national accounts data about a month after the quarter in question. These estimates are necessarily based on only partial source data, and are subject to revision as more complete data becomes available. They are then revised twice more at monthly intervals. These two monthly revisions were formerly known as the ‘preliminary’ and ‘final’ estimates, but are now simply the second and third estimates. As described by Landefeld, Seskin and Fraumeni (2008), 25% of the GDP components at the time of the release of the first estimate are trend-based data obtained from extrapolations supported by related indicator series. The proportion of trend-based data in the second and third estimates is 23% and 13% respectively. Hence these revisions typically reflect the availability of more complete source data. The series are then subject to three annual rounds of revisions (in the third quarters of the year)<sup>2</sup> to 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.<sup>3</sup>

In terms of the question we posed - why data revisions? - the need for *timely* estimates of recent economic developments necessitates that estimates are produced long before all

---

<sup>1</sup>We focus on real GDP, given its importance, and its widespread use in the literature on real-time forecasting. We also primarily consider the US, although similar considerations apply for other countries.

<sup>2</sup>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 *et al.* (2008).

<sup>3</sup>In the case of the US, the Bureau of Economic Analysis provides descriptions of the methodologies employed at: [http://www.bea.gov/methodologies/index.htm#national\\_meth](http://www.bea.gov/methodologies/index.htm#national_meth)

Table 1: Real GDP data vintages - Snapshot

DATE	13Q1	13Q2	13Q3	13Q4	14Q1	14Q2	14Q3	14Q4	15Q1
47:Q1	1770.7	1770.7	1932.6	1932.6	1932.6	1932.6	1934.5	1934.5	1934.5
47:Q2	1768.0	1768.0	1930.4	1930.4	1930.4	1930.4	1932.3	1932.3	1932.3
47:Q3	1766.5	1766.5	1928.4	1928.4	1928.4	1928.4	1930.3	1930.3	1930.3
47:Q4	1793.3	1793.3	1958.8	1958.8	1958.8	1958.8	1960.7	1960.7	1960.7
48:Q1	1821.8	1821.8	1987.6	1987.6	1987.6	1987.6	1989.5	1989.5	1989.5
48:Q2	1855.3	1855.3	2019.9	2019.9	2019.9	2019.9	2021.9	2021.9	2021.9
48:Q3	1865.3	1865.3	2031.2	2031.2	2031.2	2031.2	2033.2	2033.2	2033.2
48:Q4	1868.2	1868.2	2033.3	2033.3	2033.3	2033.3	2035.3	2035.3	2035.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12:Q4	13647.6	13665.4	15539.6	15539.6	15539.6	15539.6	15433.7	15433.7	15433.7
13:Q1	#N/A	13750.1	15583.9	15583.9	15583.9	15583.9	15538.4	15538.4	15538.4
13:Q2	#N/A	#N/A	15648.7	15679.7	15679.7	15679.7	15606.6	15606.6	15606.6
13:Q3	#N/A	#N/A	#N/A	15790.1	15839.3	15839.3	15779.9	15779.9	15779.9
13:Q4	#N/A	#N/A	#N/A	#N/A	15965.6	15942.3	15916.2	15916.2	15916.2
14:Q1	#N/A	#N/A	#N/A	#N/A	#N/A	15946.6	15831.7	15831.7	15831.7
14:Q2	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	15985.7	16010.4	16010.4
14:Q3	#N/A	16150.6	16205.6						
14:Q4	#N/A	16311.6							

the data have been collected. As indicated, data collection and refinement of official statistics ought perhaps be viewed as an ongoing process.

Consider now the quarterly vintages recorded in the Real Time Data Set for Macroeconomists (RTDSM: see Croushore and Stark (2001)). Table 1 illustrates a an excerpt of the real-time dataset for for US real GDP.<sup>4</sup> This resource has greatly facilitated real-time data analysis. The first available estimate for any quarter in the RTDSM is the ‘advance’ estimate, and this is denoted by  $y_t^{t+1}$ . The second quarterly estimate for quarter  $t$  is denoted by  $y_t^{t+2}$ , and so on. The subscript indicates the reference quarter, and the superscript the quarterly vintage (or estimate). In the case of US BEA data, the data will typically then remain unchanged unless the  $t + 3$ -vintage is an annual revision

---

<sup>4</sup>Taken from the Federal Reserve Bank of Philadelphia webpage:  
<https://www.philadelphiafed.org/research-and-data/real-time-center/>.  
 See Croushore and Stark (2001).

( $t + 3 \in Q3$ ) or a benchmark revision. In practice, a small modification to this simple seasonal pattern may occur. When benchmark revisions are anticipated to be published in January, annual revisions may not be published in the previous July.<sup>5</sup>

The table depicts only the vintages from 2013Q1 to 2015Q1 - these are the columns. Quarterly vintages for real GDP are available back to 1965. The first period for which there is data is 1947Q1, and for each vintage, data are available up to one quarter before the vintage, reflecting the delay of one quarter in publishing data (when the vintages are recorded at the quarterly frequency, as here). Note that we drop the observations from 1949Q1 to 2012Q3, inclusive, to save space. To illustrate the patterns of revision described in the preceding paragraph, consider the 2013Q4 vintage estimate of the 2013Q3 observation. This is the first estimate of reference quarter 2013Q3, and is 15790.1. In 2014Q1, this estimate is revised up to 15839.3 (a 0.3% increase), in 2014Q2 the estimate remains unaltered, and is then revised down to 15779.9 in the 2014Q3 annual revision.

Our focus will be on quarterly revisions to quarterly data, primarily to keep the notation simple. But as shown by Clements and Galvão (2017), monthly vintages can also be modelled and forecasted. It is generally the case that the models we discuss below are of the growth rates (or differences of logs) of the variable (such as GDP), and the revisions are then the differences between different vintage estimates of the growth rates. This means that we are dealing with variables that are integrated of order zero, assuming that (the log of) real GDP is integrated of order one. An exception arises when we discuss the model of Garratt *et al.* (2008) in section 5.1.

We have described the institutional nature of the revisions to US national accounts data. The task of providing timely estimates based on incomplete source data is com-

---

<sup>5</sup>For example, there are 8 benchmark revisions in the data vintages from 1965Q3 up 2010Q1. 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 - the 8 benchmark revisions and the 36 annual revisions.

mon to all government statistics offices, so that data revisions are not specific to the US. Zwijsenburg (2015) analyses the revisions to national accounts data in 18 OECD countries, in terms of regular revisions, and benchmark revisions. Regular revisions are as described earlier, and are normally the focus of interest as they are likely to be more amenable to modelling. These revisions result from the updating of the dataset used to compute the earlier estimates, and are present in all 18 OECD countries. The mean revisions for most countries are not statistically significantly different from zero - data revisions normally have no effect on the unconditional mean of quarterly growth rates. However, Zwijsenburg (2015) shows that revisions can be sizeable, with a mean absolute error of 0.18 (the average across countries for the observations for the 1993-2014 period) for revisions made up to 5 months, and of 0.40 for revisions made up to 3 years.

### **3 Forecasting methods and a model of the behaviour of the Statistical Office**

As mentioned earlier, regular data revisions can be related to the operational behaviour of the government statistical office (SO). A model of the behaviour of the statistical office is provided by Kishor and Koenig (2012), drawing on earlier contributions by Howrey (1978) and Sargent (1989). They provide models of the behaviour of the SO which directly suggest ways of forecasting the true values of the series. Before considering models of SO behaviour, it will be useful to make a distinction between news and noise revisions - section 3.1. Sections 3.2 and 3.3 then provide a simple model of behaviour, and some extensions, before showing how this approach may be used in practice in section 3.4.

### 3.1 News and Noise data revisions

Following Mankiw and Shapiro (1986), data revisions are sometimes characterized as news or noise. Data revisions are news when they add *new* information, and noise when they reduce measurement error. If data revisions are noise, they can be predicted based on the current estimate. Mankiw and Shapiro (1986) and Faust, Rogers and Wright (2005) provide empirical evidence that data revisions to US real GDP are largely news. Aruoba (2008) and Corradi, Fernandez and Swanson (2009) provide recent extensions to testing for the properties of data revisions, which may lead to more nuanced findings, but the broad classification of each series as being subject to news or noise revisions in the simple setting below is a good starting point.

The standard tests for news and noise regress a revision on either the earlier estimate or on the later estimate. For example, consider the revision between the first-estimates  $y_t^{t+1}$ , and the data available some three and a half years later  $y_t^{t+14}$  (chosen to include the three rounds of annual revisions). One can test for news and noise revisions using, respectively:

$$y_t^{t+14} - y_t^{t+1} = \alpha + \beta_{ne} y_t^{t+1} + \omega_t$$

and:

$$y_t^{t+14} - y_t^{t+1} = \alpha + \beta_{no} y_t^{t+14} + \omega_t.$$

If we reject  $\beta_{ne} = 0$ , then we reject the null hypothesis that data revisions are news, because we are able to predict revisions from knowledge of the earlier estimate,  $y_t^{t+1}$ . If we reject  $\beta_{no} = 0$ , then we reject the null hypothesis that data revisions are noise, because they are correlated with the final estimate,  $y_t^{t+14}$ . If we reject both null hypotheses, then data revisions are neither news or noise. If we fail to reject both hypotheses, we conclude that the data are not informative regarding the news /noise dichotomy.

Of course, the properties of revisions may depend on the choice of the earlier and later estimates, may vary over time, and as noted above, need not give a definitive classification. The definition of efficiency underlying news is in terms of an information set consisting only of the initial estimate, leaving open the possibility that revisions may be predictable using initial estimates of other related variables (or other sources of information, as in Clements and Galvão (2017)) so that the property of revisions being news, narrowly defined, may be less useful in multivariate settings.

To illustrate, using the standard tests, Clements (2017, Table 2, p. 427) compares the first estimates to the 15th estimates for the observations 1970Q2 to 2007Q2 for a number of variables. He finds that 9 of the 25 variables analyzed have data revisions that appear to be news, in that  $\beta_{ne} = 0$  is not rejected, but  $\beta_{no} = 0$  is rejected. 7 of the 25 variables are found to have noise revisions ( $\beta_{ne} = 0$  is rejected, but  $\beta_{no} = 0$  is not rejected).

Rather than considering the revisions between an early estimate and a later estimate (such as the preliminary and fully-revised data), Swanson and van Dijk (2006) consider the entire revision history, and identify the point at which revisions become ‘rational’ (or unpredictable). They also consider the properties of revisions separately in expansions and contractions.

### 3.2 Model of SO behaviour

Suppose that  $\tilde{y}_t$  is the true period  $t$  value of the quarterly growth rate of a macroeconomic variable subject to revisions and that it follows a simple AR(1):

$$\tilde{y}_t = f\tilde{y}_{t-1} + v_t. \tag{1}$$

There is however a delay in the publication of the true/revised values  $\tilde{y}_t$ , such that we assume it is not observed or published until  $t + 2$ , i.e.,  $y_t^{t+2} = \tilde{y}_t$ . We relax this assumption when implementing the Kishor and Koenig (2012) modelling approach in practice in section 3.4.

In period  $t + 1$ , the SO observes source data  $w_t^{t+1} = \tilde{y}_t + \eta_t$ , where  $\eta_t$  is iid $(0, \sigma_\eta^2)$ . Based on  $w_t^{t+1}$  (and possibly  $\tilde{y}_{t-1}, \tilde{y}_{t-2}, \dots; w_{t-1}^t, \dots$ ), the SO produces its first estimate of  $y_t$ , denoted  $y_t^{t+1}$ . Private forecasters do not observe  $w_t^{t+1}$ , unless the SO sets  $y_t^{t+1} = w_t^{t+1}$ .

From the perspective of a professional forecaster (PF) using only data published by the SO, the goal is to forecast future true values  $\tilde{y}_T, \tilde{y}_{T+1}, \tilde{y}_{T+2}, \dots$  based on period  $T + 1$  information,  $y_T^{T+1}, y_{T-1}^{T+1}$  ( $= \tilde{y}_{T-1}$ ), etc. Note that the superscripts denote when the value is published, while the subscripts refer to the reference quarter - i.e., the time period to which the data refers.<sup>6</sup>

A possible strategy, considered by the PF, consists of the following stages:

1. Estimate the model in equation (1) over  $t = 2$  to  $T - 1$  using  $\{\tilde{y}_t\}$ . That is, the model is estimated solely on the true data, and because of the publication delay assumed above, the estimation sample runs only to  $T - 1$  (as the observation for reference period  $T$  is a first estimate).

2. Obtain an estimate of  $\hat{y}_T$ . This is an estimate of the true value of the variable at period  $T$ , and how this should be done depends on how the SO chooses  $y_T^{T+1}$ , as discussed below.

3. Then compute forecasts as  $\tilde{y}_{T+h|T} = \hat{f}^h \hat{y}_T$  for  $h = 1, 2, \dots$

This strategy avoids what Kishor and Koenig (2012) refer to as the mixing of ‘apples and oranges’. They state that:

---

<sup>6</sup>Throughout,  $T + 1$  will be used to denote the forecast-origin-vintage. Hence the latest-vintage values of data the forecaster will have access to are  $\{\dots, y_{T-1}^{T+1}, y_T^{T+1}\}$ .

‘The problem with conventional practice is that it mixes apples and oranges. Data toward the tail end of the sample (oranges) have undergone little or no revision. Data early in the sample (apples) are heavily revised. For most series and typical sample sizes, the heavily revised data dominate estimation. Consequently, the VAR approximates the dynamic relationship between apples and apples. However, the data that are substituted in to the VAR equations to generate a forecast are end-of-sample oranges. Essentially, conventional practice constructs a cider press and then feeds oranges into it, expecting somehow to get cider’. [Kishor and Koenig (2012)]

The strategy outlined above estimates the model on fully-revised data (or ‘apples’) and also conditions the forecast on (an estimate of) a fully-revised data point. The conventional approach estimates the model on  $\{y_t^{T+1}\}$  for  $t = 2, \dots, T$ , so that all but the last observation is fully-revised, and then sets  $\hat{y}_T = y_T^{T+1}$ , i.e., conditions on a first estimate, so falling foul of the dictum not to mix apples and oranges.

For step (2) we need an estimate of  $\hat{y}_T$ . Suppose the SO simply sets  $y_T^{T+1} = w_T^{T+1}$ , so that:

$$y_T^{T+1} = \tilde{y}_T + \eta_T = f\tilde{y}_{T-1} + v_T + \eta_T,$$

and the PF assumes this is the case. That is, the SO makes no attempt to process the source data in order to generate an efficient first estimate, and the subsequent revision is noise, as defined in section 3.1. To see this, note that:

$$Cov(\tilde{y}_t, \tilde{y}_t - y_t^{t+1}) = 0,$$

but that:

$$Cov(y_t^{t+1}, \tilde{y}_t - y_t^{t+1}) = -\sigma_\eta^2 \neq 0,$$

implying that the revision  $\tilde{y}_t - y_t^{t+1}$  is uncorrelated with the true value, but predictable from  $y_t^{t+1}$ .

Suppose now that the SO does actively process the source data. Intuitively, the release of a high value  $y_T^{T+1}$  could be because of a large positive  $\eta_T$  (noise) or because of a high positive  $v_T$  (i.e., signalling a high true value  $\tilde{y}_T$ ), and some weight ought to be attributed to both of these possibilities. If  $v_t$  and  $\eta_t$  are normally distributed, then the standard signal extraction problem gives the optimal estimate of  $\tilde{y}_T$  as:

$$E(\tilde{y}_T | y_T^{T+1}, y_{T-1}^T, \tilde{y}_{T-1}, \tilde{y}_{T-2}, \dots) = f\tilde{y}_{T-1} + \gamma_N (y_T^{T+1} - f\tilde{y}_{T-1}) \quad (2)$$

where  $\gamma_N = \sigma_v^2 / (\sigma_v^2 + \sigma_\eta^2)$ , and  $\gamma_N$  is the weight on the first estimate,  $y_T^{T+1}$ , and  $(1 - \gamma_N)$  is the weight on the model predicted value,  $f\tilde{y}_{T-1}$ . When normality does not hold,  $\gamma_N$  still has the justification of minimizing:

$$E \left[ (\tilde{y}_T - [f\tilde{y}_{T-1} + \gamma (y_T^{T+1} - f\tilde{y}_{T-1})])^2 \right]$$

over  $\gamma$ .

Setting  $\hat{y}_T$  equal to the right-hand-side of (2) results in an efficient estimate, in the sense that the induced revision  $\tilde{y}_T - \hat{y}_T$  is uncorrelated with  $\hat{y}_T$ ;  $Cov(\hat{y}_T, \tilde{y}_T - \hat{y}_T) = 0$ .

Hence the release of the unprocessed source data by the SO leads to inefficient initial estimates: subsequent revisions are predictable. In our setup, defining  $\hat{y}_T$  by (2), in the second step of the three-stage forecasting strategy set out above, would result in unpredictable revisions.

### 3.3 Extensions

A number of extensions and generalizations are possible. For example, Howrey (1978) allows revisions to be serially correlated:

$$y_t^{t+1} - \tilde{y}_t = h(y_{t-1}^t - \tilde{y}_{t-1}) + w_t \quad (3)$$

and then the estimator of  $\tilde{y}_T$  is given by:

$$E(\tilde{y}_T | y_T^{T+1}, y_{T-1}^T, ; \tilde{y}_{T-1}, \tilde{y}_{T-2}, \dots) = f\tilde{y}_{T-1} + \gamma_H (y_T^{T+1} - h(y_{T-1}^T - \tilde{y}_{T-1}) - f\tilde{y}_{T-1})$$

with  $\gamma_H = \sigma_v^2 / (\sigma_v^2 + \sigma_w^2)$ .

Suppose instead, as suggested by Sargent (1989), that the SO does not announce  $y_t^{t+1} = w_t^{t+1}$ , but it filters the source data itself, and in so doing inadvertently introduces an additive random error, so that the first announcement  $y_t^{t+1}$  is given by:

$$y_t^{t+1} = f\tilde{y}_{t-1} + g(w_t^{t+1} - f\tilde{y}_{t-1}) + \xi_t,$$

where  $g = \gamma_N$  and  $\xi_t$  is the idiosyncratic error induced by the filtering.

From the perspective of the PF, who now does not observe  $w_t^{t+1}$ , we obtain:

$$\begin{aligned} y_t^{t+1} &= f\tilde{y}_{t-1} + g(\tilde{y}_t - f\tilde{y}_{t-1}) + \xi_t + g\eta_t \\ &= f\tilde{y}_{t-1} + g(\tilde{y}_t - f\tilde{y}_{t-1}) + \varepsilon_t \\ &= f\tilde{y}_{t-1} + gv_t + \varepsilon_t, \end{aligned}$$

setting  $\varepsilon_t = \xi_t + g\eta_t$ .

This suggests the optimal forecast of  $\tilde{y}_T$  is given by:

$$E(\tilde{y}_T | y_T^{T+1}, y_{T-1}^T, \tilde{y}_{T-1}, \tilde{y}_{T-2}, \dots) = f\tilde{y}_{T-1} + \gamma_S (y_T^{T+1} - f\tilde{y}_{T-1}) \quad (4)$$

where  $\gamma_S = g\sigma_v^2 / (g^2\sigma_v^2 + \sigma_\varepsilon^2)$  and, as before,  $\gamma_S$  minimizes the expected squared deviation between  $\tilde{y}_T$  and a linear combination of  $f\tilde{y}_{T-1}$  and  $y_T^{T+1}$ .

Clearly, from (4), the PF should only set  $\hat{y}_T = y_T^{T+1}$  when  $\gamma_S = 1$ . This condition requires both that the government filters the source data (as in Sargent), and does so without error. Then,  $\xi_t = 0$ , so  $\sigma_\varepsilon^2 = g^2\sigma_\eta^2$ , and using  $g = \sigma_v^2 / (\sigma_v^2 + \sigma_\eta^2)$ , we get  $\gamma_S = 1$ .

We can write the equations for the true process and the first estimate for the Sargent setup as:

$$\begin{aligned} \tilde{y}_t &= f\tilde{y}_{t-1} + v_t \\ y_t^{t+1} &= f\tilde{y}_{t-1} + g(\tilde{y}_t - f\tilde{y}_{t-1}) + (\xi_t + g\eta_t). \end{aligned}$$

Combining these, revisions are given by:

$$y_t^{t+1} - \tilde{y}_t = (g - 1)v_t + \varepsilon_t$$

so that the error in the revisions equation is correlated with that in the equation for the true process, but revisions are not serially correlated. By contrast, in Howrey's setup the errors in the revisions and true process equations are not correlated, but revisions are serially correlated.

Kishor and Koenig (2012) suggest combining these two models by allowing correlated

errors and serially correlated revisions, by specifying the revisions process as:

$$y_t^{t+1} - \tilde{y}_t = k (y_{t-1}^t - \tilde{y}_{t-1}) + (g - 1) v_t + \varepsilon_t. \quad (5)$$

Then the best estimate of  $\tilde{y}_T$  is given by

$$\hat{y}_T = f \tilde{y}_{T-1} + \gamma_K (y_T^{T+1} - k (y_{T-1}^T - \tilde{y}_{T-1}) - f \tilde{y}_{T-1}) \quad (6)$$

with  $\gamma_K = \gamma_S$ , and where ‘best’ is defined following (2) or (4).

As a consequence, the PF could use a model, which encompasses the Howrey and Sargent specifications, and combines equation (1) for the true process and equation (5) for the revisions process, written below for convenience as:

$$\begin{aligned} \tilde{y}_t &= f \tilde{y}_{t-1} + v_t \\ y_t^{t+1} - \tilde{y}_t &= k (y_{t-1}^t - \tilde{y}_{t-1}) + \varepsilon_t - (1 - g) v_t. \end{aligned} \quad (7)$$

Because of the cross-equation correlation in the error terms, efficient estimation of the two-equation system requires the seemingly-unrelated-regressions estimator (SURE). This is carried out on observations from  $t = 2$  up to  $T - 1$ . The covariance matrix of  $(v_t, \varepsilon_t - (1 - g) v_t)$  is given by:

$$Q = \begin{bmatrix} \sigma_v^2 & -(1 - g) \sigma_v^2 \\ -(1 - g) \sigma_v^2 & \sigma_\varepsilon^2 + (1 - g)^2 \sigma_v^2 \end{bmatrix}.$$

From the estimates of the elements of  $Q$ , the parameter  $\gamma_k = g \sigma_v^2 / (g^2 \sigma_v^2 + \sigma_\varepsilon^2)$  in equation (6) can be computed. Combining this with the estimates of  $f$  and  $k$  from the system in (7) provides an estimate of  $\hat{y}_T$ , as required by step (2) of the forecasting strategy.

For ease of exposition we have assumed that the first revision reveals the true value, but this clearly needs to be generalized, given the discussion in section 2 of the many rounds of revisions such data typically experience. This suggests that at time  $T + 1$  estimates are required of the true values of not just  $\tilde{y}_T$ , but of earlier periods,  $\tilde{y}_{T-1}$ ,  $\tilde{y}_{T-2}, \dots$  as well. The following section describes a practical implementation of the Kishor and Koenig (2012) approach allowing for multiple revisions.

### 3.4 The Kishor and Koenig (2012) approach in practice

For expositional purposes, we assumed in sections 3.2 and 3.3 that the true value  $\tilde{y}_t$  is observed two quarters after the observational quarter, that is,  $\tilde{y}_t = y_t^{t+2}$ . In practice, data are subject to annual revisions that may shape the true values, such that in general  $\tilde{y}_t = y_t^{t+q}$ , as described in section 2. For US data, accommodating the three rounds of annual revisions would require setting  $q = 14$ . The model described in this section is motivated by, and generalises, the simpler cases described in sections 3.2 and 3.3.

To implement the Kishor and Koenig (2012) approach in practice, we assume an  $\text{AR}(p)$  for the ‘final data’, and equations for  $q-1$  rounds of revisions. The key components are the  $t + 1$ -vintage vector  $\mathbf{y}^{t+1}$  and the true values vector  $\tilde{\mathbf{y}}_t$  :

$$\mathbf{y}^{t+1} = \begin{bmatrix} y_t^{t+1} \\ y_{t-1}^{t+1} \\ y_{t-2}^{t+1} \\ \vdots \\ y_{t-q+1}^{t+1} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{y}}_t = \begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{t-1} \\ \tilde{y}_{t-2} \\ \vdots \\ \tilde{y}_{t-q+1} \end{bmatrix} .$$

We assume that  $y_{t-q+1}^{t+1} = \tilde{y}_{t-q+1}$ :  $y_{t-q+1}^{t+1}$  is supposed to be an efficient estimate of the

true value  $\tilde{y}_{t-q+1}$ . Then the model can be written succinctly in state-space form, with measurement and state equations given by:

$$\mathbf{y}^{t+1} = \begin{bmatrix} \mathbf{I}_q & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_t \\ \mathbf{y}^{t+1} - \tilde{\mathbf{y}}_t \end{bmatrix}, \quad (8)$$

and:

$$\begin{bmatrix} \tilde{\mathbf{y}}_t \\ \mathbf{y}^{t+1} - \tilde{\mathbf{y}}_t \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{F} & 0_{q \times q} \\ 0_{q \times q} & \mathbf{K} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_{t-1} \\ \mathbf{y}^t - \tilde{\mathbf{y}}_{t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ \varepsilon_t \end{bmatrix}. \quad (9)$$

In (8),  $\mathbf{I}_q$  denotes an identity matrix of order  $q$ . The disturbances vectors are  $v_t = (v_{1t}, 0, \dots, 0)'$  and  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{q-1t}, 0)'$ . The errors in the data revision equations,  $\varepsilon_t$ , are allowed to be correlated with the disturbances to the true values,  $v_{1t}$ , as well as being contemporaneously correlated. Defining  $\mathbf{v}_t = (v_t, \varepsilon_t)'$ , we let  $E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{Q}$ .

The true values  $\tilde{y}_t$  follow an autoregression of order  $p$ , defined by the first block of (9) with:

$$\mathbf{F} = \begin{bmatrix} \mathbf{f} & 0_{1 \times (q-p)} \\ \mathbf{I}_{q-1} & 0_{(q-1) \times 1} \end{bmatrix}$$

where  $\mathbf{f} = (f_1, \dots, f_p)$  is the  $1 \times p$  coefficient vector ( $p < q$ ). The matrix  $\mathbf{K}$  describes the dynamics of  $q - 1$  data revisions  $\mathbf{y}^{t+1} - \mathbf{y}_t$ :

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & \dots & k_{1,q-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ k_{q-1,1} & \dots & k_{q-1,q-1} & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

The  $q \times 1$  vectors  $\mathbf{c}_1$  and  $\mathbf{c}_2$  are  $\mathbf{c}_1 = (c_1, 0, \dots, 0)'$  and  $\mathbf{c}_2 = (c_{21}, \dots, c_{2q-1}, 0)'$ .

The above provides a complete specification of the model, allowing for multiple revi-

sions. The unknown parameters  $\mathbf{c}_1$ ,  $\mathbf{c}_2$ ,  $\mathbf{F}$ ,  $\mathbf{K}$  and  $\mathbf{Q}$  are estimated by SURE. Because  $\tilde{\mathbf{y}}_t$  consists solely of final values, given the assumption that  $y_{t-q+1}^{t+1+i} = \tilde{y}_{t-q+1}$  for  $i \geq 0$ , the estimation sample has to end at  $t = T - q + 1$ , when  $T + 1$  is the forecast-origin data vintage. Application of the Kalman Filter then provides estimates of the fully-revised values of past observations (i.e., of the post-revision values of current and past observations,  $\tilde{y}_{T-q+2}$  up to  $\tilde{y}_T$ ). Forecasts of post-revision future observations  $\tilde{y}_{T+1}, \dots, \tilde{y}_{T+h}$  are obtained by iterating the state equation for  $\mathbf{y}_t$  using the estimate of  $\hat{\mathbf{y}}_T$  obtained from the Kalman filtering using all data through vintage  $T + 1$ .

Alternative approaches to dealing with many rounds of data revisions are proposed by Cunningham, Eklund, Jeffery, Kapetanios and Labhard (2009), and Jacobs and van Norden (2011). The model of Jacobs and van Norden (2011) is discussed in section 6.1.

## 4 Data Revisions and Forecasting with DSGE Models

Dynamic stochastic general equilibrium (DSGE) models are now routinely used for forecasting: see, e.g., Del Negro and Schorfheide (2013). Just as with the autoregressive model for the true process in section 3.4 (or vector autoregressive models more generally), forecasts will need to be generated when only early estimates of the data for the more recent time periods are available.

Galvão (2017) shows how this can be accomplished, by explicitly modelling data revisions while estimating and forecasting DSGE models. The approach aims to predict revised values of macroeconomic variables, and also computes density forecasts while making an allowance for data uncertainty. As in Kishor and Koenig (2012), Galvão (2017) estimates the forecasting model – a DSGE model in her case – with final (true

or fully-revised data). She assumes that  $y_t^{t+q} = \tilde{y}_t$ , that is, after  $q - 1$  revisions the true value is revealed. This means that at time  $t = T + 1$  we only observe true values up to  $t = T - q + 1$ , and consequently a model of data revisions (as in section 3) is used to provide estimates of the true values of the data still subject to revision.

Galvão (2017) proposes a one-step method to jointly estimate the parameters of the DSGE model (with revised data) and the parameters describing the data revisions process. Based on the model, one can compute backcasts and forecasts for the data subject to revision, including their underlying predictive density. We describe her approach below.

Define  $y_t$  as an  $n \times 1$  vector of the endogenous DSGE variables written as deviations from their steady state values. In practice,  $y_t$  may also include lagged variables. The elements of the vector  $y_t$  need not be observable, and the absence of superscripts is deliberate. Define  $\theta$  as the vector of structural parameters. The solution of the DSGE model for a given vector of parameters  $\theta$  is written as:

$$y_t = F(\theta)y_{t-1} + G(\theta)v_t \tag{10}$$

where  $v_t$  is a  $r \times 1$  vector of structural shocks, and thus, the matrix  $G(\theta)$  is  $n \times r$ . Note also that  $v_t \sim iidN(0, Q)$ , and that  $Q$  is a diagonal matrix because the shocks are regarded as being structural. Equation (10) is the state equation of the state space representation of the DSGE model.

Define  $\tilde{Y}_t = (\tilde{y}_{1,t}, \dots, \tilde{y}_{m,t})'$  as the  $m \times 1$  vector of true values of the endogenous variables in  $y_t$  that are assumed to be observable. Typically,  $m < n$  and  $m \leq r$ . The Smets and

Wouters (2007) medium-sized model has  $m = r = 7$ . The measurement equation is:

$$\tilde{Y}_t = d(\theta) + H(\theta)y_t, \quad (11)$$

that is, the set of observable variables, such as inflation and output growth, are measured without error, although Galvão (2017) also shows that the approach would work if there are measurement errors.

The DSGE is estimated using the values observed after  $q - 1$  rounds of revisions, that is, assuming that  $\tilde{Y}_{t-q+1} = Y_{t-q+1}^{t+1}$ . Then the measurement equations are:

$$Y_{t-q+1}^{t+1} = d(\theta) + H(\theta)y_{t-q+1} \text{ for } t = 1, \dots, T, \quad (12)$$

and the last  $q - 1$  observations (that is,  $Y_{t-q+2}^{t+1}, \dots, Y_t^{t+1}$ ) have to be excluded.

Define the demeaned observed revisions between first releases  $Y_t^{t+1}$  and true values  $Y_t^{t+q}$  as:

$$rev_t^{t+q,1} = (Y_t^{t+1} - Y_t^{t+q}) - M_1 \text{ for } t = 1, \dots, T - q + 1,$$

where  $M_1$  is  $m \times 1$  vector of mean revisions. This implies that we observe  $T - q + 1$  values of the full revision process to a first release at  $T + 1$ , and that the full revision process for observation  $t$  is only observed at  $t + q$  because of the statistics office data release schedule. In general, for the  $k^{th}$  release, the (demeaned) remaining revisions up to the true values are:

$$rev_t^{t+q+1-k,k} = (Y_t^{t+k} - Y_t^{t+q}) - M_k \text{ for } t = 1, \dots, T - q + v \text{ and } k = 1, \dots, q - 1.$$

The released-based approach in Galvão (2017) augments the measurement equations (12) to include a time series of first releases, second releases, and so on, as:

$$\begin{bmatrix} Y_t^{t+1} \\ Y_{t-1}^{t+1} \\ \vdots \\ Y_{t-q+1}^{t+1} \end{bmatrix} = \begin{bmatrix} d(\theta) + M_1 \\ d(\theta) + M_2 \\ \vdots \\ d(\theta) \end{bmatrix} + \begin{bmatrix} H(\theta) & 0_m & \cdots & 0_m & I_m & 0_m & \cdots & 0_m \\ 0_m & H(\theta) & \cdots & 0_m & 0_m & I_m & \cdots & \\ & & \ddots & & & & \ddots & \\ 0_m & 0_m & & H(\theta) & 0_m & 0_m & & 0_m \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-q+1} \\ rev_t^1 \\ rev_{t-1}^2 \\ \vdots \\ rev_{t-q+2}^{q-1} \end{bmatrix} \quad (13)$$

for  $t = 1, \dots, T$  and:

$$rev_t^k = (Y_t^{t+k} - \tilde{Y}_t) - M_k \text{ for } k = 1, \dots, q - 1,$$

where the  $m \times 1$  vectors  $M_v$  allow for non-zero mean data revisions.

The state equations are augmented by data revision processes as:

$$rev_t^k = K_k rev_{t-1}^k + \xi_t^k + A_k v_t, \quad \xi_t^k \sim N(0, R_k) \quad (14)$$

where the serial correlation allows for predictable revisions if the  $m \times m$  matrix  $K_k$  is non-zero. The own innovation term  $\xi_t^k$  allows for data revisions that are caused by a reduction of measurement errors, and they are assumed to be uncorrelated across variables, so  $R_k$  is diagonal. The last term  $A_k v_t$  implies that the data revisions may be caused by new information not available at the time of the current release, but included in the revised data used to compute the complete effects of the structural shocks. Because the same

vector of structural shocks  $v_t$  may lead to revisions for each one of the variables in  $Y_t^{t+k}$ , then the processes  $rev_{1,t}^k, \dots, rev_{m,t}^k$  may be correlated.

Galvão (2017) proposes a Metropolis-in-Gibbs approach to jointly estimate the DSGE parameters  $\{\theta, Q\}$  and the parameters of the revision process:

$$\{M_1, \dots, M_v, K_1, \dots, K_k, A_1, \dots, A_k, R_1, \dots, R_k\}.$$

She shows that forecasts of future values of consumption and investment growth using a Smets and Wouters (2007) DSGE model are improved by using the release-based approach in comparison with the conventional approach that employs only the last vintage of data.

## 5 Vintage-based VARs: Models in terms of observables

Whereas the Kishor and Koenig (2012) and Galvão (2017) models discussed in sections 3.4 and 4 include unobserved components (namely, the true values of variables at the time when only earlier estimates are available), other modelling approaches consider the relationships between the observables directly. A natural way to do this, when we wish to model a range of different vintages, and not just the first release (say), is to use a vector autoregression (VAR, see Sims (1980)). A number of VAR-type models based only on observables have been proposed for modelling revisions, and these are reviewed in this section.

VARs were originally used to model the dynamic relationships between inter-connected macro variables, without the need for "incredible" identifying restrictions and hard to defend assumptions about exogeneity (see Sims (1980)). It was soon recognised that such

models would have many parameters to estimate for reasonable numbers of series and lagged values, and Bayesian methods were sometimes adopted: see, e.g., Doan, Litterman and Sims (1984) and Litterman (1986), and Bańbura, Giannone and Reichlin (2010) for a more recent contribution. In terms of modelling data subject to revision, the observable variables are the data defined by both the release data, and the reference quarter, and the potential for having highly parameterised models arises. As shown below, this is dealt with in a number of different ways.

## 5.1 The Garratt *et al.* (2008) model

The first VAR-type model we consider is that of Garratt *et al.* (2008). Following Patterson (1995), Garratt *et al.* (2008) work in terms of the level (of the log) of a variable (e.g., output, denoted by  $Y$ ). The variable is assumed to be integrated of order one ( $I(1)$ : see, e.g., the textbook treatment by Banerjee, Dolado, Galbraith and Hendry (1993)) and they assume that different vintage estimates are cointegrated such that data revisions are integrated of order zero (written  $I(0)$ ). So, for example,  $Y_{t-1}^{t+1}$  and  $Y_{t-1}^t$  are both  $I(1)$  - these are the second and first estimates of reference quarter  $t - 1$ , respectively. But  $Y_{t-1}^{t+1} - Y_{t-1}^t$ , the revision between the first and second estimates of the period  $t - 1$  value, is  $I(0)$ . They model a vector of variables comprising 3 elements,  $\mathbf{Z}^{t+1} = (Y_t^{t+1} - Y_{t-1}^t, Y_{t-1}^{t+1} - Y_{t-1}^t, Y_{t-2}^{t+1} - Y_{t-2}^t)'$ . The first element of  $\mathbf{Z}^{t+1}$  is a difference across vintage and observation, and the subsequent terms are revisions to past data. The inclusion of two revisions reflects the view that a ‘revision horizon’ of two is appropriate, in the sense that revisions such as  $Y_{t-2-j}^{t+1} - Y_{t-2-j}^t$  for  $j > 0$  are supposed to be largely unpredictable, and hence are not included in the vector of variables to be modelled. This assumption also serves to limit the dimensionality of the system to be estimated.

Garratt *et al.* (2008) relate  $\mathbf{Z}^{t+1}$  to two lags of itself, where the lagging is applied to

both the vintage and reference quarter scripts:

$$\mathbf{Z}^{t+1} = \mathbf{c} + \Phi_1 \mathbf{Z}^t + \Phi_2 \mathbf{Z}^{t-1} + \boldsymbol{\varepsilon}_t \quad (15)$$

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  $i = 0, 1, 2$ . The  $\Phi_i$  are 3 by 3 matrices of coefficients with third columns consisting solely of zeros (see Garratt *et al.* (2008) for details), that is, their approach implies specific restrictions on a VAR.

A disadvantage of the formulation in (15) is that shifts in the levels of  $Y$  due to base-year changes (or to other definitional changes) at times of benchmark revisions cannot be easily handled by the GLMS model. These arise because the model is based on differences in  $Y$  across time periods and vintages. Differencing-across-vintages means that level shifts between vintages affect the resulting series. The level-shift components of the benchmark revisions are removed from the real-time data set prior to the formulation and estimation of models such as (15).

## 5.2 The Vintage-based VARs

The problems created by re-basings of the data can be circumvented by instead specifying the model in ‘same-vintage-growth rates’ (see, e.g., Clements and Galvão (2012, 2013a), Carriero, Clements and Galvão (2015)). For example, let  $y_t^{t+1} = 400 (Y_t^{t+1} - Y_{t-1}^{t+1})$  be the (approximate) quarterly percentage change (at an annual rate) for period  $t$  computed using data vintage  $t + 1$ . Because the growth rate is calculated between two data points from the same vintage, level shifts or base-year changes will have no effect (to the extent that the change is simply a re-scaling of the data).

Suppose that in addition to modelling  $y_t^{t+1}$ , the first estimate of the growth rate for

period  $t$ , we also wish to model the revisions for the next  $q - 1$  quarters. This can be accomplished by modelling the vintage  $t + 1$  values of observations  $t - q + 1$  through  $t$  as a ‘vintage VAR’ (V-VAR):

$$\mathbf{y}^{t+1} = \mathbf{c} + \sum_{i=1}^p \mathbf{\Gamma}_i \mathbf{y}^{t+1-i} + \boldsymbol{\varepsilon}^{t+1} \quad (16)$$

where  $\mathbf{y}^{t+1} = (y_t^{t+1}, y_{t-1}^{t+1}, \dots, y_{t-q+1}^{t+1})'$ ,  $\mathbf{y}^{t+1-i} = (y_{t-i}^{t+1-i}, y_{t-1-i}^{t+1-i}, \dots, y_{t-q+1-i}^{t+1-i})'$ , and  $\mathbf{c}$  is  $q \times 1$ ,  $\mathbf{\Gamma}_i$  is  $q \times q$ , and  $\boldsymbol{\varepsilon}^{t+1}$  is a  $q \times 1$  vector of disturbances. The first element of  $\mathbf{y}^{t+1}$  is the ‘new’ observation  $y_t^{t+1}$ , and subsequent estimates are the revised estimates of ‘past’ observations,  $y_{t-1}^{t+1}, \dots, y_{t-q+1}^{t+1}$ .

When  $q$  is relatively large, the autoregressive order  $p$  may be set to a low value, e.g.,  $p = 1$  (see, e.g., Clements and Galvão (2013a)). Nevertheless, if  $q$  is large, say,  $q = 14$ , in order to capture the three rounds of annual revisions to which US national accounts data are subject, there will be 14 free coefficients to estimate in each of the 14 equations when  $p = 1$  (plus an intercept in each).

The number of coefficients to be estimated can be reduced by restricting the model, by supposing 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.,  $E[(y_t^{t+n+1} - y_t^{t+n}) | \mathbf{y}^{t+n}] = 0$ , whereas  $E[(y_t^{t+i+1} - y_t^{t+i}) | \mathbf{y}^{t+i}] \neq 0$  for  $i < n$ . We can impose this restriction on the VAR, where it translates to  $E(y_{t-n}^{t+1} | y_{t-n}^t) = y_{t-n}^t$ . This is achieved by specifying  $\mathbf{\Gamma}_1$  and  $\mathbf{\Gamma}_i$  ( $i = 2, \dots, p$ ) in (16) as:

$$\tilde{\mathbf{\Gamma}}_1 = \begin{bmatrix} & \boldsymbol{\gamma}_{n \times q} & \\ \mathbf{0}_{(q-n) \times (n-1)} & \mathbf{I}_{(q-n) \times (q-n)} & \mathbf{0}_{(q-n) \times 1} \end{bmatrix}, \quad \tilde{\mathbf{\Gamma}}_i = \begin{bmatrix} \boldsymbol{\gamma}_{i, n \times q} \\ \mathbf{0} \end{bmatrix}. \quad (17)$$

If  $n$  is set equal to 2, for example, then values after the first revision are assumed to be efficient forecasts (for example, for the US this would correspond to the BEA estimate published two quarters after reference quarter constituting an efficient forecast). An unrestricted intercept is included in each equation, to accommodate non-zero mean revisions. We refer to this model as the ‘news-restricted’ vintage-based VAR, RV-VAR.

Clements and Galvão (2013) compare the forecasting performance of the RV-VAR with that of the V-VAR. They also consider a periodic specification which captures the seasonal nature of some data revisions: the annual rounds of revisions which occur in the July of each year (for US data).

Clements and Galvão (2013a) discuss the interpretation of the forecasts generated by the V-VAR and related models. Consider the forecast-origin-vintage  $T + 1$ . At this time, the information set will include all the data vintages up to and including the time- $T + 1$  vintage, i.e.,  $\mathbf{Y}^{t+1}$  for  $t = 1, 2, \dots, T$ , where  $\mathbf{Y}^{t+1} = \{\dots, y_{t-1}^{t+1}, y_t^{t+1}\}$ . The  $h$ -step ahead forecast of the vector  $\mathbf{y}^{T+1+h}$  is defined as the conditional expectation, given the model and the information set:

$$\mathbf{y}^{T+1+h|T+1} \equiv E(\mathbf{y}^{T+1+h} \mid \mathbf{Y}^{T+1}, \mathbf{Y}^T, \dots).$$

The elements of vector  $\mathbf{y}^{T+1+h|T+1}$  are  $(y_{T+h}^{T+1+h|T+1}, \dots, y_{T+h-q+1}^{T+1+h|T+1})$ , and thus provide forecasts of the first estimate of  $y_{T+h}$ , of the second estimate of  $y_{T+h-1}$ , and so on down to a forecast of the  $q^{\text{th}}$  estimate of  $T + h - q + 1$ .

Suppose we require forecasts of the ‘true’ (revised) values, that is,  $\tilde{y}_{T-1}, \tilde{y}_T, \tilde{y}_{T+1}, \tilde{y}_{T+2}$ . As before, we assume that for a reasonably large  $q$  (e.g.,  $q = 14$ , chosen to include the annual revisions), we have  $\tilde{y}_t = y_t^{t+q}$ . That is, we can equate the true values with the

values available after  $q - 1$  revisions.<sup>7</sup> We need then to consider forecasts of a subset of the elements of  $\mathbf{y}^{T+1+h|T+1}$ , namely, the last elements of each vector for  $h = 1, 2, 3, \dots, h^*$ . For the forecast origin  $T + 1$ , this gives the following set of forecasts:  $y_{T+2-q}^{T+2|T+1}$ ,  $y_{T+3-q}^{T+3|T+1}$ ,  $\dots, y_{T+1+h^*-q}^{T+1+h^*|T+1}$ . All of which are the  $q$ -th estimates. Some of these estimates will relate to the past, and others to the present, relative to the vintage-origin  $T + 1$ . When  $h^* + (1 - q) \leq 0$ , we have forecasts of past observations (or ‘backcasts’),  $\tilde{y}_T, \tilde{y}_{T-1}, \tilde{y}_{T-2}$ , etc., and for  $h^* + (1 - q) > 0$  forecasts of ‘fully-revised’ future observations:  $\tilde{y}_{T+1}, \tilde{y}_{T+2}$ , etc.

## 6 Single-equation approaches: EOS and RTV.

Single-equation approaches have also been considered, and empirically have been found to work relatively well. We use the statistical framework of Jacobs and van Norden (2011) to show that, in principle, the traditional approach is not the best way of estimating an autoregressive model for forecasting. The statistical framework of Jacobs and van Norden (2011) separately identifies news and noise data revisions, as defined in section 3.1. Their model can be used to estimate the importance of the news and noise contributions to the data revisions of a particular series, and generally this would be facilitated by specifying a relatively small number of revisions. Otherwise it may be difficult to identify the separate news and noise components with any precision. In this section we use their model as a coherent statistical framework for deriving the properties of single-equation approaches, including ignoring data revisions (as explained below) when forecasting. The clear demarcation of revisions into news and noise allows us to determine the implications of each for the relative forecast performance of the single-equation approaches, at least in

---

<sup>7</sup>This will not be literally true. More precisely, we suppose that any differences between the two are unpredictable.

population (that is, abstracting from model selection and parameter estimation issues).

Section 6.1 describes the statistical framework. Sections 6.2 and 6.3 then provide the properties of the traditional approach, and an approach suggested by Koenig, Dolmas and Piger (2003), and show that the latter is optimal, at least in population. Section 6.4 then discusses the properties of interval and density forecasts derived from the two single-equation models.

## 6.1 A News and Noise model of data revisions

The Jacobs and van Norden (2011) statistical framework for modelling data revisions is based directly on the news / noise distinction. Each release is set equal to the true value plus an error, or errors, where the errors correspond to news or noise, and are unobserved. So, for example, at period  $t+s$ , the SO releases an estimate of the value of  $y$  for reference period  $t$ , which we denote  $y_t^{t+s}$ , which is written as:

$$y_t^{t+s} = \tilde{y}_t + v_t^{t+s} + \varepsilon_t^{t+s}$$

where,  $\tilde{y}_t$  is the true value, and  $v_t^{t+s}$  and  $\varepsilon_t^{t+s}$  are the news and noise components. We allow for up to  $q$ -releases, with  $s = 1, \dots, q$ . For any given  $s$ , one or other of the news and noise components may be absent.

Jacobs and van Norden (2011) stack the  $q$ -releases of  $y_t$ , namely,  $y_t^{t+1}, \dots, y_t^{t+q}$ , in the vector  $\mathbf{y}_t = (y_t^{t+1}, \dots, y_t^{t+q})'$ , and similarly  $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{t+1}, \dots, \varepsilon_t^{t+q})'$  and  $\mathbf{v}_t = (v_t^{t+1}, \dots, v_t^{t+q})'$ , so that:

$$\mathbf{y}_t = \mathbf{i}\tilde{y}_t + \mathbf{v}_t + \boldsymbol{\varepsilon}_t \tag{18}$$

where  $\mathbf{i}$  is a  $q$ -vector of ones. In order that the news revisions are not correlated with the earlier release, namely that  $Cov(v_t^{t+s}, y_t^{t+s}) = 0$ , where  $v_t^{t+s} = y_t^{t+s} - \tilde{y}_t$ , we need

to assume a process for  $\tilde{y}_t$  which includes the news components. For example, if  $\tilde{y}_t$  is assumed to follow an AR( $p$ ), say, with iid disturbances  $R_1\eta_{1t}$ , then we need to add in the sum of  $q$  news components  $v_{i,t}$ :

$$\tilde{y}_t = \rho_0 + \sum_{i=1}^p \rho_i \tilde{y}_{t-i} + R_1\eta_{1t} + \sum_{i=1}^q v_{i,t}. \quad (19)$$

The  $v_{i,t}$  are specified as  $v_{i,t} = \sigma_{v_i}\eta_{2t,i}$  (for  $i = 1, \dots, l$ ), and both  $\eta_{1t}$  and  $\eta_{2t,i}$  are  $iid(0, 1)$ .

The news and noise components of each vintage in  $\mathbf{y}_t$  are:

$$\mathbf{v}_t = \begin{bmatrix} v_t^{t+1} \\ v_t^{t+2} \\ \vdots \\ v_t^{t+q} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^q v_{i,t} \\ \sum_{i=2}^q v_{i,t} \\ \vdots \\ v_{q,t} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_t^{t+1} \\ \varepsilon_t^{t+2} \\ \vdots \\ \varepsilon_t^{t+q} \end{bmatrix} = \begin{bmatrix} \sigma_{\varepsilon_1}\eta_{3t,1} \\ \sigma_{\varepsilon_2}\eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon_q}\eta_{3t,q} \end{bmatrix}, \quad (20)$$

where  $\eta_{3t,i}$  is  $iid(0, 1)$ . The shocks are also mutually independent, that is, if  $\eta_t = [\eta_{1t}, \eta'_{2t}, \eta'_{3t}]$ , then  $E(\eta_t) = 0$ , with  $E(\eta_t\eta'_t) = I$ .

To see how this setup delivers appropriately defined news and noise revisions, consider a few illustrative cases. The first estimate of  $y_t$ ,  $y_t^{t+1}$ , is  $y_t^{t+1} = \tilde{y}_t + v_t^{t+1} + \varepsilon_t^{t+1} = \rho_0 + \sum_{i=1}^p \rho_i \tilde{y}_{t-i} + R_1\eta_{1t} + \sigma_{\varepsilon_1}\eta_{3t,1}$ . This does not include any news component. The second estimate is:  $y_t^{t+2} = \tilde{y}_t + v_t^{t+2} + \varepsilon_t^{t+2} = \rho_0 + \sum_{i=1}^p \rho_i \tilde{y}_{t-i} + R_1\eta_{1t} + \sigma_{\varepsilon_2}\eta_{3t,2} + \sigma_{v_i}\eta_{2t,1}$ . Suppose there is no noise, so  $\eta_{3t,2} = 0$ . Then clearly  $y_t^{t+2}$  is a more accurate estimate of  $\tilde{y}_t$  than  $y_t^{t+1}$ , as it includes the news  $\sigma_{v_i}\eta_{2t,1}$ . Further, the revision between  $y_t^{t+2}$  and the true value  $\tilde{y}_t$  is uncorrelated with  $y_t^{t+2}$ , so that  $y_t^{t+2}$  is an efficient estimate:

$$Cov(y_t^{t+2} - \tilde{y}_t, y_t^{t+2}) = Cov\left(\sum_{i=2}^q v_{i,t}, v_{1,t}\right) = 0.$$

Suppose now that there is only noise, so  $v_t^{t+2} = 0$  but  $\varepsilon_t^{t+2} \neq 0$ . It follows immediately that the revisions induced by the second estimate are predictable using  $y_t^{t+2}$ :

$$\text{Cov}(y_t^{t+2} - \tilde{y}_t, y_t^{t+2}) = \sigma_{\varepsilon_2}^2.$$

News revisions imply that  $\text{var}(y_t^{t+1}) < \text{var}(y_t^{t+q})$ , while noise revisions imply that  $\text{var}(y_t^{t+1}) > \text{var}(y_t^{t+q})$ , assuming that later estimates are less ‘noisy’ ( $\sigma_{\varepsilon_1} > \sigma_{\varepsilon_q}$ ). If  $\sigma_{v_i} = 0$  and  $\sigma_{\varepsilon_i} = 0$  the  $q^{\text{th}}$  released value is the true value,  $y_t^{t+q} = \tilde{y}_t$ . The assumption that  $\tilde{y}_t$  is an  $I(0)$  stationary process ensures that  $\mathbf{y}_t$  is a stationary process from (18), as both the news and noise terms are stationary. This is a reasonable assumption when the model is applied to an  $I(0)$  transformation of the data, such as growth rates.

The model set out above implies that both noise and news revisions are zero mean, so that the unconditional mean of the underlying series  $\{\tilde{y}_t\}$  and the observed data  $\{\mathbf{y}_t\}$  are equal at  $\rho_0(1 - \rho(1))^{-1}$ . Non-zero mean revisions can easily be accommodated. Assume that each news term is instead  $v_{i,t} = \mu_{v_i} + \sigma_{v_i}\eta_{2t,i}$ , and the noise components are  $\varepsilon_t^{t+i} = -\mu_{\varepsilon_i} + \sigma_{\varepsilon_i}\eta_{3t,i}$ . The true process is now:

$$\tilde{y}_t = [\rho_0 + \sum_{i=1}^q \mu_{v_i}] + \sum_{i=1}^p \rho_i \tilde{y}_{t-i} + R_1 \eta_{1t} + \sum_{i=1}^q \sigma_{v_i} \eta_{2t,i}, \quad (21)$$

since now  $\sum_{i=1}^q v_{it} = \sum_{i=1}^q \mu_{v_i} + \sum_{i=1}^q \sigma_{v_i} \eta_{2t,i}$ . The news and noise processes of each vintage are:

$$\mathbf{v}_t = - \begin{bmatrix} \sum_{i=1}^q \mu_{v_i} \\ \sum_{i=2}^q \mu_{v_i} \\ \vdots \\ \mu_{v_q} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^q \sigma_{v_i} \eta_{2t,i} \\ \sum_{i=2}^q \sigma_{v_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v_q} \eta_{2t,q} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = - \begin{bmatrix} \mu_{\varepsilon_1} \\ \mu_{\varepsilon_2} \\ \vdots \\ \mu_{\varepsilon_q} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon_1,t} \eta_{3t,1} \\ \sigma_{\varepsilon_2,t} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon_q,t} \eta_{3t,q} \end{bmatrix}. \quad (22)$$

The statistical model can be cast in state-space form using (18) as the measurement equation and combining (21) and (22) to obtain the transition equation. The parameters can be estimated by maximum likelihood using the Kalman Filter, as described by Jacobs and van Norden (2011).

## 6.2 Estimating AR forecasting models using EOS data

Assuming the variable we wish to forecast can be described by the model set out in section 6.1, we begin with the standard or conventional approach, which effectively ignores the data revisions. We suppose the forecasting model is an autoregression. The conventional approach estimates this model on the vintage of data available at the forecast origin. From the discussion in section 3, such an approach is expected to be non-optimal, as it ‘mixes apples and oranges’. The framework outlined in section 6.1 allows us to determine why the conventional approach is not able to deliver optimal forecasts in real-time.

Consider forecasting at time  $T + 1$ . The  $T + 1$  vintage of data contains data up to  $T$ ,  $\{\dots, y_{T-1}^{T+1}, y_T^{T+1}\}$ , and the model is estimated on this data (termed End of Sample, EOS, by Koenig *et al.* (2003)). For an AR( $p$ ) we have:

$$y_t^{T+1} = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^{T+1} + e_{t,EOS} \quad (23)$$

where the unknown parameters are estimated on the observations  $t = p + 1, \dots, T$ .

Writing the model in matrix notation:

$$Y^{T+1} = \mathbf{i}\alpha_0 + \mathbf{Y}_{-1}^{T+1}\boldsymbol{\alpha} + error \quad (24)$$

where  $\mathbf{Y}_{-1}^{T+1} = [Y_{-1}^{T+1}, \dots, Y_{-p}^{T+1}]$ ,  $\mathbf{i}$  is a  $T - p$  vector of 1’s, and the vectors of observations

$Y^{T+1}$  and  $Y_{-i}^{T+1}$ ,  $i = 1, \dots, p$ , are:

$$Y^{T+1} = [y_{p+1}^{T+1}, \dots, y_{T-1}^{T+1}, y_T^{T+1}]', \quad Y_{-i}^{T+1} = [y_{p+1-i}^{T+1}, \dots, y_{T-i-1}^{T+1}, y_{T-i}^{T+1}]'$$

for  $i = 1, \dots, p$ .

Clements and Galvão (2013b) derive the population values of the least-squares estimator of the parameters in (23), when the data are generated by (18)–(20), as:

$$\begin{aligned} \boldsymbol{\alpha}^* &= (\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}} + \boldsymbol{\Sigma}_{\mathbf{v}} + \boldsymbol{\Sigma}_{\tilde{\mathbf{y}}\mathbf{v}} + \boldsymbol{\Sigma}'_{\tilde{\mathbf{y}}\mathbf{v}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{-1} (\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}} + \boldsymbol{\Sigma}'_{\tilde{\mathbf{y}}\mathbf{v}}) \boldsymbol{\rho} \\ \alpha_0^* &= (1 - \boldsymbol{\alpha}^* \mathbf{i}) \mu_{\tilde{\mathbf{y}}}, \end{aligned} \tag{25}$$

where  $\boldsymbol{\Sigma}_{\mathbf{v}}$  and  $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$  are second moment matrices of the news and noise components, and  $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}\mathbf{v}}$  is the second moment matrix between the news and the underlying process,  $\tilde{y}_t$ , and  $\mu_{\tilde{\mathbf{y}}} \equiv E(\tilde{y}_t)$  (see, Clements and Galvão (2013b, Supplementary Material) for details).

Clements and Galvão (2013b) also show that these parameter values are not optimal, that is, they are not the values that minimize the real-time out-of-sample expected squared forecast error, when the forecast is conditioned upon the forecast-origin vintage values of the data (as is standard practice). That is, if the forecast is given by  $\phi_0 + \boldsymbol{\phi}' \mathbf{y}_T^{T+1}$ , where  $\mathbf{y}_T^{T+1} = (y_T^{T+1}, \dots, y_{T-p+1}^{T+1})'$ , then setting  $\phi_0 = \alpha_0^*$ , and  $\boldsymbol{\phi} = \boldsymbol{\alpha}^*$ , does not minimize the expected squared forecast error whether the aim is to forecast  $y_{T+1}^{T+2}$  (the first estimate) or some later vintage estimate. This is perhaps not surprising, given that the majority of the data underlying the estimation are mature or revised data, whereas  $\mathbf{y}_T^{T+1}$  contains the first estimate of  $y_T$ , the second estimate of  $y_{T-1}$ , and so on. In terms of the cider press analogy in section 3.2, equation (23) is estimated on (mostly) mature data - the ‘apples’, but the forecast is conditioned on  $\mathbf{y}_T^{T+1}$  - the ‘oranges’.

We let  $\phi_0^*$  and  $\boldsymbol{\phi}^*$  denote the values of the parameters in the forecast function  $\phi_0 +$

$\phi' \mathbf{y}_T^{T+1}$  which minimize the expected squared error for forecasting  $y_{T+1}^{T+2}$ .

### 6.3 Estimating AR forecasting models using RTV data

Building on Koenig *et al.* (2003), Clements and Galvão (2013b) show that estimating the AR model using real-time-vintage (RTV) data delivers optimal estimators of the forecasting model, in population, when the forecast is to be conditioned on  $\mathbf{y}_T^{T+1}$ . RTV estimates the model:

$$y_t^{t+1} = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i}^t + e_{t,RTV} \quad (26)$$

on observations,  $t = p+1, \dots, T$ , where in contrast to EOS, the superscript denoting the vintage is not fixed at the latest-available vintage. In matrix notation:

$$Y^m = \mathbf{i}\beta_0 + \mathbf{Y}_{-1}^m \boldsymbol{\beta} + error$$

where  $Y^m$  and  $\mathbf{Y}_{-1}^m = [Y_{-1}^m, \dots, Y_{-p}^m]$  are given by:

$$Y^m = [y_{p+1}^{p+2}, \dots, y_{T-1}^T, y_T^{T+1}]', \quad Y_{-i}^m = [y_{p+1-i}^{p+1}, \dots, y_{T-i-1}^{T-1}, y_{T-i}^T]', \quad i = 1, \dots, p.$$

$Y^m$  and  $Y_{-i}^m$  contain data of a constant maturity, as indicated by the subscript  $m$ . All the observations in  $Y^m$  are first estimates, and those in  $Y_{-i}^m$  are  $i^{th}$  estimates. This means that, say, the first row of  $[Y^m: \mathbf{Y}_{-1}^m]$  contains the first estimate of  $y_{p+1}$ , the first estimate of  $y_p$ , the second estimate of  $y_{p-1}$ , etc. The last row has the same maturities: the first estimate of  $y_T$ , the first estimate of  $y_{T-1}$ , the second estimate of  $y_{T-2}$ , and so on.

Estimation of (26) results in estimates of the  $\beta$  parameters which minimize the expected squared error, at least in terms of forecasting the first estimate,  $y_{T+1}^{T+2}$ , and simple adjustments can be applied for forecasting later estimates when revisions have non-zero

means. That is  $\beta_0^* = \phi_0^*$ , and  $\beta^* = \phi^*$ .

Clements and Galvão (2013b) provide the general formula for the parameter estimator. We illustrate with the special case of an AR(1) for the true process, and for general revisions that are a combination of news and noise, when an AR(1) forecasting model is used. The optimal value of the AR parameter is:

$$\beta_1^* = \frac{\rho_1 (\sigma_{\tilde{y}}^2 - \sigma_v^2)}{\sigma_{\tilde{y}}^2 - \sigma_v^2 + \sigma_{\varepsilon_1}^2}$$

where  $\sigma_v^2 \equiv \sum_{i=1}^q \sigma_{v_i}^2$ , and  $\sigma_{\tilde{y}}^2 = Var(\tilde{y}_t)$ , and  $\beta_1^*$  is the population value of the parameter from RTV. When revisions are pure news ( $\sigma_{\varepsilon_1}^2 = 0$ ):

$$\beta_{1,news}^* = \rho_1, \quad \beta_{0,news}^* = \rho_0,$$

so RTV returns the population parameters of the true process. However,  $\beta_{1,news}^* = \rho_1$  only holds for  $p = 1$ : in general when there are news revisions the parameter vector of the underlying process  $\tilde{y}_t$  (i.e.,  $\rho$ ) is not optimal from a forecasting perspective when the forecasts are conditioned on early estimates.

For pure noise ( $\sigma_v^2 = 0$ ):

$$\beta_{1,noise}^* = \frac{\rho_1 \sigma_{\tilde{y}}^2}{\sigma_{\tilde{y}}^2 + \sigma_{\varepsilon_1}^2} \quad (27)$$

so that  $\beta_{1,noise}^* < \rho_1$  when  $\sigma_{\varepsilon_1}^2 \neq 0$ .

Consider now EOS. When revisions are news, we can show that the EOS estimator simplifies such that  $\alpha_1^* = \rho_1$ , matching the optimal value, but this is true only for the special case of  $p = 1$ .

Under noise:

$$\alpha_1^* = \frac{\rho_1 \sigma_{\tilde{y}}^2}{\sigma_{\tilde{y}}^2 + \sigma_{\varepsilon_q}^2} \quad (28)$$

An immediate implication is that  $|\alpha_1^*| > |\phi_1^*|$  if earlier revisions are larger than later revisions, as might be expected (compare (28) to (27) when  $\sigma_{\varepsilon_1}^2 > \sigma_{\varepsilon_q}^2$ ). Note that if  $\sigma_{\varepsilon_q}^2 = 0$ , so that the truth is eventually revealed when there is noise, then  $\alpha_1^* = \rho_1$  for a large estimation sample. Even so,  $\rho_1$  is not the parameter vector that minimizes the real-time squared forecast loss ( $\phi_1^* \neq \rho_1$ ).

The difference between EOS and RTV can be visualized in terms of table 1. Before the estimation, data is converted to quarterly growth rates within each column. EOS uses the column of data corresponding to the latest vintage of data available at the forecast origin. By way of contrast, RTV uses the elements in the diagonals of the data array for the left-hand-side and explanatory variables.

For example, for forecasting in 2015Q1, for an AR(2), the last RTV observation is  $y_T^{T+1} = 100[\ln(Y_{14Q4}^{15Q1}) - \ln(Y_{14Q3}^{15Q1})]$  while lags are taken from the previous vintage as  $y_{T-1}^T = 100[\ln(Y_{14Q3}^{14Q4}) - \ln(Y_{14Q2}^{14Q4})]$  and  $y_{T-2}^T = 100[\ln(Y_{14Q2}^{14Q4}) - \ln(Y_{14Q1}^{14Q4})]$ . A similar approach is followed for all  $t = p + 1, \dots, T$ . EOS would use data only from the 2015Q1 vintage for both the left-hand side values and the lags.

Finally, RTV is related to the vintage-based VAR model. Equation (26) corresponds to the first equation of the system of equations given by (16). When the lag order of the VAR is one, and the dimension of the VAR vector is  $p$ , there is an exact equivalence between the RTV model and the first equation of the VAR. RTV directly models the first release,  $y_t^{t+1}$ , as does the first equation of the VAR, whereas in addition the VAR models later estimates beyond the first.

## 6.4 Density Forecasting

Most of the literature has looked at first-moment prediction, but a few papers consider the impact on second-moment prediction, and the calculation of prediction intervals and

density forecasts. Clements (2017) and Clements and Galvão (2017) compare RTV and EOS in the context of computing predictive interval and predictive densities for short-horizon forecasting. They are mainly interested in correctly measuring the forecasting uncertainty around one-step-ahead forecasts computed in real-time, with the aim of predicting the first-release value.

A simple model of data revisions suffices to show that RTV delivers predictive densities which match the true underlying densities, while EOS delivers predictive densities that are too wide when data revisions are news, but too narrow when they are noise.

The model for data revisions is a simplified version of that in section 6.1. It assumes the true (i.e., fully-revised) values  $\tilde{y}_t$  follow an AR(1):

$$\tilde{y}_t = \rho_1 \tilde{y}_{t-1} + \eta_t + v_t, \quad |\rho_1| < 1 \quad (29)$$

where  $\eta_t$  is the underlying disturbance, and  $v_t$  is a news revision with variance  $\sigma_v^2$ , and the first estimate is given by:

$$y_t^{t+1} = \tilde{y}_t - v_t + \varepsilon_t \quad (30)$$

with  $y_t^{t+n} = y_t$  for  $n = 2, 3, \dots$ . Here  $\varepsilon_t$  is a noise revision with variance  $\sigma_\varepsilon^2$ . Then the revision  $y_t^{t+2} - y_t^{t+1} \equiv \tilde{y}_t - y_t^{t+1} = v_t - \varepsilon_t$  consists of a noise component (when  $\sigma_\varepsilon^2 \neq 0$ ) and a news component (when  $\sigma_v^2 \neq 0$ ).  $\eta_t$ ,  $v_t$  and  $\varepsilon_t$  are assumed to be mutually uncorrelated, zero-mean random variables.

Clements (2017) supposes the  $\eta_t$  are homoscedastic,  $var(\eta_t) = \sigma_\eta^2$ . Clements and Galvão (2017) also allow for conditional heteroscedasticity -  $var(\eta_t)$  follows an ARCH(1), or a GARCH(1,1), or a stochastic volatility AR(1) process. That macroeconomic volatility may be time-varying has been reported by Clark (2011), Clark and Ravazzolo (2014) and Diebold, Schorfheide and Shin (2016)), *inter alia*.

In the homoscedastic case, Clements (2017) shows that prediction intervals (calculated in the standard way, i.e., Box and Jenkins (1970)) using EOS are too wide when revisions are news, because the predictive variance is over-estimated, and that the reverse situation holds when revisions are noise. RTV delivers correctly-sized intervals. Clements and Galvão (2017) extend these results to the heteroscedastic case. When there is conditional heteroscedasticity, they show that estimating the forecasting model by RTV, with an appropriate model for the variance of the errors, will result in well-calibrated one-step-ahead predictive densities.

## 7 Evidence on the Performance of Alternative Methods of Forecasting

The relative performance of the different models and methods surveyed in this article will likely depend on the properties of the series under consideration, and in particular on the nature of the data revisions to that series. It may also depend on whether the aim is to forecast an early release, or a more mature vintage, such as the fully-revised value. Finally, relative performance may depend on whether point forecasts are required, or density forecasts or prediction intervals.

In this section we briefly summarize some of the findings in the literature to provide guidance on which method might be better in any particular instance.

Clements and Galvão (2013a), Carriero *et al.* (2015) and Galvão (2017) provide evaluations of different approaches when the goal is to forecast the *revised values*, denoted  $\tilde{y}_{T+1}, \tilde{y}_{T+2}, \dots, \tilde{y}_{T+h}$ . Clements and Galvão (2013a) compare the forecasting performance of the Kishor and Koenig (2012) (KK) approach (see section 3) and the Garratt *et al.* (2008) approach (GLMS) with the vintage-based VAR (VB-VAR) (section 5) for US

GDP growth and inflation. The models are univariate - in the sense of modelling a single variable, but allow up to 14 releases of the variable in question. The findings favour the VB-VAR, which delivers more accurate point forecasts for both variables. Carriero *et al.* (2015) show that forecasting accuracy can be improved by using their Bayesian approach. Their approach better controls the adverse effects of parameter uncertainty in such large VAR models. They are also able to allow for more than one variable, allowing for cross-equation dynamics between revisions to different macroeconomic variables (such as output growth and inflation in their application).

While these papers consider point forecasts of fully-revised values, Galvão (2017) evaluates the density forecasts of the Smets and Wouters (2007) DSGE model when data revisions are modelled, as in Kishor and Koenig (2012), and compares the findings to those obtained using the conventional approach. She finds gains in terms of logarithmic scores from the release-based approach for predicting the revised values of macroeconomic variables such as consumption and investment growth.

As well as forecasting the revised values of *future* outcomes, there are times when forecasts of the revised values of current and past observations are required. As an example, Clements and Galvão (2012) show that improved real-time estimates of output and inflation gaps result from the use of VB-VAR model ‘backcasts’.

Clements and Galvão (2013b), Clements (2017) and Clements and Galvão (2017) present forecast comparisons in terms of predicting *initial releases*, that is,  $y_{T+1}^{T+2}, y_{T+2}^{T+3}, \dots, y_{T+h}^{T+h+1}$ . Clements and Galvão (2013b) describe some of the circumstances under which RTV might be expected to outperform EOS (section 6). In particular, their findings suggest that larger gains might occur when there are explanatory variables - i.e., autoregressive-distributed lag models - as opposed to purely autoregressive models. They also suggest that ignoring data revisions, as in using EOS compared to RTV, might attract a greater

penalty in terms of accuracy when the estimation sample is small, the process is more persistent, and revisions are news. Clements and Galvão (2013b) also find that more elaborate approaches, such as that of Kishor and Koenig (2012), do not outperform RTV for modelling US output growth and inflation. This parallels findings in the (non-real-time) forecasting literature that more elaborate, complicated models do not necessarily outperform their simpler adversaries.

Clements (2017) suggests that larger relative gains might accrue to RTV when the goal is to provide well-calibrated prediction intervals. Clements and Galvão (2017) extend the results to variables subject to time-varying conditional variance, and find that RTV provides more accurate density forecasts for nominal national account variables.

## 8 Final Remarks

We have not attempted an exhaustive survey of the literature on forecasting in real time, and our main focus has been on US data. Nevertheless, we have attempted to give the reader an introduction to the types of approaches that have been proposed, and to have worked through some of these approaches in sufficient detail to lay their workings bare. We have summarised some of the evidence on the relative forecasting performance of the different approaches, but as in the macro-forecasting literature more generally, rankings across methods are unlikely to remain the same across different variables, or sample periods, and so on. In any specific instance it would seem sensible to consider a number of approaches.

That the more complex models, which attempt to jointly model the true (or revised) values along with the revisions process, are not necessarily superior to simpler approaches in terms of forecasting is not surprising, given that simple models are often found to fare

well in the forecasting literature. There are a number of possible explanations. The potential of the more complex models might be negated by the need to specify and estimate the models on relatively small historical samples. As an example, Clements and Galvão (2013b) provide a Monte Carlo study, whereby data are simulated from a vintage-based VAR model, or the Kishor and Koenig (2012) model, and the forecasting performance of RTV and EOS is compared to that of an estimated version of the model that generated the data. The authors find the estimation sample has to be relatively large for forecasts from the model assumed as data generating process to beat the simpler models. Another possible explanation stresses non-constancy or structural breaks in the process being forecast, and suggests that simpler models might exhibit greater adaptivity, or be more robust: see, e.g., Castle, Clements and Hendry (2016). Although the model of Kishor and Koenig (2012) assumes the processes for the data revisions are constant over time, this may not be the case, and if so, the potential advantages of such models may dissipate, especially when one takes into account the difficulties inherent in precisely estimating these channels unless the sample is large.

## References

- Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, **40**, 319–340.
- Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large bayesian vector autoregressions. *Journal of Applied Econometrics*, *25*(1), 71–92.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.
- Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Carriero, A., Clements, M. P., and Galvão, A. B. (2015). Forecasting with Bayesian multivariate vintage-based VARs. *International Journal of Forecasting*, *31*(3), 757–768.
- Castle, J. L., Clements, M. P., and Hendry, D. F. (2016). An overview of forecasting facing breaks. *Journal of Business Cycle Research*, **12**(1), 3–23. DOI 10.1007/s41549-016-0005-2.
- Clark, T. E. (2011). Real-time density forecasts from Bayesian Vector Autoregressions with Stochastic Volatility. *Journal of Business and Economic Statistics*, **29**, 327–341.
- Clark, T. E., and Ravazzolo, F. (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575.
- Clements, M. P. (2017). Assessing macro uncertainty in real-time when data are subject to revision. *Journal of Business & Economic Statistics*, **35:3**, 420 – 433.

- Clements, M. P., and Galvão, A. B. (2012). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business & Economic Statistics*, **30(4)**, 554–562. DOI: 10.1080/07350015.2012/707588.
- Clements, M. P., and 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. DOI: 10.1016/j.ijforecast.2011.09.003.
- Clements, M. P., and 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(3)**, 458–477. DOI: 10.1002/jae.2274.
- Clements, M. P., and Galvão, A. B. (2017). Predicting Early Data Revisions to US GDP and the Effects of Releases on Equity Markets. *Journal of Business and Economic Statistics*, **35:3**, 389 – 406.
- Clements, M. P., and Galvão, A. B. (2013). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, *29(4)*, 698–714.
- Clements, M. P., and Galvão, A. B. (2017). Data revisions and real-time probabilistic forecasting of macroeconomic variables. Discussion paper ICM-2017-01, ICMA, Henley Business School, Reading.
- Corradi, V., Fernandez, A., and Swanson, N. R. (2009). Information in the revision process of real-time datasets. *Journal of Business and Economic Statistics*, **27**, 455–467.
- 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, North-Holland.

- Croushore, D. (2011a). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.
- Croushore, D. (2011b). Frontiers of real-time data analysis. *Journal of Economic Literature*, **49**, 72–100.
- Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105**(1), 111–130.
- 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*, **30**, 173–180. doi:10.1198/jbes.2009.08171.
- Del Negro, M., and Schorfheide, F. (2013). DSGE model-based forecasting. In Elliott, G., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 2.*, pp. 57–140: Amsterdam, North-Holland.
- Diebold, F. X., Schorfheide, F., and Shin, M. (2016). Real-time forecast evaluation of DSGE models with Stochastic Volatility. Mimeo, University of Pennsylvania.
- Doan, T., Litterman, R., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**, 1–100.
- Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37** (3), 403–417.
- 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.
- Galvão, A. B. (2017). Data revisions and DSGE models. *Journal of Econometrics*,

196(1), 215–232.

- 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.
- Howrey, E. P. (1978). The use of preliminary data in economic forecasting. *The Review of Economics and Statistics*, **60**, 193–201.
- 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. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*, **30(2)**, 181–190.
- 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.
- Litterman, R. (1986). Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics*, **4**, 25–38.
- Mankiw, N. G., and Shapiro, M. D. (1986). News or noise: An analysis of GNP revisions. *Survey of Current Business (May 1986)*, US Department of Commerce, Bureau of Economic Analysis, 20–25.
- 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.

- Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy*, **97**, 251–287.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48.
- Smets, F., and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, *97*(3), 586–606.
- Swanson, N. R., and van Dijk, D. (2006). Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry. *Journal of Business and Economic Statistics*, **24**, 24–42.
- Zwijnenburg, J. (2015). Revisions of quarterly GDP in selected OECD countries. *OECD Statistics Briefing*, **July 2015 - No. 22**, 1–12.