ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Assessing Macro Uncertainty In Real-Time When Data Are Subject To Revision

January 2015

Michael P Clements
ICMA Centre, Henley Business School, University of Reading

Discussion Paper Number:
ICM-2015-02

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Assessing Macro Uncertainty In Real-Time When Data Are Subject To Revision

Michael P. Clements*

ICMA Centre

Henley Business School

University of Reading

m.p.clements@reading.ac.uk

January 6, 2015

Abstract

Model-based estimates of future uncertainty are generally based on the in-sample fit of the model, as when Box-Jenkins prediction intervals are calculated. However, this approach will generate biased uncertainty estimates in real time when there are data revisions. A simple remedy is suggested, and used to generate more accurate prediction intervals for 25 macroeconomic variables, in line with the theory. A simulation study based on an empirically-estimated model of data revisions for US output growth is used to investigate small-sample properties.

Keywords: in-sample uncertainty, out-of-sample uncertainty, real-time-vintage estimation

JEL code: C53.

*Michael P. Clements is an Associate member of the Institute for New Economic Thinking, Oxford Martin School, University of Oxford, UK.

1 Introduction

There has been much recent interest in macro-forecasting in real-time. By this we mean how the forecasting model should be specified, estimated, and the resulting forecasts evaluated, once we acknowledge that the data on which these three activities is based is subject to revision (for all but a small number of series such as interest rates and exchange rates). A number of papers have considered modelling the revisions process (see, e.g., Cunningham, Eklund, Jeffery, Kapetanios and Labhard (2009), Jacobs and van Norden (2011), Kishor and Koenig (2012)); or using single-equation models with 'real-time-vintages' (as in Koenig, Dolmas and Piger (2003), Clements and Galvão (2013b)) which we refer to as RTV-estimation and discuss below; or modelling multiple vintages of data, as with vintage-based vector autoregressive models (see, e.g., Patterson (1995, 2003), Clements and Galvão (2013a)). Other papers have considered whether assessing predictability in real-time may change the conclusions one would draw concerning putative explanatory forces, or the usefulness of estimates of the output gap as a guide to monetary policy in real time.[1] Croushore (2011a, 2011b) provide useful state-of-the-art reviews.

These related strands of research are all concerned with first-moment prediction: either improving the accuracy of forecasts of the conditional expectation, or of providing a more realistic appraisal of the accuracy of these predictions. In this paper we instead investigate the implications of data revisions for assessments of forecast uncertainty, specifically, the accuracy of prediction intervals. We show that the 'traditional' approaches to calculating prediction intervals will tend to be either too wide, when data revisions 'add news', or too narrow, when the revisions process 'removes noise'. These effects are first-order, in the sense that they do not disappear when the sample size gets large, and are not caused by non-normal errors.[2]

Clearly, the effects of data vintages on first and second-moment prediction can be side-

[1] On the former, it has been argued that the use of final-revised data may exaggerate the predictive power of explanatory variables relative to what could actually have been achieved at the time using the then available data (see, e.g, Robertson and Tallman (1998), Faust, Rogers and Wright (2003)), and on the latter see Orphanides (2001), Orphanides and van Norden (2005), Garratt, Lee, Mise and Shields (2009, 2008) Clements and Galvão (2012), *inter alia.*

[2] For example, it is often argued that Box-Jenkins prediction intervals suffer from the neglect of parameter estimation uncertainty and the possible non-normality of the underlying model's disturbances, and as a result there has been much interest in bootstrapping prediction intervals (see e.g., Thombs and Schucany (1990)). These problems persist in the present context, but even in their absence prediction intervals would be incorrectly-sized in the presence of data vintages.

stepped by considering fully-revised data, whence observations are no longer subject to revision.[3] Pseudo out-of-sample exercises use fully-revised data (e.g., the vintage of data available at the time of the study), and 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. Such exercises are useful as a way of assessing how well the model or models fit or forecast the true data, and the out-of-sample aspect guards against 'overfitting', i.e., the models capturing chance or non-recurrent sample-specific features. A real-time forecasting exercise mimics the environment a real-world forecaster faces - 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, but in addition the data are taken from the vintages that would have been available at that point in time.[4] So assuming a one period delay in data availability, at time t a forecaster will have access to the vintage-t value of the period $t-1$ observation, denoted y^t_{t-1}, and similarly the second estimate of the $t-2$ period, y^t_{t-2}, and so on. If the forecaster wishes to use only data which has been revised n times, say, the most recent data that could be used would be for $t-n-1$ (i.e., y^t_{t-n-1}, $y^t_{t-n-2}, \ldots$). It will rarely be optimal to ignore data for periods $t-n-2, \ldots, t-1$ (for large n), so the real-world forecaster will be forced to work with data subject to revision. This is the environment we seek to mimic in the real-time analysis.

Section 2 presents a simple example to motivate the concerns of the paper: the true model is a zero-mean first-order autoregressive process (AR(1)), and the true value of the process is revealed the period after the first estimate, so only the first estimate is subject to revision. Further, we ignore estimation uncertainty, so that model parameters take on their population parameters. We show that the standard approach gives an incorrect assessment of forecast uncertainty, and consequently incorrectly-sized prediction intervals. We focus on prediction intervals, but the points we make apply more generally to measures of forecast uncertainty (such as forecast densities). We then suggest a remedy that has been used for first-moment prediction, and has the virtue of simplicity and does not require that the revisions process be

[3] Of course any observation no matter how far back in time may be changed in response to far-reaching methodological changes. By fully-revised data we mean data that have undergone the initial and three anual rounds of revsions (see e.g., Landefeld, Seskin and Fraumeni (2008) for a description of the revisions process of the US Bureau of Economic Analysis NIPA data).

[4] Real-time exercises are required to provide fair assessments of the relative accuracy of the model forecasts compared to survey expectations, for example.

modelled: section 3. Section 4 provides an empirical illustration, and shows the improvements that result from using the approach advocated in this paper. We consider 25 macro variables, which exhibit different patterns of revisions, so that the conclusions we draw are reasonably general and do not rest on a few variables with (possibly) idiosyncratic features. In section 5 we present a simulation study of the two approaches to calculating prediction intervals in a controlled environment that abstracts from various factors that might affect the empirical comparisons, such as parameter non-constancies, for example. Finally, section 6 offers some concluding remarks.

2 Motivating example

Suppose the true (i.e., fully-revised) values y_t follow an AR(1):

$$y_t = \alpha y_{t-1} + \eta_t + v_t \tag{1}$$

and the estimates of y_t are given by:

$$\begin{aligned} y_t^{t+1} &= y_t - v_t + \varepsilon_t \\ y_t^{t+n} &= y_t \end{aligned}$$

for $n = 2, 3, \ldots$. We assume η_t, v_t and ε_t are mutually uncorrelated, zero-mean iid random variables. Then the revision $y_t^{t+2} - y_t^{t+1} = v_t - \varepsilon_t$ consists of a noise component (when $\sigma_\varepsilon^2 = E\left(\varepsilon_t^2\right) \neq 0$) and a news component (when $\sigma_v^2 = E\left(v_t^2\right) \neq 0$). The news/noise characterization of data revisions is due to Mankiw and Shapiro (1986). Suppose that revisions are purely news, so that $\sigma_\varepsilon^2 = 0$. Then the first estimate $y_t^{t+1} = \alpha y_{t-1} + \eta_t$ does not contain the news component v_t, but the revised estimate (which is the fully-revised value in this simple illustration) adds this term: $y_t^{t+2} = y_t = \alpha y_{t-1} + \eta_t + v_t$. A characteristic of news is that the revised value is unpredictable from information available at the time of the first estimate, or in other words, the revision $y_t^{t+2} - y_t^{t+1} = v_t$ is not systematically related to y_t^{t+1}. But the news revision clearly is correlated with the true value. For news, later estimates are more accurate estimates of the true value than earlier estimates (this follows trivially here because $y_t^{t+2} = y_t$ and so is a perfect estimate). Suppose now that revisions are solely noise, i.e., $\sigma_\varepsilon^2 \neq 0$ (but $\sigma_v^2 = 0$). For noise, the revised value removes measurement error: the revisions are predictable (based on period $t-1$ information) but are not correlated with the true value.

It follows directly that news revisions imply $var(y_t^{t+1}) < var(y_t^{t+2})$, while noise revisions imply that $var(y_t^{t+1}) > var(y_t^{t+2})$.

Consider then a sequence of forecasts made in real time, and in particular, consider the 1-step prediction interval for the period T observation made at time T, at which time the available data consists of $\left\{\ldots, \mathbf{y}_{T-2}^{T-1}, \mathbf{y}_{T-1}^{T}\right\}$, where $\mathbf{y}_{T-j-1}^{T-j} = \left[\ldots, y_{T-j-2}^{T-j}, y_{T-j-1}^{T-j}\right]'$, for $j = 0, 1, 2, \ldots$. The traditional approach is to specify and estimate the forecasting model using the period T-vintage $\left\{\mathbf{y}_{T-1}^{T}\right\}$, on the grounds that this constitutes the best available estimates of $\{\ldots, y_{T-2}, y_{T-1}\}$, irrespective of whether revisions add news, remove noise, or are some combination of the two. The use of the forecast origin vintage was referred to as end-of-sample (EOS) estimation by Koenig *et al.* (2003).

An AR(1) is estimated on the EOS data:

$$y_t^T = \beta y_{t-1}^T + e_{t,EOS}, \quad \text{for } t = \ldots, T-2, T-1 \tag{2}$$

and the forecast of y_T is $\widehat{y_T}_{,EOS} = \beta y_{T-1}^T$. As the sample gets large relative to the number of data revisions, it follows that OLS estimation of β in (2) will consistently estimate α in (1), because all but the last observation on the dependent variable ($t = T-1$) will equal the true values. This clearly holds more generally for a finite number of data revisions before the truth is revealed, and holds irrespective of the whether revisions are news or noise.[5]

2.1 News revisions

As the number of observations gets large, the estimated standard error $\hat{\sigma}_{T-1,EOS}$ from (2) will approach $\sqrt{\sigma_\eta^2 + \sigma_v^2}$. This is because $E\left(y_t^T - \beta y_{t-1}^T\right)^2 = E\left(y_t - \alpha y_{t-1}\right)^2 = E\left(\eta_t + v_t\right)^2 = \sigma_\eta^2 + \sigma_v^2$ for $t = \ldots, T-2$, and for $t = T-1$, $E\left(y_t^T - \beta y_{t-1}^T\right)^2 = E\left(y_t - v_t - \alpha y_{t-1}\right)^2 = E\left(\eta_t\right)^2 = \sigma_v^2$. The effect of the last observation will disappear as T gets large. The Box-Jenkins (BJ) $(1-\alpha)$

[5]Clements and Galvão (2013b) discuss the general case of models of order p, non-zero means in the true process and the revisions, and more than one data revision.

level prediction interval is given by:

$$\left\{\widehat{y_T}_{,EOS} + z_{\frac{\alpha}{2}}\hat{\sigma}_{T-1,EOS}, \quad \widehat{y_T}_{,EOS} + z_{1-\frac{\alpha}{2}}\hat{\sigma}_{T-1,EOS}\right\}$$

where z_γ is the γ quantile of the standard normal, $\gamma = \Phi(z_\gamma)$, and where Φ denotes the standard normal distribution function.

The expected squared error of the out-of-sample forecast is given by:

$$\begin{aligned} E\left(y_T^{T+1} - \widehat{y_T}_{,EOS}\right)^2 &= E\left(y_T - v_T - \alpha\left(y_{T-1} - v_{T-1}\right)\right)^2 \\ &= \sigma_\eta^2 + \alpha^2\sigma_v^2. \end{aligned} \tag{3}$$

Assuming that the true values follow a stationary AR(1), $|\alpha| < 1$, then the in-sample estimate of uncertainty ($\hat{\sigma}_{T-1,EOS}^2 = \sigma_\eta^2 + \sigma_v^2$) overstates the uncertainty surrounding the forecast of y_T^{T+1}. That is, the prediction intervals are too wide. The intuitive explanation is that the in-sample estimate is based on predicting the revised values, with added news relative to the first estimate, and so is accomplished with less precision than the forecasting of a first estimate (y_T^{T+1}) out-of-sample.

We have assumed that the target is the first estimate rather than the fully-revised value (in our setup, $y_T^{T+2} = y_T$).[6] The BJ interval would now under-estimate the true out-of-sample uncertainty and the actual coverage rate would be less than the nominal:

$$\begin{aligned} E\left(y_T - \widehat{y_T}_{,EOS}\right)^2 &= E\left(y_T - \alpha\left(y_{T-1} - v_{T-1}\right)\right)^2 \\ &= \sigma_\eta^2 + \left(1 + \alpha^2\right)\sigma_v^2. \end{aligned}$$

2.2 Noise revisions

Now suppose revisions reduce noise (and $\sigma_v^2 = 0$). Consider the in-sample fit of the model. For noise, $E\left(y_t^T - \beta y_{t-1}^T\right)^2 = E\left(y_t - \alpha y_{t-1}\right)^2 = E\left(\eta_t\right)^2 = \sigma_\eta^2$ for $t = \ldots, T-2$, and for $t = T-1$, $E\left(y_t^T - \beta y_{t-1}^T\right)^2 = E\left(y_t + \varepsilon_t - v_t - \alpha y_{t-1}\right)^2 = E\left(\eta_t + \varepsilon_t\right)^2 = \sigma_v^2 + \sigma_\varepsilon^2$, so for large T the in-sample error variance will be estimated as σ_η^2.

[6] Real-time forecasting exercises commonly assume that the goal is to forecast a relatively early vintage value, such as the value available one or two quarters after the reference quarter.

Consider now the out-of-sample expected squared error:

$$E\left(y_T^{T+1} - \widehat{y_T}_{,EOS}\right)^2 = E\left(y_T + \varepsilon_T - \alpha\left(y_{T-1} + \varepsilon_{T-1}\right)\right)^2$$
$$= \sigma_\eta^2 + \left(1 + \alpha^2\right)\sigma_\varepsilon^2 \quad (4)$$

which exceeds the in-sample error variance. In a reverse of the situation when revisions are news, when revisions reduce noise the in-sample error variance under-estimates the true out-of-sample uncertainty, and consequently the actual coverage of the BJ intervals will fall short of the nominal. Intuitively, when revisions remove noise, the fully-revised data used in the in-sample calculation will lead to an under-estimation of the uncertainty that characterizes the out-of-sample estimate. If instead we target the fully-revised value, $E\left(y_T - \widehat{y_T}_{,EOS}\right)^2 = \sigma_\eta^2 + \alpha^2\sigma_\varepsilon^2$, which still exceeds the in-sample estimate but to a lesser extent.

To summarize: prediction intervals will be too wide if data revisions are news (and the aim is to forecast an early vintage, otherwise they will be too narrow), but too narrow if revisions reduce noise.

In a pseudo out-of-sample forecasting exercise, e.g., using the data vintage $\left\{\mathbf{y}_{T-1}^{T+n}\right\}$, $n > 0$, to forecast y_T, there are no data vintage effects and in-sample and out-of-sample uncertainty would match save for small-sample effects and if the normal assumption were inappropriate.

A solution to the problem of obtaining correctly-sized prediction intervals in real-time is to use RTV-estimation. This was suggested by Koenig *et al.* (2003) for first-moment prediction, and further considered by Clements and Galvão (2013b) with an emphasis on autoregressive processes. In the following section we briefly discuss RTV-estimation, and show that it provides intervals with correct coverage.

3 RTV-estimation

US NIPA data are typically subject to revision for up to three and a half years after the first estimate is published. Koenig *et al.* (2003) note that EOS implies that a large part of the data used in model estimation has been revised many times, while the forecast is conditioned on data that has been just released or only revised a few times. That is, the data vector $\mathbf{y}_{T-1}^T = \left[\ldots, y_{T-2}^T, y_{T-1}^T\right]'$ comprises the first estimate of y_{T-1}, the first revision (i.e., the second estimate) of y_{T-2}, and so on up to mature data for the earlier data periods. They show that more accurate forecasts can be achieved (in principle) by not mixing mature and lightly-revised data, and instead advocate using 'real-time vintage' (RTV). The forecasting model is estimated on data of a similar maturity to the data on which the forecast is conditioned.[7]

For an AR(p) the forecast will be conditioned on early-release data ($\mathbf{y}^T_{T-1} = \left[y^T_{T-1}, y^T_{T-2}, \ldots y^T_{T-p} \right]'$. The RTV approach estimates the AR(p) on matching early-release data:

$$y^t_{t-1} = \beta_0 + \sum_{i=1}^{p} \beta_i y^{t-1}_{t-1-i} + e_{t,RTV}, \quad \text{for } t = \ldots, T-1, T, \tag{5}$$

and the forecast of y_T is $y_{T|T-1,RTV} = \beta_0 + \beta_1 y^T_{T-1} + \ldots + \beta_p y^T_{T-p}$. If we let $\beta = \left[\beta_1 \; \ldots \; \beta_p \right]'$, then Clements and Galvão (2013b) show that for a general model of data revisions, as in Jacobs and van Norden (2011), the solution (ϕ^*_0, ϕ^*) of:

$$\underset{\phi_0, \boldsymbol{\phi}}{\arg\min} E\left[\left(y^{T+1}_T - \phi_0 - \phi' \mathbf{y}^T_{T-1} \right)^2 \right] \tag{6}$$

is satisfied by the RTV-population values: $\beta_0 = \phi^*_0$, and $\beta = \phi^*$. That is, given that the forecast is conditioned on $\mathbf{y}^T_{T-1}$, RTV will deliver (in population) the values of the intercept and autoregressive parameters which minimize the expected squared error. Clements and Galvão (2013b) show that (ϕ^*_0,ϕ^*) depend on the nature of data revisions (news or noise, whether they are zero mean, etc.).

To show that prediction intervals are correctly-sized with RTV-estimation, we return to the simple example of section 2, that is, a zero-mean AR(1) with a single revision.

3.1 News revisions

The population value of β in the RTV-regression model:

$$y^t_{t-1} = \beta y^{t-1}_{t-2} + e_t$$

[7] This is Strategy 1 of Koenig *et al.* (2003), p.620, which we refer to throughout as RTV-estimation, or the use of RTV data.

is $\beta = Cov\left(y_{t-1}^{t}, y_{t-2}^{t-1}\right)/Var\left(y_{t-2}^{t-2}\right) = \alpha$.[8] Then the in-sample error variance is based on $Var\left(y_{t-1}^{t} - \alpha y_{t-2}^{t-1}\right)$ which on substituting for y_{t-1}^{t} and y_{t-2}^{t-1} is equal to $\sigma_\eta^2 + \alpha^2\sigma_v^2$. It is a simple matter to show that this equals the out-of-sample uncertainty: $E\left(y_T^{T+1} - y_{T|T-1,RTV}\right)^2 = E\left[y_T - v_T - \alpha\left(y_{T-1} - v_{T-1}\right)\right]^2 = \sigma_\eta^2 + \alpha^2\sigma_v^2$.

3.2 Noise revisions

When there are noise revisions, we can show that in population:

$$\beta = \phi^* = \alpha\frac{\sigma_y^2}{\sigma_y^2 + \sigma_\varepsilon^2}. \tag{7}$$

The in-sample error variance and the out-of-sample squared forecast error are equal, and prediction intervals based on the former will have correct conditional coverage. The in-sample error variance is:

$$Var\left(y_{t-1}^{t} - \gamma y_{t-2}^{t}\right) = Var\left[y_{t-1} + \varepsilon_{t-1} - \beta\left(y_{t-2} - \varepsilon_{t-2}\right)\right]$$

and the out-of-sample squared forecast error is:

$$Var\left(y_T^{T+1} - y_{T|T-1,RTV}\right)^2 = Var\left[y_T + \varepsilon_T - \beta\left(y_{T-1} + \varepsilon_{T-1}\right)\right].$$

The two variances are equal given the assumed stationarity of $\{y_t\}$ and $\{\varepsilon_t\}$.

Note that for both news and noise revisions, the equality of the in and out-of-sample error variances rests on forecasting the first estimate of y_T.

It follows directly that the equivalence of the in and out-of-sample error variances holds more generally for RTV-estimation. Clements and Galvão (2013b) argue that OLS estimation

[8] Note:

$$Cov\left(y_{t-1}^{t}, y_{t-2}^{t-1}\right) = E\left[\left(\alpha y_{t-2} + \eta_{t-1}\right)\left(y_{t-2} - v_{t-2}\right)\right] = \alpha\sigma_y^2 - \alpha\sigma_v^2,$$

where $\sigma_y^2 = Var\left(y_t\right)$, and

$$Var\left(y_{t-2}^{t-1}\right) = Var\left(y_{t-2} - v_{t-2}\right) = \sigma_y^2 + \sigma_v^2 - 2Cov\left(y_{t-2}, v_{t-2}\right) = \sigma_y^2 - \sigma_v^2$$

This result is a special case of the general formulae presented in Clements and Galvão (2013b). Further, the RTV parameter β only equals the autoregressive parameter of the true process (α) for the AR(1) with news, as here.

of (5) yields the same population values of the estimators as:

$$(\beta_0^*, \boldsymbol{\beta}^*) = \underset{\beta_0, \boldsymbol{\beta}}{\arg\min} E\left(y_t^{t+1} - \beta_0 - \boldsymbol{\beta}\mathbf{y}_{t-1}^t\right)^2. \tag{8}$$

A typical observation on the LHS and RHS variables in (5) is $\left\{y_t^{t+1}, \mathbf{y}_{t-1}^t = \left(y_{t-1}^t \dots y_{t-p}^t\right)'\right\}$, which is a covariance stationary process.

The estimation loss function (8) is identical to the real-time forecast loss function (6). Clements and Galvão (2013b) stress that the solutions to the two in terms of (ϕ_0, ϕ) and (β_0, β) coincide, $\beta_0^* = \boldsymbol{\phi}_\mathbf{0}^*$ and $\boldsymbol{\beta}^* = \boldsymbol{\phi}^*$, and thus RTV-estimation delivers optimal forecasts. For our purposes, note in addition that the minimized values of the functions in (6) and (8) are identical, implying in addition that the in-sample estimate of uncertainty from RTV-estimation will provide a reliable guide to out-of-sample forecasting. This holds for general revisions processes, such as that considered by Jacobs and van Norden (2011).

An alternative to using RTV data is to use models that draw on the multiple estimates of each observation which are typically available.[9] In terms of first-moment prediction, Clements and Galvão (2013b) compare the accuracy of RTV with forecasts from a vector autoregression (VAR), that models the relationships between the multiple-vintage estimates (in the spirit of recent work by Garratt *et al.* (2008, 2009)), and with the approach of Kishor and Koenig (2012), which specifies a model for the data revisions process which is estimated along side a VAR for the 'post-revision' data. They find the performance of these more elaborate models is on a par with RTV-estimation of the AR (for first-moment prediction). In this paper we do not examine the potential usefulness of these multiple-vintage models for calculating prediction intervals.

4 BJ prediction intervals for AR models estimated using RTV and EOS

We consider 25 US macro variables which are subject to data revisions. The variables are described in table 1. The data vintages are taken from the Real-Time Data Set for Macroeconomists (RTDSM) of Croushore and Stark (2001). Our first 'vintage-origin' is 1996:Q2, and

[9] Examples of multiple-vintage models include Harvey, McKenzie, Blake and Desai (1983), Howrey (1984), Patterson (1995, 2003), Jacobs and van Norden (2011), Cunningham *et al.* (2009) and Garratt *et al.* (2009, 2008).

the last is 2011:Q1, so that we have 15 years of quarterly forecast origins.[10] In order to estimate the models by EOS, we require that each data vintage provides a long enough history of past observations. We use a rolling-window forecasting scheme, where for the first vintage origin of 1996:Q2, we use data from 1984 onwards. However, for RTV, we need data vintages going back to 1984 to have data over the same historical period. That is, we require an additional 12 years of data-vintages for RTV estimation. This requirement was satisfied for the 25 variables forecast in this study.[11]

For all but one of the variables (*ruc* - the unemployment rate) we model, and evaluate forecasts of, the first difference of the natural logarithm. The variable *ruc* is modelled and forecast untransformed.

To focus on the RTV versus EOS issue, in all cases we use AR(2) models (i.e., two autoregressive lags), although the number of lags could be selected for each variable at each forecast origin using an information criterion such as BIC.[12] The real-time (EOS) forecasting performance of the AR models (estimated by EOS) has been shown to be improved by discarding the pre-Great Moderation data (Clements (2014)), which is the reason we set the initial start point of the estimation period to 1984. The actual values are taken from the vintage available one quarter after the target quarter. So for the forecast of 1996:Q2 from the first vintage origin of 1996:Q2, for example, this would be the value available in the 1996:Q3 vintage.

Table 2 records our first set of results: the relative magnitudes of the models' estimated standard errors for RTV and EOS estimation; t-statistics of the null that revisions are news; t-statistics of the null that revisions are noise; and the actual coverage rates of one-step ahead BJ intervals. The tests for news and noise are based on the revisions between the first-estimates y_t^{t+1}, and the data available some three and a half years later y_t^{t+15}, which includes the three rounds of annual regular revisions. We test for news and noise revisions using, respectively:

$$y_t^{t+1} - y_t^{t+15} = \alpha + \beta_{ne} y_t^{t+1} + \omega_t$$

[10] However, the 1999:Q4 and 2009:Q3 vintages contain missing values for many series and these forecast origins are excluded.

[11] For the Bureau of National Accounts variables the RTDSM of Croushore and Stark (2001) containing missing values for the 1996:Q1 estimates of 1995:Q4. These do not affect the EOS forecasts, because the first forecast origin is 1996:Q2, but these missing values do affect all RTV forecasts. We have simply set the values for 1995:Q4 to the 1995:Q3 values in the same (1996:Q1) vintage.

[12] We calculated the statistics reported in table 2 for models with an autoregressive order of one, and found the results were qualitatively unchanged (not reported to save space).

and:

$$y_t^{t+1} - y_t^{t+15} = \alpha + \beta_{no} y_t^{t+15} + \omega_t.$$

We find that 9 of the 25 variables have data revisions which are news, in that we do not reject $\beta_{ne} = 0$, but we do reject $\beta_{no} = 0$, at conventional levels.[13] For all of these variables the in-sample standard deviation estimated by EOS exceeds the RTV estimate. For the 7 variables for which we do not reject $\beta_{no} = 0$ but do reject $\beta_{ne} = 0$, implying noise revisions, we find the reverse - the EOS standard deviation is smaller than the RTV estimate. Hence for the 16 variables which can be categorized as news or noise, the relative magnitudes of the in-sample standard deviations are as expected given the analysis in section 3. The remaining variables cannot be characterized as purely news or noise, so it is not clear what one would expect to find.

In terms of out-of-sample performance, for around 80% of the variables the RTV intervals are more accurate than the EOS intervals, in the sense that the actual coverage rates are closer to the nominal rates. The actual coverage rates are shown for each variable for the RTV and EOS intervals in table 2, but summarizing across all variables, we find that the coverage rates of RTV intervals are closer to the nominal for 20, 21 and 18 variables for the 50%, 75% and 90% nominal intervals, respectively. Furthermore, our analysis suggests EOS-interval coverage should exceed that of RTV intervals for variables with news revisions, with the opposite holding for noise revisions. Of the 9 variables categorized as having news revisions, EOS-interval coverage is greater for either all, or all but one, of these variables (depending on the nominal interval size). Of the 7 variables with noise revisions, the RTV coverage rate is greater for all 7 variables for the 50% and 75% intervals (and for all but one for the 90% intervals).

RTV estimation is expected to generate more accurate BJ-intervals than EOS because the RTV-estimate of the in-sample standard deviation more accurately reflects the out-of-sample uncertainty. Table 2 is designed to show the relationship between the in-sample standard deviations, and the average interval coverage rates. However, obtaining the correct coverage 'on average' is a minimal requirement of a sequence of prediction intervals. To see this, let $L_{t|t-1}(p)$ and $U_{t|t-1}(p)$ denote the lower and upper limits of a 1-step ahead prediction interval

[13] Aruoba (2008) and Corradi, Fernandez and Swanson (2009) provide recent extensions to testing for the properties of data revisions.

with nominal coverage p, and let $I_t = 1$ denote a 'hit', defined as:

$$I_t = \begin{cases} 1 & \text{if } y_t \in \left(L_{t|t-1}(p), U_{t|t-1}(p)\right), \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

for a sequence of forecasts $(\{L_{t|t-1}(p), U_{t|t-1}(p)\})$ and realizations $(\{y_t\})$, $t = 1, 2, \ldots, N$. Correct unconditional coverage holds when $E(I_t) = p$, assessed by whether the sample mean $\frac{1}{N}\sum_{t=1}^{N} I_t$ is close to p. A more stringent criterion is that the occurrences of 1's and 0's are unpredictable - for a given information set $\mathcal{I}_t$ (where $\mathcal{I}_t = \{I_t, I_{t-1}, \ldots\}$ at a minimum) we require $E(I_t|\mathcal{I}_{t-1}) = p$. When $\mathcal{I}_t = \{I_t, I_{t-1}, \ldots\}$, this is equivalent to saying that $\{I_t\}$ is iid Bernoulli with parameter p. This is a joint test, and Christoffersen (1998) presents simple likelihood-based tests of the component parts: correct unconditional coverage ($E(I_t) = p$); independence (against a first-order Markov chain structure for $\{I_t\}$); as well as of the joint hypothesis of correct conditional coverage. The test for independence will have power to detect (unmodelled) changes in the volatility (because hits will tend to be clustered during the relatively low volatility periods) as well as dynamic mis-specification of the model generating the forecasts.[14]

Table 3 reports the p-values for the tests of correct unconditional coverage (UC), independence (IND) and conditional coverage (CC) for each variable, for intervals of three nominal sizes (50%, 75% and 90%). The bottom row of the table reports the number of variables for which the null of the corresponding test is rejected (when the test is conducted at the 5% level). The EOS intervals are rejected for more variables than the RTV intervals, and the rejections are mainly of correct unconditional coverage (or 'bias'), rather than of the test for independence. This indicates that many of the differences between the nominal and actual coverage rates recorded in table 2 for the EOS intervals are statistically significant.

Interval coverage rates may also differ because the interval is located about an inaccurate point forecasts, and not just because of the scale of the predictive distribution underlying the interval. However, not withstanding the superiority of RTV-estimation in principle for first-moment prediction (see Koenig *et al.* (2003) and Clements and Galvão (2013b)), for our setup

[14] As an example of the latter, prediction intervals generated by an AR(1) (say), when the data generating process is an AR(2) with gaussian disturbances, will have correct unconditional coverage but not correct conditional coverage. See Corradi and Swanson (2006) for a discussion in the context of density forecasting.

we find little to choose between the two for the majority of variables, when the point forecasts are evaluated by RMSE: see table 4.

Table 4 records the RMSEs of the point forecasts. We find that RTV improves accuracy by 5% on RMSE for forecasting output growth. But with one or two exceptions, this is at the top end of the gains to RTV, and a number of the entries exceed one, suggesting EOS is more accurate for those variables. This suggests that the RTV-intervals chiefly benefit from more accurate estimates of scale rather than location. The general point is that RTV estimation in practical forecasting may matter more for second-moment type forecasts (such as prediction intervals) than for point forecasting.

This last point is supported by the results in table 4 for bivariate ADL (autoregressive-distributed lag) models. For each of the 25 variables we generate EOS and RTV point forecasts for the 24 possible bivariate ADL models, where we include 2 lags of the dependent variable, and of the explanatory variable. This allow us to assess the relative point forecasting performance of RTV compared to EOS for models with explanatory variables which are subject to revision. The results for the ADL models are no more favourable for RTV than the results for AR models. We present the median, mean, minimum and maximum of the ratios across the 24 ADL models for each variable. There is an ADL model (i.e., an explanatory variable) for which RTV is 10% more accurate than EOS for forecasting output growth. But equally there is a variable for which EOS is 9% more accurate for forecasting output growth, and this pattern holds across the majority of variables.

5 Simulation study

We established in section 3 that BJ intervals based on RTV-estimation would be correctly-sized in the presence of data revisions, and that intervals based on EOS-estimation of the model would likely have a coverage in excess of the nominal when data revisions are news, but less than the nominal when data revisions are noise. These statements hold 'in population', that is, when we ignore parameter estimation uncertainty.[15] Even in the absence of data revisions, it is well known that BJ-intervals can be adversely affected by parameter estimation uncertainty

[15] Note that the empirical estimates of coverage rates in section 4 are generally in line with the analysis in section 3.

when the sample size is small. In addition, there will be uncertainty about the model order. Having to select appropriate model orders may affect the relative merits of RTV and EOS. Model selection issues are especially pertinent in our context because the optimal model order for RTV-estimation may differ from that using EOS-estimation. To illustrate with a simple case, consider the example in section 3, where the true values follow a zero-mean AR(1) and there is a single noise revision. Hence the data generating process is given by:

$$y_t = \alpha y_{t-1} + \eta_t \tag{10}$$

and the estimates of y_t are given by:

$$\begin{aligned} y_t^{t+1} &= y_t + \varepsilon_t \\ y_t^{t+n} &= y_t \end{aligned} \tag{11}$$

for $n = 2, 3, \ldots$. Let the population first-order RTV-regression be:

$$y_t^{t+1} = \beta y_{t-1}^t + e_{t,RTV}. \tag{12}$$

where from (7) we know that $\beta \neq \alpha$ when $E\left(\varepsilon_t^2\right) \neq 0$ (so $\beta \neq \alpha$ when there are noise revisions). Substituting for y_t^{t+1} and y_{t-1}^t from (11) into (12) yields $e_{t,RTV} = y_t - \beta y_{t-1} + \varepsilon_t - \beta\varepsilon_{t-1}$. Then it follows immediately that the first-order model is dynamically-mis-specified because the second 'lag' y_{t-2}^t (here equal to y_{t-2}) is correlated with the error term $e_{t,RTV}$.[16]

For these reasons we investigate the small-sample properties of the procedures by simulation, allowing that the appropriate model orders need to be selected by an information criterion, such as BIC. Of interest is whether RTV-estimation provides accurate intervals in these circumstances.

Simulating data with revisions requires a complete specification of the data revisions process. This requires giving values to a relatively large number of parameters, and the concomitant concern that the results of the simulation study may be specific to the set of values chosen (or

[16]
$$Cov\left(e_{t,RTV}, y_{t-2}\right) = Cov\left(y_t - \beta y_{t-1}, y_{t-2}\right) + Cov\left(\varepsilon_t - \beta\varepsilon_{t-1}, y_{t-2}\right).$$
The second covariance on the RHS is zero, and the first is only zero when $\beta = \alpha$, in which case $y_t - \beta y_{t-1} = \eta_t$.

more generally, to the range of values considered). In order to obtain sensible values for the data generating process, we use the model estimated by Jacobs and van Norden (2011) for US real output growth as our base case. They allow for data revisions to be news or noise, with and without 'spillovers'. We also experiment with a variant in which the standard deviation of the underlying shock is divided by 4, but the standard deviations of the news and noise disturbances (and all the other parameters) are left unaltered, to gauge the impact of data revisions being more prominent (than they are for real output, as estimated by Jacobs and van Norden (2011)). To save space, we do not repeat the details of their model, except to note that there are four vintage estimates of each time period, and the final is not assumed to reveal the truth (i.e., $y_t^{t+4} \neq y_t$). The estimated parameter values that we use are taken from their Table 1 (p.107). We simulate $25,000$ replications of length $T+1$,[17] and on each sample we estimate the AR model by EOS and RTV on the first T observations, and calculate a BJ prediction interval for the first estimate of the $T+1$ observation (i.e., y_{T+1}^{T+2}).

The results are recorded in table 5. For the Jacobs and van Norden (2011) data generating process (top half of table) the RTV intervals are under-sized at small T but approximately correctly-sized at larger T. The coverage rates are similar (for a given T) whether revisions are news of noise, and irrespective of whether there are spillovers: RTV-intervals are largely immune to the effects of data revisions. This is underlined by comparing these estimated coverage rates with those when there are no revisions (recorded at the foot of the table): the two sets are virtually identical.[18] By contrast, EOS-estimation is not a reliable method of generating prediction intervals, giving rise to intervals which are under-sized or over-sized, depending on T, when revisions are news. When revisions are more prominent (second half of table), the performance of the EOS intervals worsens, and the EOS intervals are clearly under-sized when there is noise. Whereas the performance of the RTV intervals is virtually unaffected. There is little appreciable effect from allowing spillovers.

Table 6 records the rejection frequencies of the tests for unconditional coverage, independence and conditional coverages. For the RTV intervals, there are some minor size distortions

[17] That is, after discarding a suitable number of initial observations to remove any dependence on initial conditions.

[18] It is well known in the literature that neglecting the uncertainty inherent in the estimated model parameters will lead to under-sized intervals, and this effect is obviously more pronounced at small T.

at the smaller sample sizes,[19] but otherwise the tests are correctly sized, while the EOS intervals are clearly inadequate, and this is flagged by the tests of coverage. For the selected Monte Carlo data-generation process parameter values, the rejections are greater for noise revisions, and for the 75% and 90% nominal intervals are in excess of 50%. We conclude that prediction intervals from RTV-estimation with BIC model-order specification generates intervals with desirable properties, whilst the traditional approach (EOS-estimation) does not.

6 Conclusions

We have shown that assessments of future macroeconomic uncertainty based on the in-sample fit of a model are likely to be misleading when the variable being modelled is subject to revision. This is because the data on which the model is estimated will be for the most part fully-revised or mature data, whereas the out-of-sample value is a first-release or only lightly-revised data point. In the context of first-moment prediction Kishor and Koenig (2012) referred to estimating the model on fully-revised data, and conditioning the forecast on only lightly-revised data, as mixing 'apples and oranges'. In the context of second-moment prediction the mismatch results instead from supposing the goodness of fit of the model on the fully-revised data (i.e., the in-sample period) is an accurate representation of the out-of-sample fit.

We have shown that a simple solution is to use real-time-vintage (RTV) data. This was proposed by Koenig *et al.* (2003) in the context of first-moment prediction. Based on the evidence for the 25 macro variables we consider in this paper, RTV-estimation is more beneficial for second-moment forecasting. Its validity in population (abstracting from parameter estimation uncertainty) is easily established. Its good forecast performance has been demonstrated, and is supported by a simulation study.

We have considered autoregressive models in this paper, but the logic of the arguments suggests that the findings carry over immediately to models with explanatory variables, so the concerns are potentially more generally applicable. For example, there has been much interest in measuring macro-uncertainty in the recent literature, driven in large part by the

[19] We generate sequences of 100 prediction intervals and actuals. The initial sample size is recorded in the table, and the model is re-estimated (by EOS or RTV) on an expanding window of data prior to calculating the interval. For the resulting vector of hits and misses, we calculate the three tests. The table records the rejection frequences across 10,000 replications of this procedure.

belief that time-varying uncertainty may play an important role in business cycle fluctuations (see, e.g., Bachmann, Elstner and Sims (2013)). Some of the approaches to measuring macro-uncertainty use 'data-rich' modelling environments, and are pseudo real-time, perhaps because of the difficulties of collecting and managing different data vintages at each point in time. Two recent closely-related contributions are Jurado, Ludvigson and Ng (2013) and Henzel and Rengel (2013). These papers are pseudo real-time, but the measures of macro uncertainty are derived from the individual variables' forecast errors, rather than the in-sample fits of the models, and so should be immune to the distorting effects described in this paper.[20] Other approaches which use instead the in-sample fit of the model - as in the standard approach to calculating prediction intervals illustrated in this paper - are likely to be misleading in the presence of data revisions.

[20] For example, Jurado *et al.* (2013) calculate forecast errors for each of a large number of variables using a factor model, and then fit a stochastic volatility model to the forecast errors to obtain individual-variable volatilty forecasts, which are then aggregated.

References

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, **40**, 319–340.

Bachmann, R., Elstner, S., and Sims, E. R. (2013). Uncertainty and economic activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics*, **5(2)**, 217–249.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**, 841–862.

Clements, M. P. (2014). Real-time factor model forecasting and offsetting the effects of instability. Working paper, ICMA Centre, University of Reading.

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 and & 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.

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.

Corradi, V., and Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, *133*(2), 779–806.

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.

Faust, J., Rogers, J. H., and Wright, J. H. (2003). Exchange rate forecasting: The errors we've really made. *Journal of International Economic Review*, **60**, 35–39.

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.

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.

Henzel, S., and Rengel, M. (2013). Dimensions of macroeconomic uncertainty: A common factor analysis. Ifo working paper series, Ifo Institute for Economic Research at the University of Munich.

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.

Jurado, K., Ludvigson, S. C., and Ng, S. (2013). Measuring Uncertainty. mimeo, Department of Economics, Columbia University.

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.

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.

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

Orphanides, A., and van Norden, S. (2005). The reliability of inflation forecasts based on output gaps in real time. *Journal of Money, Credit and Banking*, **37**, 583–601.

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.

Robertson, J. C., and Tallman, E. W. (1998). Data vintages and measuring forecast model performance. *Federal Reserve Bank of Atlanta Economic Review, Fourth Quarter*, 4–20.

Thombs, L. A., and Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, **85**, 486–492.

Table 1: Data Description

Mnemonic	Description
routputq	Real GNP/GDP
rconq	Real Personal Consumption Expenditures Total
rconndq	Real Personal Consumption Expenditures: Nondurable Goods
rcondq	Real Personal Consumption Expenditures: Durable Goods
rinvbfq	Real Gross Private Domestic Investment: Non-residential
rinvresq	Real Gross Private Domestic Investment: Residential
rexq	Real Exports of Goods and Services
rimpq	Real Exports of Goods and Services
rgq	Real Government Consumption and Gross Investment: Total
ruc	Unemployment Rate
pq	Price Index for GNP/GDP
pconq	Price Index for Personal Consumption Expenditures
pimpq	Price Index for Imports of Goods and Services
noutputq	Nominal GNP/GDP
nconq	Nominal Personal Consumption Expenditures
wsdq	Wage and Salary Disbursements
oliq	Other Labor Income
propiq	Proprietors' Income
divq	Dividends
pintiq	Personal Interest Income
tranrq	Transfer Payments
sscontrq	Personal Contributions for Social Insurance
npiq	Nominal Personal Income
ptaxq	Personal Tax and Nontax Payments
ndpiq	Nominal Disposable Personal Income

Source: The Real-Time Data Set for Macroeconomists (RTDSM), http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/, see Croushore and Stark (2001).

Table 2: RTV and EOS Box-Jenkins Prediction Intervals Coverage Rates for RTV and EOS

	Ratio RTV	t- stat	t- stat	50% interval		75% interval		90% interval	
	to EOS sd.	for 'news'	for 'noise'	RTV	EOS	RTV	EOS	RTV	EOS
routputq	0.87	2.19	-4.23	0.50	0.52	0.72	0.78	0.86	0.93
rconq	1.14	6.29	-0.16	0.52	0.48	0.74	0.66	0.84	0.88
rconndq	1.27	9.46	-0.13	0.48	0.40	0.72	0.60	0.86	0.79
rcondq	1.07	6.44	1.52	0.52	0.45	0.69	0.62	0.83	0.74
rinvbfq	1.18	2.11	-4.53	0.55	0.48	0.76	0.64	0.90	0.86
rinvresq	1.14	2.17	-3.15	0.52	0.43	0.71	0.66	0.84	0.78
rexq	1.04	3.84	-3.32	0.52	0.36	0.72	0.67	0.86	0.84
rimpq	1.13	5.72	-3.20	0.43	0.38	0.62	0.60	0.84	0.72
rgq	1.12	7.50	-1.74	0.66	0.59	0.81	0.78	0.91	0.91
ruc	1.22	0.49	-0.12	0.50	0.41	0.69	0.60	0.79	0.71
pq	1.17	2.60	-1.51	0.52	0.41	0.72	0.66	0.90	0.79
pconq	1.18	3.79	0.87	0.45	0.40	0.67	0.59	0.83	0.79
pimpq	1.10	3.23	-2.68	0.43	0.41	0.78	0.71	0.86	0.81
noutputq	0.96	0.76	-5.43	0.55	0.57	0.76	0.74	0.90	0.91
nconq	1.13	4.91	-0.67	0.62	0.57	0.86	0.74	0.91	0.88
wsdq	0.46	1.29	-10.13	0.64	0.81	0.81	0.91	0.90	0.98
oliq	0.76	1.63	-12.21	0.62	0.71	0.83	0.88	0.88	0.95
propiq	1.13	4.28	-5.74	0.62	0.69	0.86	0.83	0.95	0.90
divq	0.69	1.76	-7.44	0.76	0.86	0.81	0.90	0.88	0.93
pintiq	0.62	0.78	-9.13	0.47	0.66	0.64	0.81	0.78	0.93
tranrq	0.86	1.31	-2.17	0.59	0.55	0.76	0.79	0.83	0.86
sscontrq	1.08	1.51	-1.72	0.60	0.62	0.74	0.79	0.86	0.84
npiq	0.69	0.63	-7.81	0.59	0.74	0.78	0.84	0.86	0.91
ptaxq	0.78	-0.84	-5.59	0.59	0.71	0.83	0.84	0.86	0.88
ndpiq	0.83	1.37	-5.97	0.59	0.67	0.83	0.88	0.90	0.91

The second column reports the ratio of the estimated in-sample standard errors for RTV to EOS. (The EOS standard deviation is the average of the standard deviations calculated for each of the 60 rolling window estimation samples. The RTV sd. is calculated similarly). The third and fourth columns record the t-statistics for tests that the 1st-estimate to 15th-estimate revisions are news and noise, respectively. These tests are run once for each variable and relate to observation periods 1970:Q2 to 2007:Q2. The remaining columns record coverage rates for nominal coverages of 50%, 75% and 90%.
For both AR and EOS the model is an AR(2), estimated on forecasts (vintage) origins 1996:Q2 to 2011:Q1, using a rolling window of observations (and an initial window estimated on post 1984 observations).

Table 3: Formal tests of RTV and EOS prediction intervals

	Nominal Coverage 50%						Nominal Coverage 75%						Nominal Coverage 90%					
	RTV			EOS			RTV			EOS			RTV			EOS		
	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC
routputq	1.00	0.90	0.99	0.79	0.89	0.96	0.65	0.24	0.45	0.65	0.14	0.30	0.36	0.07	0.12	0.41	0.23	0.35
rconq	0.79	0.69	0.89	0.79	0.67	0.88	0.88	0.51	0.80	0.11	0.23	0.13	0.19	0.15	0.15	0.61	0.21	0.40
rconndq	0.79	0.67	0.88	0.11	0.40	0.20	0.65	0.08	0.20	0.01	0.69	0.05	0.36	0.89	0.65	0.02	0.71	0.05
rcondq	0.79	0.70	0.89	0.43	0.39	0.51	0.30	0.67	0.53	0.03	0.16	0.04	0.09	0.82	0.24	0.00	0.97	0.00
rinvbfq	0.43	0.65	0.66	0.79	0.52	0.79	0.88	0.08	0.21	0.06	0.20	0.07	0.93	0.01	0.03	0.36	0.37	0.44
rinvresq	0.79	0.01	0.04	0.29	0.18	0.23	0.46	0.07	0.15	0.11	0.00	0.00	0.19	0.15	0.15	0.01	0.14	0.01
rexq	0.79	0.35	0.63	0.03	0.57	0.09	0.65	0.11	0.25	0.19	0.12	0.12	0.36	0.37	0.44	0.19	0.15	0.15
rimpq	0.29	0.11	0.16	0.06	0.01	0.01	0.03	0.95	0.09	0.01	0.63	0.04	0.19	0.15	0.15	0.00	0.89	0.00
rgq	0.02	0.32	0.04	0.19	0.69	0.39	0.27	0.95	0.55	0.65	0.67	0.82	0.72	0.31	0.56	0.72	0.23	0.46
ruc	1.00	0.23	0.49	0.19	0.30	0.25	0.30	0.07	0.11	0.01	0.05	0.01	0.02	0.04	0.01	0.00	0.11	0.00
pq	0.79	0.51	0.78	0.19	0.07	0.08	0.65	0.74	0.86	0.11	0.23	0.13	0.93	0.63	0.88	0.02	0.71	0.05
pconq	0.43	0.39	0.51	0.11	0.53	0.24	0.19	0.12	0.12	0.01	0.63	0.02	0.09	0.06	0.04	0.02	0.06	0.01
pimpq	0.29	0.02	0.04	0.19	0.21	0.19	0.65	0.11	0.25	0.46	0.49	0.60	0.36	0.11	0.18	0.04	0.02	0.01
noutputq	0.43	0.20	0.32	0.29	0.98	0.57	0.88	0.56	0.83	0.88	0.17	0.39	0.93	0.10	0.26	0.72	0.41	0.67
nconq	0.06	0.78	0.17	0.29	0.60	0.50	0.04	0.37	0.07	0.88	0.48	0.77	0.72	0.41	0.67	0.61	0.87	0.87
wsdq	0.03	0.72	0.10	0.00	0.00	0.00	0.27	0.09	0.13	0.00	0.04	0.00	0.93	0.06	0.18	0.01	0.85	0.04
oliq	0.06	0.95	0.18	0.00	0.49	0.00	0.16	0.56	0.31	0.01	0.16	0.02	0.61	0.74	0.83	0.18	0.56	0.35
propiq	0.06	0.16	0.07	0.00	0.70	0.01	0.04	0.89	0.11	0.16	0.82	0.36	0.18	0.11	0.11	0.93	0.63	0.88
divq	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.61	0.00	0.01	0.41	0.01	0.03
pintiq	0.60	0.05	0.12	0.02	0.00	0.00	0.06	0.01	0.00	0.27	0.47	0.42	0.01	0.00	0.00	0.41	0.23	0.35
tranrq	0.19	0.21	0.19	0.43	0.75	0.70	0.88	0.69	0.91	0.44	0.26	0.39	0.09	0.28	0.13	0.36	0.37	0.44
sscontrq	0.11	0.00	0.00	0.06	0.03	0.02	0.88	0.37	0.66	0.44	0.19	0.31	0.36	0.04	0.08	0.19	0.10	0.11
npiq	0.19	0.21	0.19	0.00	0.00	0.00	0.65	0.00	0.01	0.08	0.10	0.06	0.36	0.28	0.37	0.72	0.41	0.67
ptaxq	0.19	0.13	0.14	0.00	0.00	0.00	0.16	0.21	0.16	0.08	0.00	0.00	0.36	0.07	0.12	0.61	0.21	0.40
ndpiq	0.19	0.30	0.25	0.01	0.32	0.02	0.16	0.82	0.36	0.01	0.02	0.00	0.93	0.10	0.26	0.72	0.41	0.67
5% Rejs	3.00	5.00	5.00	9.00	7.00	10.00	3.00	3.00	3.00	9.00	5.00	11.00	2.00	5.00	5.00	9.00	2.00	8.00

The elements of the table (bar those in the bottom row) are p-values of the the null of correct unconditional coverage (UC), independence (IND) and conditional coverage (CC). The bottom row elements are the numbers of variables for which the p-values in the corresponding columns are less than 0.05.

Table 4: RTV and EOS forecasts with AR models and ADL

	RMSFE	Ratio of RMSFE of RTV to EOS				
	AR	AR	ADL_{med}	ADL_{mea}	ADL_{min}	ADL_{max}
routputq	0.53	0.95	0.98	0.98	0.90	1.09
rconq	0.52	1.07	1.08	1.07	0.97	1.14
rconndq	0.85	1.03	1.03	1.04	0.99	1.14
rcondq	2.90	1.00	1.01	1.01	0.97	1.07
rinvbfq	2.49	1.00	1.02	1.03	0.96	1.11
rinvresq	3.30	0.95	0.96	0.97	0.93	1.11
rexq	2.48	0.98	0.98	0.99	0.94	1.10
rimpq	2.56	1.01	1.03	1.03	0.94	1.12
rgq	0.75	1.02	1.05	1.08	1.00	1.63
ruc	0.29	1.03	1.04	1.05	1.00	1.12
pq	0.25	0.99	1.00	1.02	0.92	1.40
pconq	0.44	1.00	1.00	1.01	0.97	1.09
pimpq	2.51	1.03	1.04	1.05	1.00	1.14
noutputq	0.55	0.96	0.98	0.99	0.93	1.16
nconq	0.69	0.97	1.00	1.00	0.96	1.07
wsdq	0.57	0.78	0.78	0.78	0.62	0.92
oliq	0.50	0.98	0.96	0.96	0.85	1.16
propiq	1.22	1.13	1.14	1.17	1.07	1.77
divq	7.09	1.32	1.36	1.36	1.25	1.46
pintiq	1.29	1.00	0.96	0.95	0.82	1.11
tranrq	1.66	0.98	0.98	0.99	0.92	1.06
sscontrq	1.56	1.00	1.01	1.01	0.96	1.06
npiq	0.54	0.99	1.00	1.00	0.81	1.10
ptaxq	4.94	1.03	1.04	1.04	0.95	1.15
ndpiq	0.81	1.09	1.09	1.09	0.96	1.31

Notes. The table reports the RMSEs for a rolling forecasting scheme, with initial window beginning in 1984:Q1. Out-of-sample forecast period is 1996:Q2 to 2011:Q1.

Table 5: Monte Carlo of small-sample coverage rates of EOS and RTV intervals

	50%		75%		90%	
	EOS	RTV	EOS	RTV	EOS	RTV
Jacobs and van Norden (2011) parameter values						
A. News, no spillovers						
15	0.45	0.43	0.68	0.66	0.83	0.81
25	0.49	0.47	0.73	0.71	0.88	0.86
50	0.51	0.48	0.76	0.73	0.91	0.88
200	0.53	0.50	0.78	0.75	0.92	0.90
B. Noise, no spillovers						
15	0.43	0.43	0.65	0.66	0.81	0.81
25	0.46	0.46	0.70	0.70	0.85	0.86
50	0.47	0.48	0.72	0.73	0.88	0.88
200	0.49	0.50	0.74	0.75	0.89	0.90
C. News & spillovers						
15	0.46	0.43	0.70	0.66	0.84	0.81
25	0.51	0.47	0.75	0.71	0.89	0.86
50	0.53	0.49	0.78	0.73	0.92	0.88
200	0.54	0.50	0.79	0.75	0.93	0.90
D. Noise & spillovers						
15	0.43	0.43	0.65	0.66	0.81	0.81
25	0.45	0.46	0.70	0.70	0.85	0.86
50	0.47	0.48	0.72	0.73	0.87	0.88
200	0.49	0.50	0.74	0.75	0.89	0.90
Disturbance s.d. multiplied by 0.25						
A. News, no spillovers						
15	0.46	0.42	0.69	0.64	0.84	0.79
25	0.52	0.47	0.76	0.70	0.90	0.86
50	0.54	0.48	0.79	0.73	0.92	0.88
200	0.56	0.50	0.80	0.75	0.94	0.89
B. Noise, no spillovers						
15	0.38	0.43	0.59	0.65	0.75	0.81
25	0.40	0.47	0.63	0.71	0.79	0.86
50	0.41	0.48	0.65	0.73	0.81	0.89
200	0.43	0.50	0.66	0.75	0.82	0.89
C. News & spillovers						
15	0.48	0.42	0.71	0.64	0.85	0.80
25	0.53	0.47	0.77	0.70	0.91	0.86
50	0.55	0.48	0.80	0.73	0.93	0.88
200	0.57	0.50	0.82	0.74	0.94	0.89
D. Noise & spillovers						
15	0.37	0.43	0.58	0.65	0.74	0.81
25	0.40	0.47	0.62	0.71	0.79	0.86
50	0.41	0.48	0.64	0.73	0.81	0.89
200	0.42	0.50	0.65	0.75	0.82	0.89

When there are no revisions, the coverage rates are: 0.43, 0.46, 0.48, 0.50 (for a nominal 50%, for estimation samples 15 to 200); 0.66, 0.70, 0.73, 0.75 (for a nominal 75%); and 0.82, 0.86, 0.88, 0.89 (for a nominal 90%).

Table 6: Monte Carlo of tests of RTV and EOS prediction intervals

	RTV									EOS								
	50%			75%			90%			50%			75%			90%		
	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC	UC	IND	CC
	A. News, no spillovers																	
15	0.09	0.06	0.07	0.11	0.06	0.09	0.12	0.06	0.11	0.14	0.06	0.10	0.14	0.07	0.14	0.12	0.05	0.10
25	0.08	0.06	0.07	0.09	0.06	0.08	0.09	0.05	0.08	0.16	0.06	0.13	0.16	0.06	0.16	0.15	0.04	0.11
50	0.07	0.06	0.06	0.07	0.06	0.07	0.07	0.05	0.07	0.18	0.06	0.13	0.20	0.06	0.18	0.19	0.03	0.13
200	0.06	0.05	0.05	0.06	0.05	0.06	0.05	0.04	0.05	0.21	0.05	0.15	0.25	0.06	0.22	0.24	0.02	0.16
	B. Noise, no spillovers																	
15	0.08	0.05	0.06	0.09	0.05	0.08	0.10	0.05	0.09	0.45	0.05	0.34	0.64	0.06	0.55	0.67	0.07	0.62
25	0.07	0.05	0.05	0.07	0.05	0.07	0.07	0.05	0.07	0.42	0.05	0.33	0.60	0.05	0.51	0.64	0.07	0.59
50	0.06	0.05	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.40	0.05	0.30	0.58	0.05	0.48	0.61	0.07	0.55
200	0.05	0.05	0.05	0.05	0.06	0.05	0.04	0.04	0.04	0.36	0.06	0.27	0.55	0.05	0.44	0.58	0.07	0.52
	C. News & spillovers																	
15	0.09	0.06	0.07	0.11	0.06	0.09	0.12	0.06	0.11	0.19	0.06	0.15	0.20	0.07	0.19	0.17	0.05	0.13
25	0.08	0.06	0.07	0.08	0.06	0.08	0.09	0.06	0.09	0.22	0.06	0.17	0.24	0.07	0.22	0.22	0.04	0.15
50	0.07	0.06	0.06	0.07	0.06	0.06	0.07	0.05	0.07	0.25	0.05	0.18	0.29	0.06	0.26	0.27	0.03	0.18
200	0.06	0.05	0.05	0.06	0.06	0.06	0.05	0.04	0.04	0.29	0.05	0.21	0.35	0.06	0.31	0.34	0.02	0.23
	D. Noise & spillovers																	
15	0.07	0.05	0.06	0.09	0.06	0.08	0.10	0.06	0.09	0.48	0.06	0.38	0.68	0.06	0.59	0.71	0.07	0.66
25	0.07	0.05	0.06	0.08	0.06	0.07	0.07	0.05	0.07	0.46	0.05	0.36	0.65	0.05	0.55	0.69	0.07	0.64
50	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05	0.44	0.06	0.34	0.63	0.06	0.53	0.66	0.07	0.60
200	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.04	0.05	0.42	0.06	0.31	0.61	0.05	0.50	0.64	0.07	0.58

The table reports the proportion of rejections of the null for tests carried out at a 5% significance level. The data-generating process is based on the Jacobs and van Norden (2011) parameter values, but with the underlying disturbance standard deviation multiplied by 0.25.