Data Revisions and Real-time Probabilistic Forecasting of Macroeconomic Variables

Michael P. Clements
ICMA Centre
Henley Business School
University of Reading
m.p.clements@reading.ac.uk

Ana Beatriz Galvao
Warwick Business School
University of Warwick
ana.galvao@wbs.ac.uk

January 10, 2017

Abstract

Macroeconomic data are subject to revision over time as later vintages are released, yet the usual way of generating real-time out-of-sample forecasts from models effectively makes no allowance for this form of data uncertainty. We analyze a simple method which has been used in the context of point forecasting, and does make an allowance for data uncertainty. This method is applied to density forecasting in the presence of time-varying heteroscedasticity, and is shown in principle to improve real-time density forecasts. We show that the magnitude of the expected improvements depends on the nature of the data revisions.

Keywords: real-time forecasting, inflation and output growth predictive densities, real-time-vintages, time-varying heteroscedasticity.

JEL code: C53.

1 Introduction

Decision makers employ probabilistic forecasts of macroeconomic variables to compute the probability of future outcomes of interest as an aid to determining which course of action to take. For example, based on density forecasts, one can quantify the probability of sluggish growth (say lower than 1%) and/or of deflation to support a monetary policy decision. This paper considers whether it is important to make an allowance for data uncertainty when computing probabilistic forecasts in real-time, given that most macroeconomic variables are subject to revision.

Clements (2017) considers the impact of data revisions on the assessment of macroeconomic forecasting uncertainty. He shows that the standard real-time approach, which estimates the forecasting model on the vintage of data available at the forecast origin, will likely give an inaccurate assessment of the uncertainty surrounding future values of the variables, especially of the early-vintage estimates of those values. Clements (2017) is mainly concerned with the coverage rates of prediction intervals obtained from autoregressive models with constant variance disturbances.

The recent literature on macroeconomic forecasting strongly supports the use of models with stochastic volatility to improve density forecasts in real time (see, for example, Clark (2011), Clark and Ravazzolo (2015) and Diebold, Schorfheide and Shin (2016)). Although the papers use real-time data, that is, macroeconomic time series actually available at the time the forecast is computed, they do not explicitly consider the impact of data revisions on the measurement of forecasting uncertainty, which underlies their probabilistic forecasts. In general, the conventional real-time data approach of using the vintage of data available at the forecast origin fails to make an allowance for data uncertainty, in the sense that it takes the data as given and does not allow for the consequences of the data being revised over time.

In this paper, we consider the impact of data uncertainty when computing one-step-ahead probabilistic forecasts in real-time using models with constant disturbance variances, and models with disturbances characterized by time-varying volatility. We compare two ways of using the real-time dataset to estimate the model and compute density forecasts. The first one is the conventional approach that uses the latest vintage of the time series available at each point in time. This makes no allowance for data uncertainty. It is sometimes known as end-of-sample, abbreviated to EOS. The second is the use of real-time-vintage (RTV) data, advocated by Koenig, Dolmas and Piger (2003) and Clements and Galvão (2013b) for point forecasting, and shown by Clements (2017) to be a simple and effective way of delivering more accurate assessments of forecasting uncertainty. RTV is designed to provide accurate forecasts in the presence of such revisions.[1]

We use both absolute tests for correct specification and relative measures of density fore-

[1] For brevity, we refer to 'the EOS approach' or 'EOS estimation' etc. as simply EOS, and similarly for RTV.

casting performance to compare the two approaches.[2] We show that RTV maximizes the out-of-sample real-time log score, that is, it delivers predictive densities 'nearer' to the true densities of the first-release values. Hence in empirical comparisons of density forecasting accuracy using log score, we would expect RTV to be more accurate than the conventional approach. This advantage of using RTV holds if data revisions are assumed to be either news (add new information) or noise (reduce measurement error) in the sense of Mankiw and Shapiro (1986). It also holds if the forecasting model allows for time-varying volatilities. We also establish that the loss to using the conventional approach identified by Clements (2017) is diminished by the presence of conditional heteroscedasticity in some circumstances.

We consider three commonly-used models of time-varying volatility: ARCH and GARCH models (see, *inter alia*, Engle (1982) and Bollerslev (1986)) and stochastic volatility (SV) models (e.g., Shephard (1996), Kim, Shephard and Chib (1998)). Although we are mainly concerned with the implications of data uncertainty for density forecasting, our results may also have implications for the literature on measuring macroeconomic uncertainty, as in Jurado, Ludvigson and Ng (2015) and Henzel and Rengel (2013). Jurado *et al.* (2015) use stochastic volatility models to measure forecasting uncertainty in a large set of macroeconomic time series which are in fact subject to revision, although they disregard the effects of data revisions by using the data vintage available at the time the study is undertaken. That is, their study is pseudo real-time, in the sense that at all but the latest forecast origin use is made of a vintage of the data that would not have been available.

Our analysis of the relative performance of RTV and EOS relies on simple autoregressive models, but we also compare VAR models in the empirical exercise. In the context of point forecasting 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 modelling multiple vintages of data, as with vintage-based vector autoregressive models (see, e.g., Patterson (1995, 2003), Clements and Galvão (2013a)). It might be of interest to apply some of these models to density forecasting, but we do not consider that here.

We also show that although RTV is superior to EOS in principle (in terms of log scores), the relative gains to RTV are only likely to be appreciable when revisions are noise (as opposed to news), when they are large (compared to the variance of the underlying disturbance), and when the process is reasonably persistent. Hence we are able to characterise the conditions under which revisions are expected to matter, and when they can reasonably be ignored, and the conventional approach to real-time forecasting can be used.

A confounding factor is that (unmodelled) breaks in the variance of the process for the true values may mask the benefits of RTV relative to EOS. This is relevant for empirical forecasting

[2] For a survey of these techniques, see Corradi and Swanson (2006).

comparisons which span such breaks, of which the Great Moderation is often regarded as a key example, dating an important downward shift in volatility (see, e.g., McConnell and Perez-Quiros (2000) and Sensier and van Dijk (2004)).

We also show using a simulation exercise that real-time density forecasting improvements from using RTV in comparison to EOS also hold for alternative loss functions, such as the CRPS and a quantile-forecast-based function.

Whether the expected gains to RTV based on our analysis materialize in practice is ultimately an empirical matter. We find that the neglect of data uncertainty (e.g., via EOS) may matter less when variables are subject to time-varying heteroscedasticity.

The plan of the remainder of the paper is as follows. Section 2 shows that RTV improves probabilistic forecasts in comparison with EOS when forecasting with autoregressive models, assuming that data revisions are either news or noise. Section 2 also provides measures of the expected magnitude of the differences in log score as a function of the data generating process parameters and the characteristics of the revisions. Section 3 provides some insights into the effects of time-varying volatility on real-time density forecasting in the presence of data revisions. The last part of Section 3 measures the impact of conditional heteroscedasticity on the relative performance of EOS and RTV for ARCH and GARCH models. Section 4 considers absolute and relative tests for density forecasting accuracy in a simulation exercise for samples of the size typically available in practice. We establish which statistical tests are most likely to discriminate between EOS and RTV. Section 5 reports the results of a real-time forecasting exercise for US macroeconomic time series subject to data revisions, comparing RTV and EOS one-step-ahead probabilistic forecasts. Our exercise evaluates forecasts for eight quarterly time series, and six monthly time series, using the real-time datasets made available by the Philadelphia Fed and St Louis Federal Reserve Bank. We also consider vector autoregressive models with stochastic volatility as suggested by Clark (2011). Section 6 offers some concluding remarks.

2 Improving Probabilistic Forecasts in Real Time

Clements and Galvão (2013b) show that if real-time data is reorganized into 'real-time vintages' for model estimation, instead of employing the conventional end-of-sample approach, the real-time accuracy of point forecasts from autoregressive models may be improved. In this section, we show that RTV delivers improved density forecasts in terms of log score, in comparison with the conventional approach (known as end-of-sample, or EOS), at least in population: see subsection 2.2. We also allow for an alternative density forecast loss functions when we compare RTV and EOS: subsection 2.3. We then evaluate the impact of breaks in the variance on the relative performance of both EOS and RTV, in section 2.4, and finally in subsection 2.5 allow for models with higher-order autoregressive lags than the AR(1).

2.1 Real-Time data approaches for autoregressive models: EOS and RTV

Consider the simple case of a forecaster using an autoregressive model of order one (AR(1)) for forecasting in real time. If the forecaster employs the latest-available vintage, that is, EOS, she will estimate the regression model given by:

$$y_t^{T+1} = \beta_0^{EOS} + \beta_1^{EOS} y_{t-1}^{T+1} + e_t^{EOS}, \quad \text{for } t = 2, .., T \tag{1}$$

where y_t^{T+1} is the vintage $T+1$ estimate of the reference period t value, and where t run from 2 up to T. We assume the data are published with a one period (month, or quarter) delay. The one-step ahead point forecast is:

$$\mu_{T+1|T}^{EOS} = E[y_{T+1}| \left\{ y_t^{T+1} \right\}_{t=1}^{t=T}] = \beta_0^{EOS} + \beta_1^{EOS} y_T^{T+1}$$

and the one-step-ahead forecast variance is:

$$\sigma_{T+1|T}^{2,EOS} = var \left[y_{T+1}| \left\{ y_t^{T+1} \right\}_{t=1}^{t=T} \right] = var(e_{T+1}^{EOS}) = var(e_t^{EOS}),$$

where the last equality assumes stationarity, and we abstract from parameter estimation uncertainty by using the population values of the parameters defined by (1) rather than their estimates. [3]

The problems with EOS when data are subject to revision has been analyzed by a number of authors for point forecasting (see Koenig *et al.* (2003), Kishor and Koenig (2012) and Clements and Galvão (2013b)): the model estimation relates the fully-revised values of the predictand to fully-revised values of the predictor(s), whereas the real-time prediction problem is to generate a forecast (possibly of an early estimate) using early estimates of the predictor variables. A simple solution is RTV, whereby the estimation of the model is set up to mimic the real-time prediction problem, so that in-sample early estimates of the LHS variable are related to early estimates of the RHS variable(s). Clements (2017) notes that such an approach delivers correctly-sized intervals (estimates of uncertainty more generally) in the case of autoregressive models.

If the forecaster has access to $T-1$ past vintages of y, that is, she has access to a real-time database, then RTV for the AR(1) model is given by estimation of:

$$y_t^{t+1} = \beta_0^{RTV} + \beta_1^{RTV} y_{t-1}^{t} + e_t^{RTV}, \text{ for } t = 2, .., T \tag{2}$$

[3] The '$T+1|T$' subscripts on μ and σ^2 indicate forecasts of the period $T+1$ (the first data estimate of which will not be released until $T+2$) made using data up to and including reference period T (from the $T+1$-vintage). We make this explicit because the forecast origin could be specified in terms of the time-period to which the data refers or the time when it is released (the vintage) - here the T in $|T$ refers to the former.

where y_t^{t+1} and y_{t-1}^t are time series of first releases.[4] When forecasting in real-time, the loss function the forecaster faces implies that forecasts are conditioned on initial releases. RTV estimates the parameters of the model using an objective function defined on the early-vintage estimates of the predictand and predictor(s), and so matching the out-of-sample loss function for assessing forecast accuracy. The one-step ahead point forecast and forecast variance are:

$$\begin{aligned} \mu_{T+1|T}^{RTV} &= \beta_0^{RTV} + \beta_1^{RTV} y_T^{T+1} \\ \sigma_{T+1|T}^{2,RTV} &= var(e_{T+1}^{RTV}) = var(e_t^{RTV}) \end{aligned}$$

where the last equality in the expression for the variance assumes stationarity.

2.2 Comparing EOS and RTV based on Log Scores

In order to compute density forecasts, the forecaster uses:

$$N(\mu_{T+1|T}, \sigma_{T+1|T}^2).$$

We assume the objective is to maximize the expected log score. This will minimize the Kullback-Leibler distance between the model predictive density and the true density (see, e.g., Lee, Bao and Saltoglu (2007) and Hall and Mitchell (2009)). This is obviously equivalent to minimizing the negative of the expected log score:

$$E[-\ln(p_{T+1|T}(y_{T+1}))] = E\left[\frac{(y_{T+1} - \mu_{T+1|T})^2}{2\sigma_{T+1|T}^2} + \frac{1}{2}\ln(\sigma_{T+1|T}^2) + 0.5\ln(2\pi)\right]. \tag{3}$$

The log score computed analytically for a normal predictive density as above is equivalent to the Dawid-Sebastiani score function. It is a proper score function, meaning that the optimal forecast is to deliver the true density function - there is no incentive to gameplay (as surveyed in Gneiting and Katzfuss (2014)). We generally assume that the one-step-ahead predictive densities of the EOS and RTV forecasting models are normally distributed. However, the derivations that follow should also hold for a Dawid-Sebastiani score function for non-gaussian predictive densities.

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

$$y_t = \phi y_{t-1} + \eta_t + v_t, \ |\phi| < 1 \tag{4}$$

where η_t is the underlying disturbance, and v_t is a news revision, and the first estimate is given

[4] For an AR(2), for example, there would be an additional regressor, y_{t-2}^t, which as indicated would again be a vintage-t value, and so would constitute a second-release. And so on for additional lags.

by:

$$y_t^{t+1} = y_t - v_t + \varepsilon_t \tag{5}$$

with $y_t^{t+n} = y_t$ for $n = 2, 3, \ldots$. Here ε_t is a noise revision. Then the revision $y_t^{t+2} - y_t^{t+1} \equiv y_t - y_t^{t+1} = v_t - \varepsilon_t$ consists of a noise component (when $\sigma_\varepsilon^2 = E\left(\varepsilon_t^2\right) \neq 0$) and a news component (when $\sigma_v^2 = E\left(v_t^2\right) \neq 0$), see e.g., Mankiw and Shapiro (1986). We assume η_t, v_t and ε_t are mutually uncorrelated, zero-mean random variables.

News revisions (with $\sigma_\varepsilon^2 = 0$) are characterized by the revision being uncorrelated with the first estimate:

$$Cov\left(y_t^{t+2} - y_t^{t+1}, y_t^{t+1}\right) = Cov\left(v_t, \alpha y_{t-1} + \eta_t\right) = 0,$$

and the revised estimate - the fully-revised estimate here - adds the news v_t. Later estimates are more accurate estimates of the true value than earlier estimates (here, $y_t^{t+2} = y_t$) and have larger unconditional variance than earlier estimates. Conversely, noise revisions remove measurement error: the revisions are predictable (based on period $t-1$ information) but are not correlated with the true value, i.e.:

$$Cov\left(y_t^{t+2} - y_t^{t+1}, y_t^{t+1}\right) = Cov\left(-\varepsilon_t, y_t + \varepsilon_t\right) = -\sigma_\varepsilon^2,$$

but:

$$Cov\left(y_t^{t+2} - y_t^{t+1}, y_t\right) = Cov\left(-\varepsilon_t, y_t\right) = 0.$$

Hence later estimates have smaller unconditional variances.

Assume the following homoscedastic processes:

$$\begin{aligned} \eta_t &= \sigma_\eta \xi_{1t} \\ v_t &= \sigma_v \xi_{2t} \\ \varepsilon_t &= \sigma_\varepsilon \xi_{3t} \\ \xi_{it} &\sim iidN(0,1) \text{ for } i = 1, 2, 3. \end{aligned} \tag{6}$$

In order to emphasize the differences between news and noise revisions, we suppose revisions are either news or noise. In the case of pure news revisions, we have $\sigma_\varepsilon = 0$, such that $y_t^{t+1} = y_t + rev_t$ and $rev_t = -\sigma_v \xi_{2t}$. In the case of pure noise revisions, $\sigma_v = 0$ and $rev_t = \sigma_\varepsilon \xi_{3t}$. We define δ as the relative size of the data revision process, that is:

$$\delta = \frac{var(rev_t)}{\sigma_\eta^2},$$

implying that if data revisions are news, $\delta = \sigma_v^2/\sigma_\eta^2$, and if data revisions are noise, $\delta = \sigma_\varepsilon^2/\sigma_\eta^2$.

Our setup is a simplified version of the statistical model of Jacobs and van Norden (2011):

they allow for l-revisions ($l > 1$); the possibility that the truth is not eventually revealed; and for more general processes for the true data. The statistical framework described in this section is general enough to bring out the key features of news versus noise, and of EOS versus RTV.

2.2.1 News revisions

Under pure news revisions in the model described by equations (4), (5) and (6), the EOS and RTV forecasts of the mean and the variance are given by:

$$\begin{aligned}
\mu_{T+1|T}^{EOS} &= \phi y_T^{T+1} = \phi(y_T - \sigma_v \xi_{2T}) \\
\sigma_{T+1|T}^{2,EOS} &= \sigma_\eta^2 + \sigma_v^2 = \sigma_\eta^2(1+\delta) \\
\\
\mu_{T+1|T}^{RTV} &= \phi y_T^{T+1} = \phi(y_T - \sigma_v \xi_{2T}) \\
\sigma_{T+1|T}^{2,RTV} &= \sigma_\eta^2 + \phi^2\sigma_v^2 = \sigma_\eta^2(1+\phi^2\delta)
\end{aligned} \tag{7}$$

where we use the result in Clements and Galvão (2013b) that under news revisions and $p = 1$, $\beta_0^{EOS} = \beta_0^{RTV} = 0$ and $\beta_1^{EOS} = \beta_1^{RTV} = \phi$. The expressions for the variances are derived in Clements (2017), and they imply that $\sigma_{T+1|T}^{2,EOS} > \sigma_{T+1|T}^{2,RTV}$, that is, for news revisions EOS over-states future uncertainty.

Proposition 1 *The difference between EOS and RTV log score, $\Delta score^{News}$, assuming that the target is the initial release y_{T+1}^{T+2} and data revisions are pure news in the model described by equations (4), (5) and (6), is*

$$\Delta score^{News} = \frac{1}{2}\left[\frac{\delta(\phi^2-1)}{1+\delta} + \ln[(1+\delta)/(1+\phi^2\delta)]\right], \tag{8}$$

and because $\delta \geq 0$ and $|\phi| < 1$, $\Delta score^{News} \geq 0$, that is, EOS density forecasts are further away from the true density in comparison with RTV forecasts. Proof in Appendix A.1.

The expression for $\Delta score^{News}$ suggests that if there are no revisions so $\delta = 0$, then $\Delta score^{News} = 0$. From (7) it can be seen that this results from EOS and RTV delivering the same predictive density when $\delta = 0$, by construction. Figure 1 shows how $\Delta score^{News}$ varies with ϕ and δ. The values clearly show that $\Delta score^{News}$ decreases with ϕ and increases with δ.

2.2.2 Noise revisions

Under pure noise, $\beta_1^{EOS} = \phi$, but $\beta_1^{RTV} \neq \phi$, as shown in Clements and Galvão (2013b). We write $\beta_1^{RTV} = B\phi$, where B is given by:

$$B = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\varepsilon^2} = \frac{\sigma_\eta^2/(1-\phi^2)}{\sigma_\eta^2/(1-\phi^2) + \sigma_\eta^2\delta} = \frac{(1-\phi^2)^{-1}}{((1-\phi^2)^{-1} + \delta)}.$$

This implies that

$$\begin{aligned}
\mu_{T+1|T}^{EOS} &= \phi y_T^{T+1} = \phi(y_T + \sigma_\varepsilon \xi_{3T}) \\
\mu_{T+1|T}^{RTV} &= B\phi y_T^{T+1} = B\phi(y_T + \sigma_\varepsilon \xi_{3T}).
\end{aligned}$$

Because we use mainly revised data to compute $\sigma_{T+1|T}^{2,EOS}$, in the case of pure noise revisions, we have the EOS forecast of the variance as simply:

$$\sigma_{T+1|T}^{2,EOS} = \sigma_\eta^2 \tag{9}$$

The one-step-ahead variance computed with the RTV approach is:

$$\sigma_{T+1|T}^{2,RTV} = \sigma_\eta^2(1 + \delta + \varrho), \tag{10}$$

where detailed derivations are in Appendix A.2, and

$$\varrho = [\phi^2(B-1)^2/(1-\phi^2) + \delta B^2\phi^2].$$

Because $B < 1$ and $|\phi| < 1$, then $\varrho > 0$, implying that for the same δ and ϕ, $\sigma_{T+1|T}^{2,RTV}$ for noise is greater than $\sigma_{T+1|T}^{2,EOS}$ for news. Clements (2017) shows that RTV accurately reflects the one-step uncertainty about y_{T+1}^{T+2}, but EOS leads to an under-estimation of future uncertainty when there are noise revisions. Note also that if there are no revisions ($\delta = 0$), then $\varrho = 0$ since $B = 1$.

Proposition 2 *The difference between EOS and RTV log score, $\Delta score^{Noise}$, assuming that the target is the initial release y_{T+1}^{T+2} and data revisions are pure noise in the model described by equations (4), (5) and (6), is*

$$\Delta score^{Noise} = \frac{1}{2}\left[(\delta(1+\phi^2)) - \ln(1+\delta+\varrho)\right], \tag{11}$$

and because $\delta \geq 0$ and $|\phi| < 1$, $\Delta score^{Noise} \geq 0$, that is, EOS density forecasts are further away from the true density in comparison with RTV forecasts. Proof in Appendix A.3

Figure 1 shows $\Delta score^{Noise}$ values for different values of ϕ and δ. They clearly show that, as in the case of news revisions, $\Delta score^{Noise}$ increases with δ, but in contrast now also increases with ϕ.

Comparing the plots for news and noise suggests larger differences in EOS and RTV log scores are likely to be observed empirically when data revisions are noise. (Note the different scales of the vertical axes). For example, if $\delta = 2$, differences in log score are between 0.46 and 1.1 for noise revisions are noise but only between 0.004 and 0.22 for news.

2.3 Alternative Score Functions

Although the log score is arguably the most popular score, other proper score functions could be used to compare the relative performance of RTV and EOS density forecasts. The continuous ranked probability score (CRPS, due to Epstein (1969)) is one such scoring rule, and uses the predictive cumulative density function in place of the probability density function. The CRPS generalizes the absolute error to density forecasts, so it is easier to link to point forecasting accuracy, and it is more robust to outliers. In the case of normal predictive densities, the CRPS closed form expression:

$$CRPS(F_{T+1|T}(y_{T+1})) = \sigma_{T+1|T}\left[\frac{(y_{T+1}-\mu_{T+1|T})}{\sigma_{T+1|T}}\left(2\Phi\left(\frac{y_{T+1}-\mu_{T+1|T}}{\sigma_{T+1|T}}\right)-1\right)+2\varphi\left(\frac{y_{T+1}-\mu_{T+1|T}}{\sigma_{T+1|T}}\right)-\frac{1}{\sqrt{\pi}}\right],$$

where $\Phi(.)$ is the CDF of the standard normal distribution, and $\varphi(.)$ is the PDF of the standard normal distribution (see, e.g., Gneiting and Katzfuss (2014)).

A proper scoring rule for interval forecasts based on quantiles $q^{\frac{\alpha}{2}}_{T+1|T}$ and $q^{1-\frac{\alpha}{2}}_{T+1|T}$ is:

$$q\_loss(y_{T+1};\alpha) = (q^{1-\frac{\alpha}{2}}_{T+1|T} - q^{\frac{\alpha}{2}}_{T+1|T}) + \frac{2}{\alpha}(q^{\frac{\alpha}{2}}_{T+1|T} - y_{T+1})I(y_{T+1} < q^{\frac{\alpha}{2}}_{T+1|T}) + \frac{2}{\alpha}(y_{T+1} - q^{1-\frac{\alpha}{2}}_{T+1|T})I(y_{T+1} > q^{1-\frac{\alpha}{2}}_{T+1|T}),$$

where smaller scores are better, and the first term is a penalty for the length of the interval. The quantiles for distribution function F are such that $\gamma = F(q^{\gamma})$ for $\gamma \in (0, 1)$.

Table 1 records the differences between EOS and RTV for these score functions, for particular values for ϕ and δ. We assume that $\sigma_\eta = 0.753$, which is calibrated to US quarterly GDP. The results are based on a large sample ($T = 500$), and 2,000 replications of the Monte Carlo. The quantile loss is computed for $\alpha = 0.10$. Table 1 also compares the simulated log score difference with the analytical expressions given by (8) and (11). The results suggest that the analytical results for $\Delta score^{News}$ and $\Delta score^{Noise}$ provide good predictions for the observed differences in log score. Evidently the sample size is sufficiently large that the impact of parameter estimation uncertainty on the simulated values is small. The two density score functions suggest that RTV delivers better probabilistic forecasts for all the parameter values considered. Similarly, RTV generates better intervals judged by q_loss. The relative differences in the performance of EOS and RTV vary in a similar way across the parameter space for each of the three metrics.

2.4 Impact of Breaks

In this subsection, we evaluate the effect of a break in the variance of the underlying disturbances (η_t) of the data generating process on the relative merits of RTV and EOS. This is motivated by the evidence of a downward break in the volatility of macroeconomic series in the 1980s, as reported by McConnell and Perez-Quiros (2000) and Sensier and van Dijk (2004), *inter alia*, usually called the Great Moderation. The new process for η_t in (4) is amended to:

$$\eta_t = (\sigma_\eta + I(t \geq \tau)\sigma_b)\xi_{1t}$$

We set the break date as $\tau = 0.95T$ and we look at breaks that increase the variance $\sigma_b = 0.5\sigma_\eta$ and breaks that decrease the variance $\sigma_b = -0.5\sigma_\eta$.

Table 2 presents results for $T = 200$, for the same set of parameter values as in Table 1. To highlight the effect of the break, we also record the analytical and simulated differences in EOS and RTV log scores when there is no break. The forecasting models for both the break and 'no break' cases are as in (1) and (2).

Table 2 shows that breaks that occur towards the end of sample (as here) clearly affect the relative performance of EOS and RTV measured by log score. An upward shift in the variance of the underlying process when revisions are news improves the relative performance of EOS, to the extent that EOS is now preferred on log score. However, a downward shift further improves RTV relative to EOS, compared to the no-break scenario. These results arise because the EOS-forecasting device is subject to two forms of mis-specification, which partially offset when the variance increases. As established in section 2.2, the first is that EOS forecasts of variance over-estimate the variance when revisions are news (in the absence of breaks). Secondly, a forecasting model will generate under-predictions when there is an unexpected and unmodelled upward shift in the mean of the target variable. Hence the two forms of mis-specification will offset for unmodelled upward shifts.

For noise revisions the relative performance of EOS is improved (and for some parameter values dominates RTV) when the variance decreases. The rationale is as above, noticing that EOS under-estimates the variance when revisions are noise.

In summary, in the absence of unmodelled breaks or shifts in the underlying variance of the process, RTV would be expected to outperform EOS for one-step-ahead density forecasts of initial releases. However, breaks in variance are a confounding factor. Breaks that occurred some time in the past might be dealt with by using only more recent, post-break observations to estimate the model (see, for example, Pesaran and Timmermann (2007)) or by using rolling windows of observations (as opposed to a recursive scheme) to allow some adaptivity of the model to the changed environment. Both of these are routinely used for first-moment prediction. We leave for future research a study of their efficacy for second-moment prediction in the presence of data revision uncertainty.

2.5 Longer Autoregressive Orders

The previous results are based on the simplifying assumption of setting the autoregressive orders of the data generating process and the forecasting models to one. However, Clements (2017) shows that the optimal RTV order may exceed that of the data generation process for noise revisions. In this subsection, we re-run the simulation results of section 2.3 allowing selection of the EOS and RTV models' lag orders using an information criterion. We also consider an AR(2) for the true process in second set of simulations.

Table 3A shows the differences between EOS and RTV in terms of log score, CRPS, and quantile loss, with $\alpha = 0.10$, based on 5,000 simulations with $T = 200, 500$, assuming that the data generating process (DGP) is as in Table 1, but the forecasting model orders are selected by the Bayesian Information Criterion (BIC) assuming a maximum autoregressive order of 4. As a consequence, by comparing the results of Table 3A with Table 1 we can assess the impact of uncertainty concerning forecasting model lag orders on the density forecasts. The observed differences in log score are not very different from the analytical values obtained ignoring parameter uncertainty. Indeed in the case of news revisions, BIC chooses a lag order of one for both EOS and RTV on 98-99% of the replications. In the case of noise revisions, the information criteria picks $p = 2$ for RTV when the process is persistent and revisions relatively large ($\phi = 0.9$, $\delta = 1$) for both sample sizes. The consequence of allowing an additional lag is larger gains to RTV in terms of log score, CRPS and the quantile loss function. Evidently, although fixing $p = 1$ enables readily interpretable analytical derivations, these under-estimate the relative advantage of RTV when revisions are noise.

Table 3B shows similar information to Table 3A, but for an AR(2) for the data generating process in equation (4). The parameters are set as $\phi_1 = \phi/2$ and $\phi_2 = \phi/2$. Interestingly the relative gains to RTV over EOS with $T = 200$ are larger than for $T = 500$ if data revisions are noise, but $T = 200$ gains are smaller when data revisions are news. This suggests that when the sample size is short, we should expect to find more evidence in favour of RTV forecasts when data revisions are noise. For the case that $\phi = 0.1$ ($\delta = 1$), the information criteria in around 90% of replications picks $p = 1$ instead of $p = 2$ for both EOS and RTV. Instead when $\phi = 0.9$, and sample size is large, the information criteria choose $p = 3$ in 15% of the replications using RTV.

In summary, the findings of this subsection are that our results are applicable to AR(p) models, in that they correctly indicate the superiority of RTV over EOS, and that the gains to RTV are greater for noise revisions compared to news revisions (for a given δ). However, even in population (abstracting from parameter estimation uncertainty) the assumption that the RTV model order is the same as that of the data generating process does not hold for noise revisions, and this is reflected in the simulation study when the model order is determined by BIC, and longer lags are selected. Empirically, the model orders will depend on a range of factors, as illustrated by the Monte Carlo using an AR(2) for the true process.

3 Real-time Probabilistic Forecasting with Conditional Heteroscedastic Models

The effect of data revisions and time-varying volatility on probabilistic forecasts are investigated for ARCH, GARCH and SV models. We assume that the data revisions process is such that the true value is revealed after one revision, and the target forecast is of a one-step-ahead first-release value. We start by describing the data generating process with conditional heteroscedasticity, and then establish the forecasting performance of RTV and EOS in these circumstances. We also use a simulation exercise to measure the impact of conditional heteroscedasticity on the relative forecasting performance of EOS and RTV, and on their prediction interval coverage rates.

3.1 Adding Conditional Heteroscedasticity

As before, the process for the true (revised values) is given by equation (4). Assume now that η_t is conditionally heteroscedastic, and follows an ARCH(1) process:

$$\begin{aligned} \eta_t &\sim N(0, h_t) \\ h_t &= w_0 + w_1 \eta_{t-1}^2. \end{aligned} \tag{12}$$

We assume that the processes for news v_t and noise revisions ε_t are as before, that is, they are white noise with constant variances. The unconditional variance implied by the ARCH(1) process is set as $w_0/(1-w_1) = \sigma_\eta^2$. In the absence of data revisions, $\eta_t = y_t - \phi y_{t-1}$ and the process for the squared errors is:

$$(y_t - \phi y_{t-1})^2 = \sigma_\eta^2 + w_1[(y_{t-1} - \phi y_{t-2})^2 - \sigma_\eta^2] + z_t \tag{13}$$

where z_t is the error in forecasting the squared errors with h_t, that is, $z_t = \eta_t^2 - h_t$. The resulting AR(1) process for the squared errors $\eta_t^2 = (y_t - \phi y_{t-1})^2$ of the ARCH model is described in, e.g., Taylor (2005, p. 199). Note that $var(\eta_t^2)$ is only finite if $3w_1^2 < 1$, so we normally expect w_1 to be small.

We could instead assume that the conditional variance of η_t follows a GARCH (1,1) process, that is,

$$h_t = w_0 + w_1\eta_{t-1}^2 + w_2 h_{t-1}. \tag{14}$$

Then the implied process for the squared errors is

$$(y_t - \phi y_{t-1})^2 = \sigma_\eta^2 + (w_1 + w_2)[(y_{t-1} - \phi y_{t-2})^2 - \sigma_\eta^2] + z_t - w_2 z_{t-1},$$

following Taylor (2005, p. 203) and using $\sigma_\eta^2 = w_0/(1 - w_1 - w_2)$. Note that volatility process is stationary if $|w_1 + w_2| < 1$, but $var(\eta_t^2)$ is finite only if $(2w_1^2 + (w_1 + w_2)^2) < 1$, implying that as before w_1 should typically be small.

An alternative approach to modelling conditional heteroscedasticity popular in the macroeconomic forecasting literature (e.g., Primiceri (2005), Cogley and Sargent (2005), Clark (2011)) is to assume that:

$$\begin{aligned} \eta_t &= \sqrt{h_{\eta t}}\xi_{1t};\ \xi_{1t} \sim iidN(0,1) \\ \ln h_{\eta t} &= \kappa_{0\eta} + \kappa_{1\eta}\ln h_{\eta t-1} + \sigma_1\varsigma_{1t};\ \varsigma_{1t} \sim iidN(0,1). \end{aligned} \tag{15}$$

This stochastic volatility (SV) model for η_t can be cast in the following state-space representation (assuming no data revisions):

$$\begin{aligned} \ln(y_t - \phi y_{t-1})^2 &= \ln h_{\eta t} + \ln \xi_{1t}^2 \\ \ln h_{\eta t} &= \kappa_{0\eta} + \kappa_{1\eta}\ln h_{\eta t-1} + \sigma_1\varsigma_{1t}, \end{aligned}$$

where the first equation is the measurement equation and both the log conditional variance ($\ln h_{\eta t}$) and the log squared disturbance ($\ln \xi_{1t}^2$) are unobserved components. Harvey, Ruiz and Shephard (1994) classify the above representation of a stochastic volatility model as a linear state-space with non-Gaussian measurement errors (since $\ln \xi_{1t}^2$ is not Gaussian). The state-space representation implies that:

$$\begin{aligned} \ln(y_t - \phi y_{t-1})^2 &= \kappa_{0\eta} + \kappa_{1\eta}\ln(y_{t-1} - \phi y_{t-2})^2 + \sigma_1\varsigma_{1t} + \ln \xi_{1t}^2 - \kappa_{1\eta}\ln \xi_{1t-1}^2 \\ &= \ln(\sigma_\eta^2) + \kappa_{1\eta}[\ln(y_{t-1} - \phi y_{t-2})^2 - \ln(\sigma_\eta^2)] + \sigma_1\varsigma_{1t} + \ln \xi_{1t}^2 - \kappa_{1\eta}\ln \xi_{1t-1}^2, \end{aligned}$$

which is an ARMA(1,1) process for σ_1 small.

The SV model treats volatility as a latent variable. Relative to the GARCH(1,1) model, it implies an ARMA (1,1) representation for the log, as opposed to the levels, of the squared errors. Note that $var(\eta_t^2)$ is stationary if $|\kappa_{1\eta}| < 1$. The unconditional variances are:

$$\begin{aligned} E[\ln(y_t - \phi y_{t-1})^2] &= \ln(\sigma_\eta^2) \\ E[(y_t - \phi y_{t-1})^2] &= \exp\{\ln(\sigma_\eta^2) + 0.5[(\sigma_1^2/(1-\kappa_{1\eta}^2)]\}, \end{aligned}$$

and in the above computation we are excluding the effects of $\ln \xi_{1t}^2$, as when the method of moments is employed to estimate stochastic volatility models (as in Jacquier, Polson and Rossi (1994)).

In the following we assume that conditional heteroscedasticity only affects the innovations to the truth (η_t) in equation (4). The empirical literature provides evidence of changes in volatility for true values of macroeconomic series (as in McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004)), but does not provide strong evidence of conditional heteroscedasticity in the data revision processes.[5] Allowing conditional heteroscedasticity in the data revision process (v_t or ε_t) would complicate the analysis. For example, if both η_t and v_t followed ARCH processes, then the process for the squared error will be the sum of two AR(1) processes. The resultant process will be an ARMA(2,1) process, see e.g., Hamilton (1994, p. 107), and a forecasting model which simply adds an ARCH(1) to the innovations would be mis-specified. Given the additional difficulties from allowing conditional heteroscedasticity in the data revision process and the lack of empirical evidence for this phenomenon, we do not consider this extension here.

3.2 Forecasting with Real-Time Vintages

Section 2 shows that RTV delivers a better log score than EOS for homoscedastic AR processes. In this subsection, we modify the forecasting model to add conditional heteroscedasticity. We then argue that by estimating the model by maximizing the likelihood, with data of the same maturity as that in the real-time log score function, we will obtain a one-step-ahead predictive density that maximizes log score.

A conditional heteroscedastic AR(1) model with RTV is:

$$\begin{aligned} y_t^{t+1} &= \beta_0^{RTV} + \beta_1^{RTV} y_{t-1}^t + e_t^{RTV} \text{ for } t = 2,..,T \\ e_t^{RTV} &\sim N\left(0, h_t^{RTV}\right), \end{aligned} \tag{16}$$

and the conditional variance may follow either an ARCH(1) specification:

$$h_t^{RTV} = w_0^{RTV} + w_1^{RTV}\left(e_{t-1}^{RTV}\right)^2, \tag{17}$$

[5] Swanson and van Dijk (2006) provide evidence that data revisions may change over the business cycle through nonlinear dynamics in the revision process, while maintaining the assumption of constant variance.

a GARCH (1,1):

$$h_t^{RTV} = w_0^{RTV} + w_1^{RTV}\left(e_{t-1}^{RTV}\right)^2 + w_2^{RTV} h_{t-1}^{RTV} \tag{18}$$

or stochastic volatility:

$$\log h_t^{RTV} = \kappa_0^{RTV} + \kappa_1^{RTV} \log h_{t-1}^{RTV} + \sigma^{RTV} \varsigma_t. \ \varsigma_t \sim iidN(0,1). \tag{19}$$

Let θ denote the vector of parameters for each specification. In the case, for example, of ARCH(1) models, we have $\theta = (\beta_0^{RTV}, \beta_1^{RTV}, w_0^{RTV}, w_1^{RTV})'$. The parameters of the forecasting model are estimated by maximizing the log likelihood function with respect to the vector of parameters θ. In the case of stochastic volatility models, because the measurement errors are non-Gaussian, the Kalman filter will provide only an approximation of the unobserved $\log h_t^{RTV}$, but if employed to build the likelihood, we obtain quasi-maximum likelihood estimators as discussed in Harvey et al (1994). For the ARCH and GARCH specifications, estimation requires:

$$\max_\theta LL(\theta; y_1^2, ..., y_{T-1}^T, y_T^{T+1}) = -\frac{1}{2}\sum_{t=2}^{T}\left(\ln(2\pi h_t^{RTV}) + \frac{(y_t^{t+1} - \beta_0^{RTV} - \beta_1^{RTV} y_{t-1}^t)^2}{h_t^{RTV}}\right) \tag{20}$$

The average real-time log score computed assuming that we are targeting the first release y_{T+i}^{T+1+i} for $i = 1, ..., P$, where P is the number of observations in the out-of-sample period is:

$$LS(p_{t+1|t}(y_{t+1})) = -\frac{1}{2}\sum_{i=0}^{P-1}\left(\ln(2\pi\sigma^2_{T+1+i|T+i}) + \frac{\left(y_{T+i}^{T+1+i} - \mu_{T+1+i|T+i}(y_{T+i-1}^{T+i})\right)^2}{\sigma^2_{T+1+i|T+i}}\right). \tag{21}$$

The above equations imply that if we use RTV data, the data maturity employed to obtain the parameters that maximize $LL(\theta; y_1^2, ..., y_{T-1}^T, y_T^{T+1})$ is the same as the one employed to compute $LS(p_{t+1|t}(y_{t+1}))$. A similar intuitive argument is behind the results in Clements and Galvão (2013b), who show that RTV yields optimal estimates in population in a MSE sense if a general model of data revisions (as in Jacobs and van Norden (2011)) is allowed. This is also behind the analytical results in section 2 for the autoregressive model, since parameter values that minimize the sum of in-sample squared errors (OLS) also maximize the likelihood.

As a consequence, we expect that allowing for time-varying conditional variances in the data generating process and in the forecasting model will not change the expected dominance of RTV over EOS on log score ($\Delta score^{News} > 0$ and $\Delta score^{Noise} > 0$), at least discounting any small-sample estimation effects. We also expect that the model in (16) will deliver empirical predictive interval coverages equal to the nominal values. We collect these results in the following proposition.

Proposition 3 *If the data generating process is given by (4) and (5) with the η_t process given by either (12), (14) or (15), and we use accordingly one of (17), (18) and (19) estimated by RTV, then the data maturity in the real-time out-of-sample log score (21) is the same as the one employed for estimating the forecasting model parameters in (20). As a consequence, RTV delivers the one-step-ahead predictive density that is nearest to the true density, that is, density forecasts which maximize the log score.*

Note that in general the RTV parameters will be not equal to those of the underlying processes for the mean equation for the true values (y_t) or for the conditional variance equation for η_t. For an AR(1) and news revisions, the mean parameters coincide, but those for the conditional variance differ, and for higher-order models, and for noise revisions even with first-order models, the parameters of the mean equation are different.

The estimated parameters of the conditional variance h_t^{RTV} will be such that the implied unconditional variance will be:

$$E[h_t^{RTV}] = \sigma_\eta^2 + \phi^2\sigma_v^2 \tag{22}$$

if data revisions are news, and

$$E[h_t^{RTV}] = \sigma_y^2[(1-\phi^2) + \phi^2(B-1)^2] + (1+B^2\phi^2)\sigma_\varepsilon^2 \tag{23}$$

if data revisions are noise, exactly matching the expressions for $\sigma^{2,RTV}_{T+1|T}$ for news and noise in the homoscedastic case (equations (7) and (10) in section 2.2). (In (23), $\sigma_y^2 = \sigma_\eta^2\left(1-\phi^2\right)^{-1}$).

3.3 Forecasting with the End-of-Sample Approach

In the previous section, we explained the use of RTV to obtain one-step-ahead variances that match real-time out-of-sample expected squared forecasting errors, and in so doing maximize the log score of the predictive density. In this section, we show the mismatch between the predicted variances and out-of-sample values for EOS.

The EOS forecasting model is also modified to allow for conditional heteroscedasticity in the disturbances. As before we employ an autoregressive model of order 1 for the conditional mean for simplicity. Results are easily extended to an AR(p). The forecasting model estimated with EOS data is:

$$\begin{aligned} y_t^{T+1} &= \beta_0^{EOS} + \beta_1^{EOS} y_{t-1}^{T+1} + e_t^{EOS} \text{ for } t = 2,..,T \\ e_t^{EOS} &\sim N\left(0, h_t^{EOS}\right) \end{aligned} \tag{24}$$

where

$$h_t^{EOS} = w_0^{EOS} + w_1^{EOS}\left(e_{t-1}^{EOS}\right)^2 \tag{25}$$

in the case of an ARCH (1) model, but the conditional variance follows:

$$h_t^{EOS} = w_0^{EOS} + w_1^{EOS} \left(e_{t-1}^{EOS}\right)^2 + w_2^{EOS} h_{t-1}^{EOS} \tag{26}$$

in the case of a GARCH model, and, finally, in the case of stochastic volatility, we have:

$$\ln h_t^{EOS} = \kappa_0^{EOS} + \kappa_1^{EOS} \ln h_{t-1}^{EOS} + \sigma^{EOS} \varsigma_t;\ \varsigma_t \sim iidN(0,1). \tag{27}$$

3.3.1 Noise Revisions

Assume that data revisions are noise. This implies that the population values of the EOS estimates coincide with those in (4) and (12) if we ignore the influence of the latest observation (y_{T-1}^T), as will be reasonable for large sample sizes. Explicitly, with an ARCH(1) model for the conditional variance, $\left(\beta_0^{EOS}, \beta_1^{EOS}, w_0^{EOS}, w_1^{EOS}\right) = (0, \phi, w_0, w_1)$, and:

$$\begin{aligned}
\mu_{T+1|T}^{EOS} &= \phi y_T^{T+1} \\
\sigma_{T+1|T}^{2EOS} &= w_0^{EOS} + w_1^{EOS} (\eta_T + \varepsilon_T)^2 \\
&= w_0 + w_1 \eta_T^2 + w_1 \varepsilon_T^2 + 2 w_1 \eta_T \varepsilon_T \\
&= \sigma_\eta^2 + w_1 \varepsilon_T^2 + w_1 (\eta_T^2 - \sigma_\eta^2) + 2 w_1 \eta_T \varepsilon_T
\end{aligned}$$

where the second line substitutes $(y_T^{T+1} - \phi y_{T-1}^{T+1}) = (y_T + \varepsilon_T) - \phi y_{T-1} = \eta_T + \varepsilon_T$, under the assumption of pure noise revisions.

These results imply that the unconditional variance is:

$$E\left(\sigma_{T+1|T}^{2EOS}\right) = \sigma_\eta^2 + w_1 \sigma_\varepsilon^2 = \sigma_\eta^2 (1 + w_1 \delta), \tag{28}$$

which can be compared to the homoscedastic case, (9). In the homoscedastic case, the EOS forecast of the variance under-estimates the true uncertainty by $(\delta + \varrho)\,\sigma_\eta^2$, from subtracting (9) from (10). Allowing for heteroscedasticity in the underlying disturbances, and using EOS with an ARCH(1) error, results in a difference between the true and forecast variance of:

$$\sigma_\eta^2 \,(1 + \delta + \varrho) - \sigma_\eta^2 (1 + w_1 \delta) = \sigma_\eta^2 \,(\delta + \varrho - w_1 \delta)$$

which is smaller than $(\delta + \varrho)\,\sigma_\eta^2$. Hence, the costs in terms of interval coverage of using EOS, compared to RTV, will typically be lessened when there is heteroscedasticity, and this is modelled. Because conditional heteroscedasticity does not affect the conditional mean forecats, we can use these results to suggest that EOS logarithm scores will be improved relatively to the RTV scores. We measure the impact of these changes on density forecasting performance in section 3.4.

In Appendix B, we show similar results for GARCH$(1,1)$ and SV models. For both cases we can show that in principle the transmission of the forecasting origin noise should improve the under-coverage of EOS forecasts.

3.3.2 News Revisions

Suppose now we have the data generating process given by (4) and (12), but data revisions are pure news. Because conditional heteroscedasticity has no impact on the conditional mean parameters, we know that the population parameters in equation (24) will be $\beta_0^{E0S} = 0$ and $\beta_1^{E0S} = \phi$. The ARCH process is then estimated for the composite error $e_t^{EOS} = \eta_t + v_t$. As a consequence, $w_0^{EOS} \neq w_0$ and $w_1^{EOS} \neq w_1$. We show in Appendix A.1.4 that:

$$w_0^{EOS} = \left(\sigma_\eta^2 + \sigma_v^2\right)\left(1 - w_1^{EOS}\right), \tag{29}$$

$$w_1^{EOS} = w_1 \frac{var(\eta_t^2)}{var(\eta_t^2) + 2\sigma_v^4 + 4\sigma_\eta^2\sigma_v^2} \tag{30}$$

where w_1 is the ARCH slope parameter (see (12)) and $var(\eta_t^2)$ depends on σ_η^2 and w_1 (see the Appendix). From (30) it follows that $w_1^{EOS} < w_1$, such that w_1^{EOS} is reduced relative to w_1, and the degree to which this occurs increases in the size of the revision δ (recall that $\sigma_v^2 = \delta\sigma_\eta^2$).

The conditional predictive variance is:

$$\begin{aligned}
\sigma_{T+1|T}^{2EOS} &= w_0^{EOS} + w_1^{EOS}\left(e_T^{EOS}\right)^2 \\
&= w_0^{EOS} + w_1^{EOS}\eta_T^2 \\
&= (\sigma_\eta^2 + \sigma_v^2) + w_1^{EOS}(\eta_T^2 - (\sigma_\eta^2 + \sigma_v^2))
\end{aligned}$$

The second line uses $e_T^{EOS} = y_T^{T+1} - \phi y_{T-1}^{T+1} = y_T - v_T - \phi y_{T-1} = \eta_T$.

Averaging the variance forecasts over the out-of-sample period corresponds to taking the unconditional expectation, resulting in:

$$\begin{aligned}
E[\sigma_{T+1|T}^{2EOS}] &= (\sigma_\eta^2 + \sigma_v^2) - w_1^{EOS}\sigma_v^2 \\
&= \sigma_\eta^2(1 + (1 - w_1^{EOS})\delta)
\end{aligned}$$

The term $w_1^{EOS}\sigma_v^2$ is the effect of the first release data at the forecasting origin when computing the one-step-ahead predictive variance. As shown, $w_1^{EOS} < w_1$, and w_1 will be small for a process with finite fourth moments. As consequence, only a small reduction in the expected variance is likely.

For news revisions, from (7), the true uncertainty less the EOS forecast variance is given by

$\sigma_\eta^2\delta(\phi^2-1)$ in the homoscedastic case. Allowing for heteroscedasticity, the difference is now:

$$\sigma_\eta^2\left(1+\phi^2\delta\right)-\sigma_\eta^2(1+(1-w_1^{EOS})\delta)=\sigma_\eta^2\delta\left(\phi^2-\left(1-w_1^{EOS}\right)\right)$$

This implies that the over-coverage of EOS interval forecasts will be smaller in the heteroscedastic case in comparison with the homoscedastic if $w_1^{EOS}<\min\left\{1,\left|2\left(\phi^2-1\right)\right|\right\}$. For $\phi^2<0.5$, the condition reduces to $w_1^{EOS}<1$, which holds automatically for a stationary ARCH. The condition will hold for all but processes with large w_1^{EOS} parameters, which are normally not very realistic as argued previously. As a consequence, it seems reasonable to assume that the presence of heteroscedasticity will reduce the costs of EOS, relative to RTV, for news revisions, as we also established in section 3.3.1 for noise revisions.

The results for the GARCH and SV models are given in Appendix B. They also suggest that EOS will over-estimate the coverage of predictive intervals when revisions are news. The EOS estimates of the parameters of the conditional variance equation of the data generating process will reflect the news revisions as well as the underlying disturbances, even though Clements and Galvão (2013b) show that the EOS estimates of the conditional mean parameters are equal to those of the data generating process in large samples when $p=1$.

3.4 Measured effects on log score and interval coverage

We use a 'large sample' simulation study to compare EOS and RTV 'in population', to illustrate the magnitude of the effects and differences between EOS and RTV described in the previous two sections. We estimate EOS and RTV forecasting models for large T ($T=1000$). We only consider ARCH and GARCH models, since they are relatively simple to estimate compared to the computationally-intensive estimation procedures required for the SV model. Hence it is feasible to use a large T, and to run a large number of replications (5000). We consider SV models in the small-sample evaluation study of the next section. Accurate estimation of SV models requires computationally-intensive methods due to the non-Gaussianity of their state-space representation.

We consider the data generating process given by (4) and (5) using ARCH errors as in equation (12) and GARCH errors as in (14). In the case the DGP has ARCH errors we estimate both forecasting models (24) and (16), and if it has GARCH errors, we use (26) and (18).

We simulate data for the ARCH(1) processes using the following expressions:

$$h_t=\sigma_\eta^2+w_1(h_{t-1}-\sigma_\eta^2)+w_1h_{t-1}(\xi_{t-1}^2-1);\ \xi_t\sim iidN(0,1)$$

and in the case of a GARCH(1,1):

$$h_t = \sigma_\eta^2 + (w_1 + w_2)(h_{t-1} - \sigma_\eta^2) + w_1 h_{t-1}(\xi_{t-1}^2 - 1);\ \xi_t \sim iidN(0,1).$$

The parameters are set such that $\sigma_\eta = 0.753$, $\sigma_v^2 = \sigma_\varepsilon^2 = \delta\sigma_\eta^2$ as in Table 1.

Figure 2A provides results of a simulation exercise with the ARCH(1) model and $w_1 = 0.1$. The Figures in the top panel present the differences between the EOS-RTV difference in log score obtained for the ARCH and the AR model for $T = 1,000$. These differences in $\Delta score^{News}$ and $\Delta score^{Noise}$ are presented for different values of ϕ and δ. The values for news revisions (left panel) are very small. The largest differences are around -0.004 indicating that although we should expect that ARCH reduces the gap between EOS and RTV when data revisions are news, the effect on the relative performance of EOS to RTV is very small. A similar plot computed for noise revisions tells a different story. EOS probabilistic forecasts show clear improvements (relative to RTV) in the presence of (modelled) ARCH in the variance of the true disturbances. Relative improvements from using EOS data increase with ϕ, proportionally to analytical values in Figure 1.

Figure 2B allows us to assess the impact of ARCH and GARCH parameter values on the EOS-RTV log score difference. In Figure 2B, we assume $\delta = 1$, so each line represents a new set of DGP parameters/forecasting models. The Figures present simulation values for AR models, two ARCH(1) specifications ($w_1 = 0.1$ and $w_1 = 0.2$) and two GARCH(1,1) specifications ($w_1 = 0.1$ and $w_2 = 0.8; w_1 = 0.2$ and $w_2 = 0.7$). For the case of news revisions, it is really hard to see differences between the values, except for the case of $\phi = 0.1$ and GARCH($w_1 = 0.2$ and $w_2 = 0.7$). As a consequence, we conclude that in the case of news our analytical predictions for the difference between EOS and RTV log score provide a good approximation even in the presence of conditional heteroscedasticity. In the case of noise revisions, Figure 2B shows a reasonably-sized improvement (up to 10%) in relative EOS performance if we compare ARCH($w_1 = 0.2$) with AR. This improvement decreases with w_1, and is smaller for GARCH models.

In summary, conditional heteroscedasticity does have an impact on relative EOS to RTV probabilistic forecasting performance. But the effect is likely to be small, unless data revisions are noise. For noise revisions the findings suggest that the effect of ARCH will be greater than of GARCH.

Figure 2C provides results for the same specifications as in Figure 2B. Instead of log score, we now show results for effective coverage rates of 90% predictive intervals. The values are the EOS excess coverage in comparison with nominal values. Results for RTV for the same exercise suggest correct coverage (approximately 90%) in all specifications considered, so they are not shown. In the case of news revisions, the disadvantage of employing EOS to compute the predicted intervals does not change much across specifications. This follows our previous results suggesting that predictive variances do not change too much when dealing with news revisions. In the case of noise revisions, there is a reduction in the EOS under-coverage when ARCH and GARCH models are employed instead of an autoregressive model. This reduction, however, is not very large.

4 Simulation Exercise in Small Samples

In section 3 we primarily compared EOS and RTV in terms of their unconditional performance. We showed that one of the effects of conditional heteroscedasticity in the model is to reduce the difference between the EOS unconditional variance forecast and the true (average) level of uncertainty, which will translate into superior unconditional density forecasts. However, given the nature of the data generating process - the presence of time-varying heteroscedasticity - a key component of any forecast evaluation exercise is an assessment of the conditional performance, as this bears on how well the forecasting methods capture the time-variation (as stressed by Clements and Taylor (2003), amongst others).

Section 3.4 has provided some indication as to how EOS and RTV might perform in real-time in terms of log score and interval coverage, given large estimation samples, and so effectively abstracting from model estimation uncertainty. In this section, the simulation study investigates how both approaches perform in small samples, as guidance for our empirical exercises in section 5, and we also consider tests which focus on the conditional aspect of forecast performance.

We set $T = 200$ and $P = 100$, where P is the out-of-sample period, and the total number of observations generated is $T + P$. We re-estimate the forecasting models using rolling windows of size T over the out-of-sample period. This design mimics our setup in the empirical exercise. It also allows the evaluation of a sequence of one-step ahead forecasts, and of the potential relationships between elements of those sequences which might signal non-optimality.

We consider the data generating process given by (4) and (5) using homoscedastic errors (AR) as in equation (6), ARCH errors as in equation (12), GARCH errors in η_t as in (14), and stochastic volatility as in equation (15). In the case the DGP is homoscedastic we estimate EOS and RTV models as in (1) and (2), if it has ARCH errors we estimate (25) and (17), if it has GARCH errors, we use (26) and (18), and, finally, if errors have SV, we use (27) and (19).

In short, the model matches the form of conditional volatility in the data generating process, but no attempt is made to forecast using a model that incorporates modelling of the data revision process. This typifies the neglect of data revisions inherent in the standard approach - EOS - whereby data revisions are effectively ignored.

As before, we set $\sigma_\eta = 0.753$, and for the case of ARCH/GARCH models, we use $w_1 = 0.4$ and $w_2 = 0.5$, and simulate data as described in section 3.4. For the stochastic volatility data generating process, we simulate data using:

$$\ln h_{\eta t} = (1 - \kappa_{1\eta}) \ln(\sigma_\eta^2 - 0.5\lambda_\eta^2) + \kappa_{1\eta} \ln h_{\eta t-1} + (1 - \kappa_{1\eta}^2)^{1/2} \lambda_\eta \varsigma_{1t}, \tag{31}$$

which ensures the unconditional variance is fixed at $\sigma_\eta^2 = \exp(\ln h_\eta)$, where $\lambda_\eta^2 = \sigma_1^2/(1 - \kappa_{1\eta}^2)$. We set $\kappa_{1\eta} = 0.9$ and $\lambda = 0.1$. These parameters are compatible with the notion that the volatility of macroeconomic variables changes slowly over time so the process has a large autoregressive parameter and a relatively small variance for the disturbances. The values of ϕ are set to $\phi = 0.1, 0.5, 0.9$ if data revisions are large ($\delta = 1$), and $\phi = 0.5$ if data revisions are small ($\delta = 0.3$).

The forecasting models with stochastic volatility are estimated by Gibbs sampling using the modification proposed by Chan (2013) to draw volatility series based on a mixture of normals approximation to the underlying non-Gaussian state-space (following Kim *et al.* (1998)). Conditional draws for the stochastic volatility parameters use normal-inverse-gamma priors and closed-form conditional distributions.

4.1 Tests of Equal Accuracy, Interval Coverage and Independence, and PITs-based test

Next we describe the statistical tests that may detect differences between RTV and EOS, and specifically the expected improvements from using RTV. We then present estimates of the rejection rates of applying these tests.

We start by looking at tests of equal accuracy, and consider different loss functions. In addition to log scores, we also look at the CRPS and the quantile loss function with $\alpha = 0.10$ as described in section 2.3. For each loss function, we use the Diebold and Mariano (1995) t-statistic, and compute p-values assuming the statistic has a normal distribution under the null, following Diebold and Mariano (1995). We use a one-sided-test design such that rejections favour the RTV model against the EOS model. One-sided tests are normally suggested in this context (as in Clark and McCracken (2013)) because of the low power in small samples of tests of equal accuracy.

We also consider whether the predictive intervals have correct coverage on average and conditionally, following Christoffersen (1998). We use the lower $q_{T+1|T}^{\frac{\alpha}{2}}$ and upper $q_{T+1|T}^{1-\frac{\alpha}{2}}$ limits of the prediction interval of y_{T+1}, made at time T, with nominal coverage α. Let $I_t = 1$ denote a 'hit', defined as:

$$I_t = \begin{cases} 1 & \text{if } y_{T+1} \in \left(q_{T+1|T}^{\frac{\alpha}{2}}, q_{T+1|T}^{1-\frac{\alpha}{2}}\right), \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

for a sequence of forecasts, $\left\{q_{T+1+i|T+i}^{\frac{\alpha}{2}}, q_{T+1+i|T+i}^{1-\frac{\alpha}{2}}\right\}$, and realizations, $\{y_{T+1+i}\}$, $i = 0, 1, \ldots, P-1$. Correct unconditional coverage holds when $E(I_t) = \alpha$, correct conditional coverage requires that $E(I_t|\mathcal{I}_{t-1}) = \alpha$, i.e., $\{I_t\}$ is iid Bernoulli with parameter α. This is a joint test of $E(I_t) = \alpha$ and of independence of $\{I_t\}$ (e.g., against a first-order Markov chain structure structure).

We also examine tests for calibration and independence based on probability integral transforms (PIT).[6] The PITs are computed using the predicted cumulative density function at the realization, that is, $P_{T+1+i|T+i}(y_{T+1+i})$ for $i = 0, ..., P-1$. In the case of AR and GARCH models we use the normal CDF and predictions for $\mu_{T+1+i|T+i}$ and $\sigma^2_{T+1+i|T+i}$ to compute the PITs values. In the case of stochastic volatility models, we have a set of draws from the predictive density, so we compute PITs values numerically. We apply a test suggested by Berkowitz (2001) to evaluate the null hypothesis of correct calibration and independence of the PITs. This implies that we test whether the inverse normal transformation of the values $P_{T+1+i|T+i}(y_{T+1+i})$ has a mean zero, variance one, and no serial correlation of order 1, using an LR statistic and chi-squared critical values.

4.2 Analysis of Simulation Results

We consider tests with a nominal significance level of 10% since this is the typical level employed to evaluate forecasts when the out-of-sample is not very large ($P = 100$ in our case). Table 4 presents rejection frequencies at the 10% level for the following: equal accuracy tests in terms of log score, CRPS and q_loss; tests for correct unconditional coverage (UC), independence (ID) and conditional coverage (CD) of 50% and 90% prediction intervals; and tests for calibration and independence of density forecasts (PITs: Berkowitz (2001)). We have four sub-tables for the different combinations of ϕ and δ.

Our assessment of the results in table 4 starts with the ability of the tests to detect differences between RTV and EOS in terms of forecast performance. Firstly, the rejections rates of equal accuracy tests clearly diminish in ϕ for news revisions, but are increasing in ϕ for noise. Rejection rates on log scores tend to be higher than using CRPS: the lower rejection rates of EOS using CRPS are especially marked when data revisions are noise. This may be because the CRPS loss function favourably offsets narrower intervals (a feature of EOS when there is noise) against lower than expected coverage, resulting in fewer rejections of EOS. For this reason it may be preferable to use log score to evaluate the relative performance of EOS and RTV.

Secondly, it is clear that EOS and RTV forecasting performances are not very different if data revisions are relatively small, corresponding to $\delta = 0.3$ in our design (see Table 4A), as opposed to $\delta = 1$. But even for $\delta = 0.3$, many of the tests have reasonable power to reject the EOS forecasts when revisions are noise, although the rejection rates are markedly lower for the EOS forecasts when revisions are news. This is consistent with the large gains to using RTV, instead of EOS, when data revisions are noise, shown in Figure 1.

Thirdly, the conditional aspect of the comparison (given by the ID tests) appears to add relatively little: the rejections of the EOS intervals are primarily caused by incorrect (uncondi-

[6] Evaluation based on the PIT dates back to Rosenblatt (1952), with recent contributions including Shephard (1994), Kim *et al.* (1998) and Diebold, Gunther and Tay (1998).

tional) coverage, and the rejection rates of correct conditional coverage are rarely very different. However, for $\delta = 1$ and $\phi = 0.1$, the rejection rate of ID when there is GARCH is nearly 50% (0.46 in the table), whereas that for SV is markedly lower (0.15). For higher values of ϕ the power of the ID test is seen to diminish.

Note there are no circumstances in which RTV is actually worse than EOS, suggesting the primacy of RTV. That said, there are circumstances when the two are very similar. Naturally, when revisions are small, but also if revisions are news and the process is persistent.

In summary, if data revisions are large enough we should be able to find improvements in probabilistic forecasting in real time from using RTV instead of EOS data, regardless of whether there is conditional heteroscedasticity. The two tests which are most likely to flag up the difference between RTV and EOS are the test for equal accuracy based on log score differences, and the predictive interval coverage rate tests.

5 Empirical Real-Time Forecasting Comparisons

In this section, we compare EOS and RTV for autoregressive models with constant volatility, with ARCH(4) volatility and with stochastic volatility. We also evaluate Bayesian VARs with stochastic volatility. As in the previous sections, we are interested in one-step-ahead forecasts where the realization is a first release, as is typical in nowcasting exercises (see, e.g., Giannone, Reichlin and Small (2008) and Bańbura, Giannone and Reichlin (2011)).

We evaluate forecasts of eight quarterly macroeconomic time series and six monthly macroeconomic time series. Their acronyms and descriptions are in Table 5. These time series were obtained from the Philadelphia Fed real-time dataset, with the exception of the two monthly series as described in the note to Table 5. We use data from 1985, such that the time series cover 30 years from 1985 to 2014. Table 5 provides information on the characteristics of the data revisions to each series, to enable an assessment of the extent to which the empirical results in this section are consistent with the analysis of the previous sections. Recall a key finding was that differences between RTV and EOS would likely be larger for noise, as opposed to news, revisions. Moreover, for processes with noise revisions, the difference between RTV and EOS is expected to increase in the degree of persistence of the underlying process, and in the relative variability of revisions.

We test whether data revisions are either news or noise by assuming that the revision is the difference between the release published 14 quarters after the observational quarter y_t^{t+14} and the first release y_t^{t+1}. This includes initial revisions and three rounds of annual revisions. We use equivalent values for the case of monthly series, that is, the difference between release published 42 months after the observational month and the first release. All series are in first differences in logs, except the trade balance (XM), which is in levels.

Motivated by the discussion of news and noise revisions in section 2.2, we test for news and noise revisions using, respectively, the auxiliary regressions

$$y_t^{t+14} - y_t^{t+1} = \alpha + \beta y_t^{t+1} + \epsilon_t \tag{33}$$

and:

$$y_t^{t+14} - y_t^{t+1} = \alpha + \beta y_t^{t+14} + \epsilon_t. \tag{34}$$

In (33) the null is that revisions are news, H_0: $\beta = 0$, in which case the revision is not predictable given the initial release, y_t^{t+1}. In (34) the null is that revisions are noise, H_0: $\beta = 0$, in which case the revision is not correlated with the later value (or the true estimate).

The results suggest that we can identify revisions as either news or noise for most of the variables, but in some cases, data revisions might be a combination of both. A variable is classified as having news revisions if we fail to reject the null in (33) but reject $\beta = 0$ in (34). And the reverse set of test outcomes indicates noise. Rejection in both (33) and (34), or alternatively, failure to reject in both, gives an indeterminate classification (in terms of the news-noise dichotomy). Table 5 also shows the following ratios: $var(y_t^{t+n} - y_t^{t+1})/var(y_t^{t+1})$ and $var(y_t^{t+n} - y_t^{t+1})/var(y_t^{t+n})$, where $n = 14$ for quarterly series and $n = 42$ for monthly series. These ratios are a measure of the relative size of data revisions compared to the variability of the series. When data revisions are identified as noise, we calculate the ratio relative to the variability of the revised revised data, and for news, relative to the first release. Based on these calculations, the most sizeable noise revisions are for government spending (G) and hours (Ho). The largest news revisions are for nominal wages, GDP and industrial production. For all cases except for nominal wages, the magnitude of these empirical data revisions is comparable to those of the simulated data of section 4 when $\delta = 0.3$. Data revisions for monthly series tend to be smaller than for quarterly series.

5.1 Univariate Models

For the forecasting exercises that follow, we set the in-sample period as 1985-1999, and the out-of-sample period as 2000-2014 (details are available in the table notes). Models are estimated with rolling windows of $T = 60$ ($T = 180$) observations with quarterly (monthly) data. The out-of-sample period with quarterly (monthly) series has $P = 60$ ($P = 180$) observations.

Table 6 presents results for three forecasting models. An AR(p) with $p = 1$ for quarterly series and $p = 3$ for monthly series, the same model augmented by ARCH(4) for the volatility, and the model augmented with stochastic volatility estimates as described in section 4. The log scores presented in Table 6 are only for EOS. Table 6 addresses whether density forecasts can be improved by adding conditional heteroscedasticity in the volatility of the disturbances. The last two columns show the t-statistic of the test of equal log scores with the AR(p) under the null, with negative values showing that the model with conditional heteroscedasticity performs better. These results indicate that the inclusion of ARCH improves forecasts in few cases, but stochastic volatility improves forecasts in many cases, for the quarterly and monthly series.

Tables 7 to 10 show results for EOS versus RTV using three different models: AR(p), AR(p) + ARCH(4), and AR(p) + SV. For quarterly data, we set $p = 1$ and $p = 3$ for monthly data. Results are robust to setting $p = 2$ for quarterly data and $p = 12$ for monthly data (results available on request). Table 7 compares point forecasts using RMSFEs. The t-statistic is for a Diebold-Mariano test, as also used in section 4. Table 8 compares density forecasts using log scores and the test for equal accuracy: the section 4 simulation results suggested log score as the loss function with the greater power to discriminate between RTV and EOS. Significant negative t-statistics favour RTV models. Table 9 provides empirical coverages for 90% nominal intervals and reports p-values of the test for conditional coverage as described in section 4. Table 10 provides p-values of the test for calibration and independence of density forecasts (Berkowitz test), again as described in section 4.

Figure 3 shows the measures of forecasting uncertainty at each forecasting origin in the out-of-sample period computed with the AR-SV model for quarterly series. Figure 4 shows the results for monthly series. Figure 5 shows the histogram of the PITs computed for the AR-SV with quarterly data. It compares the results for EOS and RTV.

Although the main focus of this paper is density forecasting we include RMSFEs results and accuracy tests for completeness. Results in Table 7 suggest that for many series, we find that RTV-based *point* forecasts are statistically better than EOS, particularly when looking at ARCH models for quarterly nominal series (W, P, PCE), as well as for the real series C and XM. There are two cases where EOS outperforms RTV across the three model specifications in the table: for forecasting changes in government spending (G) and hours (Ho).

The log score results in Table 8 indicate fewer rejections in favour of RTV modelling when we consider densities. But RTV in general improves forecasts for the three quarterly nominal series (W, P, PCE) when employing the AR-SV model. As in the case of point forecasting, RTV fares significantly worse than EOS for G and Ho. This is a puzzle because these two variables have the larger noise revisions which ought to favour RTV over EOS.

Figures 3 and 4 present the time series of the one-step ahead predictive variances estimated using the AR(p)-SV model with EOS and RTV data. These results lend support to the analytical results which indicate $\sigma^{2,EOS}_{T+1|T} > \sigma^{2,RTV}_{T+1|T}$ when data revisions are news, and $\sigma^{2,EOS}_{T+1|T} < \sigma^{2,RTV}_{T+1|T}$ for noise. These figures can also help us to understand why RTV is worse than EOS in the case of G and Ho. In both cases, the predictive variances (both EOS and RTV) are seen to decline throughout the first part of the out-of-sample period. The simulation results in section 2.4 suggest that if data revisions are noise (as it is indicated in both cases) and there is a break that reduces the variance of the underlying process, we could observe EOS delivering a better performance than RTV.

The coverage and conditional coverage test results in Table 9 indicate that the autoregressive model with constant variance is normally not able to deliver good prediction intervals, but if conditional heteroscedasticity is added, we can find a model with good performance. For the quarterly series, the best model is normally the AR(1)-SV model, except in the case of consumption. Moreover, EOS and RTV in these cases tend to be reasonably similar, except in the case of the nominal variables (W, P) where the coverage rates and the size of the p-values give support to RTV. Our analytical work suggests that RTV and EOS might be more closely aligned when the series is characterized by conditional heteroscedasticity. For the monthly series, the AR(3)-ARCH(4) is the best model for payroll employment and sales, while the AR(3)-SV is better for the remaining variables. However, in all these cases, the conditional coverage test suggest EOS and RTV perform similarly.

The test for calibration and independence of the PITs in Table 10 indicates the null hypothesis is rejected for EOS, but not RTV, for some of the quarterly series. In addition to the nominal wages and the deflator, the RTV improvements are also found for GDP and the trade balance. Figure 5 shows histograms of the PITs as an aid to interpreting these results. The histogram is based on five equal-sized bins, so under the null hypothesis we expect to find 20% of the PITs observations in each bin (given that the PITs are uniformly distributed). The improvements found for GDP are mainly related to the right tail of the distribution, with the lower predictive variances of the RTV approach resulting in more observations in the right tail of the predictive density, and consequently more PITs in the right-most bin. In the case of nominal wages, EOS delivers an inverted-V shaped histogram, suggesting that the predictive variances are too large, while the RTV histogram is more uniformly distributed.

5.2 Multivariate Model

In addition to these univariate models, we also evaluate the use of RTV and EOS in a VAR with four variables. This is motivated by Clark (2011), who shows the importance of allowing for SV for generating accurate real-time density forecasts of inflation. Following Clark (2011), we use a small Bayesian vector autoregression (BVAR), consisting of four variables: the first differences of the logs (growth rates) of GDP (Y), the GDP deflator (P), and consumption (C), and the level of the fed fund rate (R). Relative to Clark (2011), we use C in place of the unemployment series, so we use only data from the real-time dataset analyzed thus far. The Fed fund rate series is of course not subject to revisions. We use a BVAR model with stochastic volatility.[7] In contrast with the SV models considered earlier, this model assumes that the variance follows a random walk, that is, $\kappa_0^{EOS} = \kappa_0^{RTV} = 0$ and $\kappa_1^{EOS} = \kappa_1^{RTV} = 1$. This means that the only parameter of the conditional volatility process allowed to differ between EOS and RTV is the

[7] We use the code for the Carriero, Clark and Marcellino (2016) paper made available on M. Marcellino's webpage.

innovation variance to the volatility equation. This approach is able to capture slow-moving changes in the conditional variance without assuming an equilibrium value.

When applying the RTV approach to the BVAR for the vector of endogenous variables $\mathbf{y}_t$, the LHS variable consists of vectors of observations $\left\{\mathbf{y}_t^{t+1}\right\}_{t=p+1}^{t=T}$, that is, the time series of first releases for each variable. The right-hand side variables consist of:

$$\mathbf{x}_{t-1}^t = (1, \mathbf{y}_{t-1}^t, \mathbf{y}_{t-2}^t, ..., \mathbf{y}_{t-p}^t), \text{ for } t = p+1, ..., T,$$

that is, all lags are taken from the vintage at t, so for more than one lag ($p > 1$), at least partially-revised data are used. As a consequence, data employed in the estimation of the model and in the evaluation of forecasts has the same maturity, as discussed in section 3. In the case of EOS data, all data is from the latest-available release, that is, $\mathbf{y}_t^{T+1}$ and $\mathbf{x}_{t-1}^{T+1}$, for $t = p+1, ..., T$.

Table 11 presents measures of the forecasting performance of RTV and EOS using the BVAR with stochastic volatility and setting $p = 4$. The table includes log scores, coverage and p-values of test for conditional coverage and the Berkowitz test for each endogenous variable in the VAR, employing the same forecasting design as in Tables 7 to 10. The coverage and log score values are not very different from the ones obtained with the AR(1)-SV model. Based on the coverage rates of nominal 90% intervals, RTV performs better for output growth, inflation and the Fed fund rate, in that the actual coverage rate is closer to the nominal, but of these three variables only for the Fed fund rate are the EOS intervals rejected in terms of correct conditional coverage. The null of correct conditional coverage is rejected for the EOS intervals for consumption. Whereas the interval results lend some support to RTV, the EOS density forecasts are preferred on log score, although the differences between the EOS and RTV log scores are not statistically significant for GDP and inflation.

In summary, the results are rather mixed for the four variable BVAR, and do not clearly favour either RTV or EOS. As established analytically, significant differences between EOS and RTV might not be expected in the presence of conditional heteroscedasticity, and when revisions are characterized by news, as opposed to noise. We note that none of the three series subject to revision are unambiguously subject to noise revisions (see table 5).

6 Conclusions

In this paper, we analyze a simple method with the potential of improving the accuracy of real-time probabilistic forecasts of macroeconomic variables subject to revisions. We show that in principle, when there are data revisions, RTV provides predictive densities which more closely approximate the true unknown density, compared to the standard approach to forecasting, which is simply to estimate the model on the vintage of data available at the forecast origin.

The analytical results indicate that the improvements to RTV are likely to be larger when data revisions can be characterized as primarily reducing noise as opposed to adding in news (in terms of the distinction due to Mankiw and Shapiro (1986)), when revisions are large, and when the underlying process is persistent. We also establish that unmodelled shifts or breaks in variance can either accentuate or diminish the relative advantage of RTV depending on whether the shift is upward or downward. Finally, our analytical results suggest that the costs of using EOS instead of RTV are reduced if there is time-varying volatility, and this is modelled.

Empirically, we find evidence of improvements from using RTV for univariate model forecasts of some nominal quarterly variables (nominal wage and GDP deflator inflation), where the model is specified with an SV error term. The results for a multivariate model using four key variables were less encouraging, although we argue this is consistent with our findings for variables whose revisions are not clearly noise.

References

Bańbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting, chapter 7. In Clements, M. P., and Hendry, D. F. (eds.), *Forecasting Handbook*, pp. 193–224: Oxford University Press.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, **19(4)**, 465–474.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **51**, 307–327.

Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, *176*(2), 162–172.

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

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., and McCracken, M. (2013). *Advances in Forecast Evaluation*, Vol. 2 of *Handbook of Economic Forecasting*, pp. 1107–1201: Elsevier.

Clark, T. E., and Ravazzolo, F. (2015). 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*. Forthcoming.

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 Taylor, N. (2003). Evaluating prediction intervals for high-frequency data. *Journal of Applied Econometrics*, **18**, 445 – 456.

Cogley, T., and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post World War II US. *Review of Economic Dynamics*, **8**, 262–302.

Corradi, V., and Swanson, N. R. (2006). Predictive density evaluation. In Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 197–284: Elsevier, Horth-Holland.

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.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts: With applications to financial risk management. *International Economic Review*, **39**, 863–883.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.

Diebold, F. X., Schorfheide, F., and Shin, M. (2016). Real-time forecast evaluation of DSGE models with Stochastic Volatility. Mimeo, University of Pennsylvania.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, **8**, 985–987.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting GDP and Inflation: The real-time informational content of macroeconomic data releases. *Journal of Monetary Economics*, **55**, 665–676.

Gneiting, T., and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, **1**, 125–51.

Hall, S. G., and Mitchell, J. (2009). Recent developments in density forecasting. In Mills, T. C., and Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 199–239: Palgrave MacMillan.

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

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review* *of Economic Studies*, **61**, 247–64.

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.

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.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, **12**, 371–89.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, *65*(3), 361–393.

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.

Lee, T.-H., Bao, Y., and Saltoglu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, *26*(3), 203–225.

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.

McConnell, M. M., and Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980s?. *American Economic Review*, **90**, 1464–1476.

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.

Pesaran, M. H., and Timmermann, A. (2007). Selection of estimation window in the presence of structural breaks. *Journal of Econometrics*, **137(1)**, 134–161.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, **72**, 821–852.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, **23**, 470–472.

Sensier, M., and van Dijk, D. (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics*, *86*(3), 833–839.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, **81**, 115–131.

Shephard, N. G. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E. (eds.), *Time Series Models: In econometrics, finance and other fields*. London: Chapman and Hall.

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.

Taylor, S. J. (2005). *Asset Price Dynamics, Volatility,and Prediction*: Princeton University Press.

A Appendix

A.1 Proof of Proposition 1

The difference between EOS and RTV log score is computed as:

$$\begin{aligned}
\Delta score^{News} &= E[-\ln(p^{EOS}_{T+1|T}(y^{T+2}_{T+1}))] - E[-\ln(p^{RTV}_{T+1|T}(y^{T+2}_{T+1}))] \\
&= E[-\ln(p^{EOS}_{T+1|T}(y_{T+1} - \sigma_v \xi_{2T+1}))] - E[-\ln(p^{RTV}_{T+1|T}(y_{T+1} - \sigma_v \xi_{2T+1}))] \\
&= \left[\frac{1+\phi^2\delta}{2(1+\delta)} + \frac{1}{2}\ln(\sigma^2_\eta(1+\delta))\right] - \left[\frac{1}{2} + \frac{1}{2}\ln\sigma^2_\eta(1+\phi^2\delta)\right] \\
&= \frac{1}{2}\left[\frac{1+\phi^2\delta}{1+\delta} + \ln(1+\delta) - 1 - \ln(1+\phi^2\delta)\right] \\
&= \frac{1}{2}\left[\frac{\delta(\phi^2-1)}{1+\delta} + \ln[(1+\delta)/(1+\phi^2\delta)]\right],
\end{aligned}$$

and we need to show that $\Delta score^{News} \geq 0$ in order to establish the dominance of RTV over EOS on log score. If we take the derivative of the expression in brackets with respect to ϕ^2, we find $\delta/(1+\delta) - \delta/(1+\phi^2\delta)$. Because $(1+\delta) > (1+\phi^2\delta)$, since $\phi^2 < 1$ and $\delta \geq 0$, the derivative is always negative. This means that the minimum value of $\Delta score^{News}$ will be at the maximum value of ϕ^2, that is, $\phi^2 \approx 1$. Based on the expression above, it is clear that if $\phi^2 = 1$, $\Delta score^{News} = 0$. If $\Delta score^{News}$ is equal to zero at its minimum, then for values such that $0 \leq \phi^2 < 1$, we have $\Delta score^{News} \geq 0$.

A.2 Computation of RTV Predictive Variance under Noise Revisions

The expression (10) is calculated as the in-sample error variance of RTV, as:

$$\begin{aligned}
\sigma^{2,RTV}_{T+1|T} &= var(y_{T+1} + \sigma_\varepsilon \xi_{3T+1} - B\phi y_T - B\phi\sigma_\varepsilon \xi_{3T}) \\
&= var(y_t) + B^2\phi^2 var(y_t) + (1+B^2\phi^2)\sigma^2_\varepsilon - 2B\phi Cov\,(y_t y_{t-1}) \\
&= \sigma^2_y(1 + B^2\phi^2 - 2B\phi^2) + (1+B^2\phi^2)\sigma^2_\varepsilon \\
&= \sigma^2_y[(1-\phi^2) + \phi^2(B-1)^2] + (1+B^2\phi^2)\sigma^2_\varepsilon \\
&= \sigma^2_\eta/(1-\phi^2)[(1-\phi^2) + \phi^2(B-1)^2] + \sigma^2_\eta\delta(1+B^2\phi^2) \\
&= \sigma^2_\eta + \sigma^2_\eta(\phi^2(B-1)^2/(1-\phi^2)) + \sigma^2_\eta\delta + \sigma^2_\eta\delta B^2\phi^2 \\
&= \sigma^2_\eta + \delta\sigma^2_\eta + \sigma^2_\eta(\phi^2(B-1)^2/(1-\phi^2)) + \sigma^2_\eta\delta B^2\phi^2 \\
&= \sigma^2_\eta(1+\delta+\varrho),
\end{aligned}$$

where the second line assumes stationarity.

A.3 Proof of Proposition 2

The difference between EOS and RTV log score is computed as:

$$\begin{aligned}
\Delta score^{Noise} &= E[-\ln(p^{EOS}_{T+1|T}(y^{T+2}_{T+1}))] - E[-\ln(p^{RTV}_{T+1|T}(y^{T+2}_{T+1}))] \\
&= E[-\ln(p^{EOS}_{T+1|T}(y_{T+1} + \sigma_\varepsilon \xi_{3T+1}))] - E[-\ln(p^{RTV}_{T+1|T}(y_{T+1} + \sigma_\varepsilon \xi_{3T+1}))] \\
&= \left[\frac{1}{2}(1+\delta(1+\phi^2)) + \frac{1}{2}\ln(\sigma^2_\eta)\right] - \left[\frac{1}{2} + \frac{1}{2}\ln(\sigma^2_\eta(1+\delta+\varrho))\right] \\
&= \frac{1}{2}\left[(1+\delta(1+\phi^2)) - 1 - \ln(1+\delta+\varrho)\right] \\
&= \frac{1}{2}\left[(\delta(1+\phi^2)) - \ln(1+\delta+\varrho)\right].
\end{aligned}$$

To show that $\Delta score^{Noise} \geq 0$, we use the concavity of the logarithm function. But first note that we can rewrite ϱ as:

$$\begin{aligned}
\varrho &= [\phi^2(B-1)^2/(1-\phi^2) + \delta B^2 \phi^2] \\
&= \phi^2(1-\phi^2)^{-1}\left[\frac{\delta^2}{((1-\phi^2)^{-1}+\delta)^2}\right] + \delta\phi^2\left[\frac{(1-\phi^2)^{-2}}{((1-\phi^2)^{-1}+\delta)^2}\right] \\
&= \frac{\phi^2(1-\phi^2)^{-1}\delta(\delta + (1-\phi^2)^{-1})}{((1-\phi^2)^{-1}+\delta)^2} \\
&= \frac{\phi^2(1-\phi^2)^{-1}\delta}{((1-\phi^2)^{-1}+\delta)} = \phi^2\delta B
\end{aligned}$$

Recall that $x \geq \ln(1+x)$ if $x \geq 0$. In the case that $B = 1$, $\varrho = \delta\phi^2$, and then $\delta(1+\phi^2) > \ln(1+\delta(1+\phi^2))$. When $\sigma^2_\varepsilon > 0$, then $B < 1$, and we have:

$$\Delta score^{Noise} = \frac{1}{2}(\delta(1+\phi^2)) - \ln(1+\delta(1+\phi^2 B)).$$

Since $\phi^2 B < \phi^2$, then it must be the case that $\Delta score^{Noise} \geq 0$, establishing the dominance of RTV over EOS on log score for noise revisions.

A.4 Computation of w^{EOS}_0 and w^{EOS}_1 for News Revisions

Our objective is to compute the population values of the parameters of the variance equation in (24) when the data generating process is given by (4) and (12) and data revisions are pure news.

Firstly, it follows that $E\left((e^{EOS}_t)^2\right) = E\left[(\eta_t + v_t)^2\right] = \sigma^2_\eta + \sigma^2_v = w^{EOS}_0\left(1 - w^{EOS}_1\right)^{-1}$. Note that:

$$e_t^{EOS} = y_t^{T+1} - \phi y_{t-1}^{T+1} = y_t - \phi y_{t-1} = \eta_t + v_t$$

for $t < T$. When $t = T$, the actual value is a first release (and does not contain the news component), so that $e_T^{EOS} = \eta_T$.

Secondly, w_1^{EOS} is equal to the first-order autocorrelation coefficient of $(e_t^{EOS})^2$, that is,

$$w_1^{EOS} = \frac{Cov\left((e_t^{EOS})^2, (e_{t-1}^{EOS})^2\right)}{Var\left((e_t^{EOS})^2\right)}.$$

The numerator is then $Cov\left((e_t^{EOS})^2, (e_{t-1}^{EOS})^2\right) = Cov\left(\eta_t^2 + v_t^2 + 2\eta_t v_t, \eta_{t-1}^2 + v_{t-1}^2 + 2\eta_{t-1}v_{t-1}\right) = Cov\left(\eta_t^2, \eta_{t-1}^2\right) = w_1 Var\left(\eta_t^2\right)$ [because all the other covariances can be shown to be equal to zero.] The denominator is $Var\left((e_t^{EOS})^2\right) = Var\left(\eta_t^2 + v_t^2 + 2\eta_t v_t\right) = Var\left(\eta_t^2\right) + Var\left(v_t^2\right) + 4Var\left(\eta_t v_t\right)$ [because the covariance terms can all be shown to equal zero], where $Var\left(v_t^2\right) = E\left[\left(v_t^2 - E\left(v_t^2\right)\right)^2\right] = E\left(v_t^4\right) - \left(E\left(v_t^2\right)\right)^2 = 3\sigma_v^4 - \sigma_v^4 = 2\sigma_v^4$ and $4Var\left(\eta_t v_t\right) = 4E\left[\left(\eta_t v_t - E\left(\eta_t v_t\right)\right)^2\right] = 4E\left(\eta_t^2 v_t^2\right) = 4\sigma_\eta^2 \sigma_v^2$. These results imply that

$$w_1^{EOS} = w_1 \frac{var\left(\eta_t^2\right)}{var\left(\eta_t^2\right) + 2\sigma_v^4 + 4\sigma_\eta^2 \sigma_v^2}$$

For $var\left(\eta_t^2\right)$, we can use the results for kurtosis in Taylor (2005), that is,

$$var(\eta_t^2) = 3\sigma_\eta^4 \left[\frac{1 - {w_1}^2}{1 - 3w_1^2}\right]$$

for an ARCH(1) model.

B EOS forecasting with GARCH and SV Models

In sections 3.3.1 and 3.3.2 we provide results for EOS for time-varying heteroscedasticity given by an ARCH(1) model. In this appendix, we extend those results to a GARCH$(1,1)$ and a stochastic volatility model.

B.1 Noise Revisions

We assume the data generating process includes a GARCH$(1,1)$ model for the underlying disturbance η_t, and that the EOS forecasting model contains a GARCH$(1,1)$ model for it's error:

$$\begin{aligned}\sigma^{2EOS}_{T+1|T} &= w_0^{EOS} + w_1^{EOS}\left(e_T^{EOS}\right)^2 + w_2^{EOS}h_T^{EOS} \\ &= w_0 + w_1(\eta_T^2 + \varepsilon_T^2) + w_2 h_T^{EOS} \\ &= w_0 + w_1\varepsilon_T^2 + (w_1\xi_T^2 + w_2)h_T\end{aligned}$$

where $\xi_T \sim iidN(0,1)$, and we have substituted $\eta_T^2 = \xi_T^2 h_T$ to obtain the last line. In the above expression we have used the fact that the GARCH model estimates in the EOS approach are equal in population to their values in the DGP, and that $h_T^{EOS} = h_T$.

The unconditional variance is then:

$$\begin{aligned}E[\sigma^{2EOS}_{T+1|T}] &= \sigma_\eta^2 + w_1\sigma_\varepsilon^2 + (w_1 + w_2)(h_T - \sigma_\eta^2) \\ &= \sigma_\eta^2(1 + w_1\delta),\end{aligned}$$

which is the same as expression (28) for the ARCH model, with similar implications for EOS/RTV relative performance.

In the case of stochastic volatility models, we also assume that EOS returns the DGP values in large samples: $\kappa_1^{EOS} = \kappa_{1\eta}$ and that $\sigma^{EOS} = \sigma_1$. This implies that the predictive EOS variance is:

$$\begin{aligned}E[\sigma^{2EOS}_{T+1|T}] &= \exp\{\ln\sigma_\eta^2 + \kappa_{1\eta}E[\ln(\eta_T^2 + \varepsilon_T^2) - \ln(\sigma_\eta^2)] + 0.5[(\sigma_1^2/(1 - \kappa_{1\eta}^2)]\} \\ &= \exp\{\ln\sigma_\eta^2 + \kappa_{1\eta}\ln\sigma_\varepsilon^2 + 0.5[(\sigma_1^2/(1 - \kappa_{1\eta}^2)]\}.\end{aligned}$$

The inclusion of stochastic volatility diminishes the under-estimation of the true predictive variance by the the EOS approach since it includes now a term that depends on σ_ε^2. Note however that RTV unconditional predictive variance will be:

$$E[\sigma^{2RTV}_{T+1|T}] = \exp\{\ln\sigma_\eta^2 + \ln\sigma_\varepsilon^2 + \ln\varrho\sigma_\eta^2 + 0.5[(\sigma^{2,RTV}/(1 - (\kappa_1^{RTV})^2)]\},$$

and as a consequence, we still expect EOS to under-estimate the variance on average.

B.2 News Revisions

In the case of GARCH (1,1) EOS model, we have that $w_0^{EOS} \neq w_0$, $w_1^{EOS} \neq w_1$ and $w_2^{EOS} \neq w_2$, since the GARCH$(1,1)$ model for EOS is effectively estimated for $\eta_t + v_t$, rather than for η_t, as in the data generation process. The predicted variance is then:

$$\begin{aligned}\sigma^{2EOS}_{T+1|T} &= w_0^{EOS} + w_1^{EOS}\eta_T^2 + w_2^{EOS}h_T^{EOS} \\ &= (\sigma_\eta^2 + \sigma_v^2) + w_1^{EOS}(\eta_T^2 - (\sigma_\eta^2 + \sigma_v^2)) + w_2^{EOS}(h_T^{EOS} - (\sigma_\eta^2 + \sigma_v^2))\end{aligned}$$

The first line has substituted $e_T^{EOS} = \eta_T$. The unconditional variance is:

$$\begin{aligned} E[\sigma^{2EOS}_{T+1|T}] &= (\sigma^2_\eta + \sigma^2_v) - w_1^{EOS}\sigma^2_v \\ &= \sigma^2_\eta(1 + (1 - w_1^{EOS})\delta), \end{aligned}$$

which is the same as in the case of an ARCH(1) model such that implications for EOS/RTV relative performance are similar.

In the case of stochastic volatility models, the unconditional predictive EOS variance is:

$$\begin{aligned} E[\sigma^{2EOS}_{T+1|T}] &= \exp\{\ln(\sigma^2_\eta + \sigma^2_v) + \kappa_1^{EOS} E[\ln(\eta^2_T) - \ln(\sigma^2_\eta + \sigma^2_v)] + 0.5[(\sigma^{2EOS}/(1 - (\kappa_1^{EOS})^2)]\} \\ &= \exp\{\ln(\sigma^2_\eta) + (1 - \kappa_1^{EOS})\ln(\sigma^2_v) + 0.5[(\sigma^{2EOS}/(1 - (\kappa_1^{EOS})^2)]\}, \end{aligned}$$

which should reduce the over-estimation of the variance by EOS since $\kappa_1^{EOS} > 0$, but the value is still larger than the RTV unconditional variance:

$$E[\sigma^{2RTV}_{T+1|T}] = \exp\{\ln\sigma^2_\eta + 0.5[(\sigma^{RTV,2}/(1 - (\kappa_1^{RTV})^2)]\}.$$

Table 1: Simulation values for differences between EOS and RTV forecasting performance in terms of logscore, CRPS and quantile loss function for a large sample size (T=500) with AR(1) forecasting models.

News					
δ	ϕ	Analytical Logscore	Simulated Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.009	0.007	0.090	0.036
1	0.1	0.094	0.092	1.480	0.478
1	0.5	0.048	0.048	0.760	0.258
1	0.9	0.002	0.002	0.005	0.015
Noise					
δ	ϕ	Analytical Logscore	Simulated Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.033	0.032	0.490	0.147
1	0.1	0.157	0.148	1.440	0.602
1	0.5	0.244	0.235	3.430	0.939
1	0.9	0.412	0.397	4.810	1.577

Note: Values are computed using 5,000 monte carlo replications. Quantile loss uses level of 10%. δ is the data revision size (0.3 is small and 1 is large), and ϕ is the value of the autoregressive coefficient.

Table 2: Simulation values for differences between EOS and RTV forecasting performance in terms of in logscore, including results with a break (T=200) with AR(1) forecasting models.

News					
δ	ϕ	Analytical	Simulated	With Break: $0.5\sigma_\eta$	With Break: $-0.5\sigma_\eta$
0.3	0.5	0.009	0.0059	-0.0821	0.0693
1	0.1	0.094	0.0876	-0.1886	0.2816
1	0.5	0.048	0.0446	-0.1243	0.1605
1	0.9	0.002	0.0009	-0.0280	0.0204
Noise					
δ	ϕ	Analytical	Simulated	Break: $0.5\sigma_\eta$	Break: $-0.5\sigma_\eta$
0.3	0.5	0.033	0.0391	0.1736	-0.0578
1	0.1	0.157	0.1606	0.4127	-0.0145
1	0.5	0.244	0.2631	0.5160	0.0854
1	0.9	0.412	0.4498	0.7410	0.2470

Note: Values are computed using 2,000 monte carlo replication. The break date is set as 0.95T.

Table 3A: Simulation values for differences between EOS and RTV forecasting performance in terms of logscore, CRPS and quantile loss function with AR(p) forecasting models. DGP has p=1.

News							
		T=200			T=500		
δ	ϕ	Logscore	CRPS (*100)	Quantile Loss	Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.006	0.119	0.038	0.009	0.129	0.039
1	0.1	0.086	1.457	0.462	0.086	1.387	0.446
1	0.5	0.042	0.729	0.231	0.041	0.747	0.233
1	0.9	-0.001	0.035	-0.016	0.002	0.039	0.013
Noise							
		T=200			T=500		
δ	ϕ	Logscore	CRPS (*100)	Quantile Loss	Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.039	0.478	0.160	0.028	0.439	0.109
1	0.1	0.175	1.636	0.713	0.152	1.414	0.664
1	0.5	0.263	3.507	1.108	0.237	3.243	1.049
1	0.9	0.481	8.330	2.044	0.440	8.228	1.940

Note: See notes of Table 1. Data generating process has p=1, but we allow the AR(p) forecasting models to pick up the lag order using SIC assuming pmax=4.

Table 3B: Simulation values for differences between EOS and RTV forecasting performance in terms of logscore, CRPS and quantile loss function with AR(p) forecasting models. DGP has p=2

News							
		T=200			T=500		
δ	ϕ	Logscore	CRPS (*100)	Quantile Loss	Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.012	0.289	0.056	0.015	0.219	0.061
1	0.1	0.086	1.431	0.461	0.098	1.543	0.499
1	0.5	0.078	1.555	0.419	0.086	1.407	0.437
1	0.9	0.049	0.888	0.265	0.057	0.937	0.310
Noise							
		T=200			T=500		
δ	ϕ	Logscore	CRPS (*100)	Quantile Loss	Logscore	CRPS (*100)	Quantile Loss
0.3	0.5	0.028	0.190	0.114	0.018	0.184	0.068
1	0.1	0.172	1.528	0.696	0.141	1.365	0.604
1	0.5	0.195	1.814	0.797	0.162	1.844	0.711
1	0.9	0.249	3.135	1.035	0.211	2.986	0.934

Note: See notes of Table 1. Data generating process has p=2 where $\phi 1 = \phi/2$ and $\phi 2 = \phi/2$, but we allow the AR(p) forecasting models to pick up the lag order using SIC assuming pmax=4. Analytical difference in logscore is computed using only the first order AR coefficient ($\phi/2$).

Table 4. Proportion of rejections: tests of equal density accuracy, density accuracy and predictive interval accuracy for rolling window forecasting with T= 200 and P=100 for different revision sizes (δ) and AR(1) parameter (ϕ)

Table 4A: $\delta = 0.3;\ \phi = 0.5$

	News				Noise			
	AR	ARCH	GARCH	SV	AR	ARCH	GARCH	SV
logscore_t	0.22	0.19	0.28	0.31	0.42	0.46	0.60	0.35
CRPS_t	0.18	0.17	0.28	0.29	0.27	0.30	0.33	0.25
Quantl_t	0.20	0.18	0.27	0.32	0.34	0.33	0.39	0.30
	EOS							
PITs	0.27	0.27	0.41	0.35	0.66	0.66	0.91	0.62
UC_50	0.17	0.17	0.33	0.21	0.36	0.30	0.42	0.40
ID_50	0.07	0.16	0.17	0.11	0.10	0.17	0.19	0.12
CC_50	0.17	0.20	0.32	0.21	0.31	0.32	0.43	0.33
UC_90	0.24	0.21	0.31	0.34	0.59	0.65	0.84	0.60
ID_90	0.07	0.16	0.13	0.04	0.14	0.26	0.35	0.10
CC_90	0.16	0.24	0.32	0.25	0.57	0.65	0.83	0.52
	RTV							
PITs	0.09	0.09	0.12	0.10	0.09	0.14	0.17	0.08
UC_50	0.08	0.08	0.09	0.10	0.08	0.09	0.10	0.08
ID_50	0.09	0.11	0.09	0.13	0.10	0.11	0.11	0.12
CC_50	0.09	0.09	0.10	0.12	0.09	0.11	0.12	0.09
UC_90	0.07	0.10	0.13	0.09	0.08	0.10	0.12	0.10
ID_90	0.11	0.14	0.11	0.09	0.11	0.14	0.13	0.12
CC_90	0.08	0.11	0.11	0.07	0.07	0.10	0.11	0.10

Table 4B: $\delta = 1;\ \phi = 0.1$

	News				Noise			
	AR	ARCH	GARCH	SV	AR	ARCH	GARCH	SV
logscore_t	0.83	0.85	0.90	0.88	0.87	0.91	0.96	0.84
CRPS_t	0.77	0.79	0.89	0.83	0.57	0.63	0.67	0.57
Quantl_t	0.78	0.82	0.91	0.84	0.77	0.80	0.85	0.74
	EOS							
PITs	0.99	0.95	0.93	0.99	0.99	0.99	1.00	0.98
UC_50	0.90	0.92	0.94	0.92	0.84	0.64	0.79	0.92
ID_50	0.11	0.35	0.46	0.15	0.08	0.37	0.43	0.12
CC_50	0.84	0.91	0.97	0.91	0.79	0.71	0.83	0.81
UC_90	0.98	0.88	0.84	0.98	0.99	0.98	1.00	0.99
ID_90	0.01	0.19	0.18	0.04	0.11	0.55	0.69	0.10
CC_90	0.92	0.90	0.86	0.95	0.98	0.98	1.00	0.99
	RTV							
PITs	0.10	0.07	0.10	0.08	0.08	0.11	0.14	0.08
UC_50	0.07	0.08	0.10	0.11	0.10	0.09	0.11	0.08
ID_50	0.11	0.07	0.09	0.13	0.11	0.07	0.11	0.12
CC_50	0.09	0.09	0.12	0.15	0.11	0.10	0.11	0.09
UC_90	0.07	0.09	0.11	0.08	0.08	0.09	0.13	0.10

ID_90	0.12	0.12	0.14	0.09	0.12	0.12	0.14	0.12
CC_90	0.09	0.09	0.12	0.08	0.06	0.11	0.14	0.10

Table 4C: $\delta = 1;\ \phi = 0.5$

	News				Noise			
	AR	ARCH	GARCH	SV	AR	ARCH	GARCH	SV
logscore_t	0.58	0.50	0.59	0.69	0.95	0.96	0.96	0.95
CRPS_t	0.53	0.48	0.63	0.63	0.70	0.72	0.73	0.76
Quantl_t	0.52	0.48	0.59	0.65	0.87	0.85	0.86	0.97
	EOS							
PITs	0.84	0.75	0.79	0.92	1.00	1.00	1.00	1.00
UC_50	0.61	0.63	0.74	0.70	0.91	0.80	0.90	0.92
ID_50	0.10	0.15	0.18	0.11	0.07	0.27	0.28	0.12
CC_50	0.55	0.60	0.74	0.64	0.87	0.80	0.90	0.81
UC_90	0.75	0.64	0.69	0.87	1.00	1.00	1.00	0.99
ID_90	0.02	0.19	0.15	0.03	0.10	0.51	0.55	0.10
CC_90	0.62	0.64	0.68	0.73	0.99	0.99	1.00	0.99
	RTV							
PITs	0.10	0.11	0.14	0.09	0.13	0.15	0.17	0.11
UC_50	0.09	0.08	0.08	0.11	0.10	0.09	0.11	0.08
ID_50	0.11	0.09	0.09	0.13	0.10	0.09	0.12	0.08
CC_50	0.10	0.11	0.09	0.15	0.10	0.11	0.11	0.08
UC_90	0.09	0.09	0.13	0.08	0.09	0.10	0.14	0.09
ID_90	0.13	0.14	0.13	0.09	0.10	0.13	0.11	0.12
CC_90	0.09	0.09	0.13	0.08	0.07	0.11	0.11	0.10

Table 4D: $\delta = 1;\ \phi = 0.9$

	News				Noise			
	AR	ARCH	GARCH	SV	AR	ARCH	GARCH	SV
logscore_t	0.13	0.06	0.07	0.18	0.98	0.98	0.97	0.99
CRPS_t	0.13	0.07	0.08	0.18	0.73	0.74	0.70	0.76
Quantl_t	0.14	0.07	0.07	0.19	0.96	0.90	0.89	0.97
	EOS							
PITs	0.16	0.18	0.22	0.16	1.00	1.00	1.00	1.00
UC_50	0.11	0.10	0.13	0.10	0.98	0.94	0.98	0.99
ID_50	0.10	0.09	0.09	0.11	0.08	0.13	0.12	0.10
CC_50	0.12	0.10	0.13	0.11	0.97	0.92	0.96	0.97
UC_90	0.14	0.12	0.15	0.19	1.00	1.00	1.00	1.00
ID_90	0.10	0.12	0.14	0.06	0.10	0.32	0.28	0.11
CC_90	0.11	0.13	0.16	0.14	1.00	1.00	1.00	1.00
	RTV							
PITs	0.09	0.13	0.18	0.12	0.55	0.50	0.52	0.45
UC_50	0.09	0.08	0.08	0.06	0.11	0.10	0.10	0.11
ID_50	0.10	0.09	0.11	0.08	0.11	0.09	0.11	0.10
CC_50	0.09	0.09	0.11	0.06	0.12	0.10	0.10	0.11

UC_90	0.07	0.09	0.13	0.10	0.11	0.11	0.14	0.11
ID_90	0.14	0.12	0.12	0.14	0.14	0.14	0.10	0.10
CC_90	0.08	0.08	0.11	0.10	0.12	0.11	0.14	0.08

Note: The equal accuracy tests between EOS and RTV are based on logscore, CRPS and Quantile loss function. PITs is the Berkowitz test for calibration and independence of density forecast based on probability integral transforms (PIT), UC (coverage), ID (independence) and CC (conditional coverage) are tests for predictive intervals of coverage 50% and 90%. Results are based on 200 replications. Data generating process varies for each forecasting model (AR, ARCH(1), GARCH(1,1) and SV). Details on DGP parameter values in section 4.

Table 5: Characteristics of Data Revisions.

		Final is yt(t+14)		Data revisions size	
		H0: News	H0: Noise	To First	To Final
Quarterly Data Set (1985Q2-2015Q1)					
Y	real GDP	0.10	-6.49	0.60	0.38
C	real Cons.	2.48	-2.70	0.35	0.37
I	real invest.	4.24	-1.51	0.23	0.25
XM	net exports	-2.00	-2.96	0.04	0.04
G	real gov spending	4.49	-1.39	0.38	0.50
W	nominal wages	-0.10	-6.21	2.15	0.68
P	gdp deflator	3.89	-3.11	0.40	0.42
PCE	pce deflator	1.93	-0.79	0.11	0.11
Monthly Data Set (1985M1-2014M12)					
IP	Ind. Prod.	1.15	-8.97	0.45	0.35
PEMP	payroll emp.	0.97	-5.24	0.25	0.22
HS	housing starts	4.23	-1.39	0.18	0.20
S	Nominal sales	-0.10	-4.81	0.27	0.21
Ho	hours	7.58	1.37	0.34	0.60
EMP	civilian emp.	5.67	0.36	0.18	0.23

Notes: Entries in the first two columns are t-tests for the null hypothesis indicated. Yellow indicates data revisions are mainly news and blue that they are mainly noise. All series are in first differences, except the trade balance which is in levels. Real time data from the Philadelphia Fed real-time dataset, except nominal sales and civilian employment which are from ALFRED.

Table 6: Logscore values and improvements from adding ARCH and SV to the EOS approach

	AR	AR + ARCH	AR + SV	AR vs ARCH	AR vs SV
Quarterly series					
Y	0.780	4.8	6.7	-1.281	-1.292
C	0.681	1.4	9.5	-0.376	-1.925
I	2.256	6.7	10.1	-0.671	-0.780
XM	5.728	15.8	49.3	-0.852	-2.103
G	1.251	-4.6	0.9	1.418	-0.536
W	1.185	37.9	19.9	-3.887	-2.799
P	0.252	-8.2	2.5	1.502	-0.641
PCE	1.056	-34.1	42.7	1.199	-1.155
Monthly series					
IP	0.951	3.5	5.3	-0.517	-1.278
PEMP	-0.961	9.3	8.0	-3.761	-3.465
HS	3.538	4.5	12.3	-0.570	-1.845
S	1.453	20.9	13.4	-2.468	-2.382
Ho	0.397	0.2	0.2	-0.082	-0.432
EMP	0.197	4.6	4.6	-1.249	-1.715

Note: Out-of-sample forecasting origins: quarterly 2000Q2-2015Q1; monthly 2000M1-20014M12. Rolling windows of 60 (quarterly) or 180 (monthly) observations (sample starts in 1985). One-step ahead forecasts. Forecasts errors computed with first release data. The “AR” column has –logscore values and in the “AR+ARCH” and “AR+SV” columns the values are the (100*) difference in logscore, where positive values indicate forecast improvements. At 10% significance level, a one-side test rejects the null of equal accuracy in favour of the more accurate alternative if t-stat is <-1.282.

Table 7: EOS and RTV RMSFEs and t-statistic for test of equal accuracy.

	AR(1)			AR(1) + ARCH(4)			AR(1) + SV		
	EOS	RTV	H0: equal	EOS	RTV	H0: equal	EOS	RTV	H0: equal
Y	0.510	0.5	-0.289	0.539	3.0	-1.155	0.509	1.3	-0.664
C	0.457	0.3	-0.079	0.474	4.1	-1.427	0.457	-0.1	0.026
I	2.178	-2.9	0.540	2.076	-9.5	1.587	2.106	-6.2	1.089
XM	41.298	1.6	-2.148	41.586	2.9	-1.760	42.180	-18.2	0.945
G	0.832	-6.6	2.717	0.838	-7.4	3.150	0.833	-7.9	2.902
W	0.647	29.7	-1.832	0.566	19.4	-1.858	0.577	22.0	-1.702
P	0.272	2.5	-0.900	0.275	5.1	-1.629	0.273	1.8	-0.729
PCE	0.449	0.0	-0.001	0.460	3.7	-1.467	0.434	0.3	-0.296
	AR(3)			AR(3) + ARCH(4)			AR(3) + SV		
	EOS	RTV	H0: equal	EOS	RTV	H0: equal	EOS	RTV	H0: equal
IP	0.612	0.5	-0.376	0.626	3.0	-1.736	0.607	0.6	-0.422
PEMP	0.093	-1.0	0.435	0.092	-1.2	0.482	0.092	0.1	-0.046
HS	7.778	0.1	-0.201	7.843	0.5	-0.702	7.906	-0.1	0.154
S	1.003	1.1	-0.839	1.017	3.4	-1.314	1.042	3.6	-2.141
Ho	0.358	-18.0	2.762	0.351	-4.8	1.587	0.360	-7.6	2.080
EMP	0.280	-0.9	0.922	0.284	-0.2	0.143	0.284	-0.1	0.180

Notes: Out-of-sample forecasting origins: quarterly 2000Q2-2015Q1; monthly 2000M1-20014M12. Rolling windows of 60 (quarterly) or 180 (monthly) observations (sample starts in 1985). One-step ahead forecasts. Forecasts errors computed with first release data. The "EOS" column has RMSFEs and the "RTV" column has the proportional gain (reduction) in RMSFE in percentage such that positive values indicate RTV improvements.

Table 8: EOS and RTV Logscores and t-statistic for test of equal accuracy.

	AR(1)			AR(1) + ARCH(4)			AR(1) + SV		
	EOS	RTV	H0: equal	EOS	RTV	H0: equal	EOS	RTV	H0: equal
Y	0.780	-0.3	0.043	0.732	-4.0	0.553	0.713	-0.2	0.041
C	0.681	-1.8	0.726	0.667	-6.2	0.997	0.587	-3.7	1.307
I	2.256	-1.3	0.172	2.189	1.0	-0.089	2.155	-0.5	0.109
XM	5.728	40.3	-2.410	5.569	18.1	-1.124	5.234	-0.4	0.049
G	1.251	-9.4	2.921	1.297	-6.1	1.331	1.242	-6.5	2.661
W	1.185	52.7	-3.073	0.806	18.5	-1.413	0.986	37.8	-2.833
P	0.252	9.1	-1.616	0.334	-43.3	0.842	0.227	8.1	-1.371
PCE	1.056	9.8	-1.156	1.397	56.0	-1.099	0.630	4.7	-1.195
	AR(3)			AR(3) + ARCH(4)			AR(3) + SV		
	EOS	RTV	H0: equal	EOS	RTV	H0: equal	EOS	RTV	H0: equal
IP	0.951	-0.4	0.198	0.916	-1.4	0.331	0.898	2.3	-1.346
PEMP	-0.961	-1.6	0.565	-1.054	1.7	-0.668	-1.041	1.9	-0.942
HS	3.538	0.3	-0.334	3.493	-2.6	1.029	3.415	-1.6	1.551
S	1.453	-2.7	0.486	1.243	-13.9	0.875	1.319	-2.3	0.288
Ho	0.397	-18.7	4.527	0.395	-4.5	1.218	0.395	-6.8	1.947
EMP	0.197	4.1	-1.436	0.151	3.9	-1.774	0.151	2.4	-1.291

Notes: as Table 6. The "EOS" column has –logscore values and the "RTV" column has the (100*) difference in logscore, where positive values indicate forecast improvements.

Table 9: Excess Coverage and Conditional Coverage Test for 90% Predictive Intervals.

	AR(1)				AR(1) + ARCH(4)				AR(1) + SV			
	EOS	RTV	EOS_CC	RTV_CC	EOS	RTV	EOS_CC	RTV_CC	EOS	RTV	EOS_CC	RTV_CC
Y	0.05	0.02	0.10	0.10	0.03	-0.05	0.03	0.16	0.03	-0.03	0.32	0.46
C	0.03	0.05	0.32	0.10	0.00	0.02	0.87	0.63	0.02	0.05	0.63	0.10
I	0.03	0.00	0.03	0.00	0.02	-0.03	0.10	0.12	0.05	0.03	0.10	0.32
XM	-0.20	-0.10	0.00	0.07	-0.23	-0.10	0.00	0.06	-0.02	0.03	0.36	0.49
G	-0.02	-0.05	0.90	0.16	-0.03	-0.03	0.71	0.46	-0.02	-0.03	0.90	0.71
W	0.03	-0.02	0.03	0.07	0.05	-0.02	0.10	0.40	0.05	-0.03	0.10	0.46
P	-0.12	-0.05	0.02	0.35	-0.12	-0.08	0.01	0.15	-0.08	-0.02	0.15	0.90
PCE	-0.15	-0.08	0.00	0.01	-0.07	-0.10	0.27	0.06	-0.07	-0.07	0.12	0.12
	AR(3)				AR(3) + ARCH(4)				AR(3) + SV			
	EOS	RTV	EOS_CC	RTV_CC	EOS	RTV	EOS_CC	RTV_CC	EOS	RTV	EOS_CC	RTV_CC
IP	-0.01	-0.01	0.01	0.03	-0.06	-0.07	0.02	0.00	0.02	-0.02	0.72	0.71
PEMP	0.00	0.04	0.07	0.00	0.02	0.03	0.59	0.28	0.01	0.03	0.02	0.00
HS	-0.11	-0.09	0.00	0.00	-0.06	-0.06	0.07	0.04	-0.02	-0.03	0.62	0.39
S	0.04	0.02	0.00	0.00	0.02	-0.02	0.56	0.40	0.02	0.01	0.00	0.00
Ho	-0.01	0.04	0.87	0.03	-0.01	0.03	0.66	0.12	-0.01	0.03	0.53	0.28
EMP	-0.02	0.01	0.06	0.47	-0.03	-0.01	0.18	0.97	-0.02	0.01	0.40	0.47

Note: The first two columns show excess coverage. The following two columns are p-values of the of the conditional coverage test. Entries in blue indicate that the null hypothesis of predictive intervals with correct coverage is not rejected at 10% level. Sample period as in Table 6.

Table 10: Berkowitz test for calibration and independence of PITs; entries are p-values.

	AR(1)		AR(1) + ARCH(4)		AR(1) + SV	
	EOS	RTV	EOS	RTV	EOS	RTV
Y	0.12	0.16	0.05	0.02	0.04	0.24
C	0.32	0.13	0.06	0.06	0.30	0.08
I	0.07	0.21	0.03	0.01	0.21	0.01
XM	0.00	0.00	0.00	0.01	0.01	0.72
G	0.30	0.06	0.31	0.06	0.49	0.22
W	0.00	0.22	0.00	0.15	0.00	0.27
P	0.00	0.02	0.00	0.00	0.01	0.12
PCE	0.00	0.00	0.00	0.00	0.00	0.01
	AR(3)		AR(3) + ARCH(4)		AR(3) + SV	
	EOS	RTV	EOS	RTV	EOS	RTV
IP	0.16	0.01	0.08	0.01	0.54	0.41
PEMP	0.00	0.00	0.01	0.00	0.00	0.00
HS	0.00	0.00	0.00	0.00	0.06	0.05
S	0.02	0.00	0.01	0.00	0.01	0.03
Ho	0.08	0.00	0.03	0.03	0.09	0.02
EMP	0.01	0.20	0.03	0.06	0.07	0.10

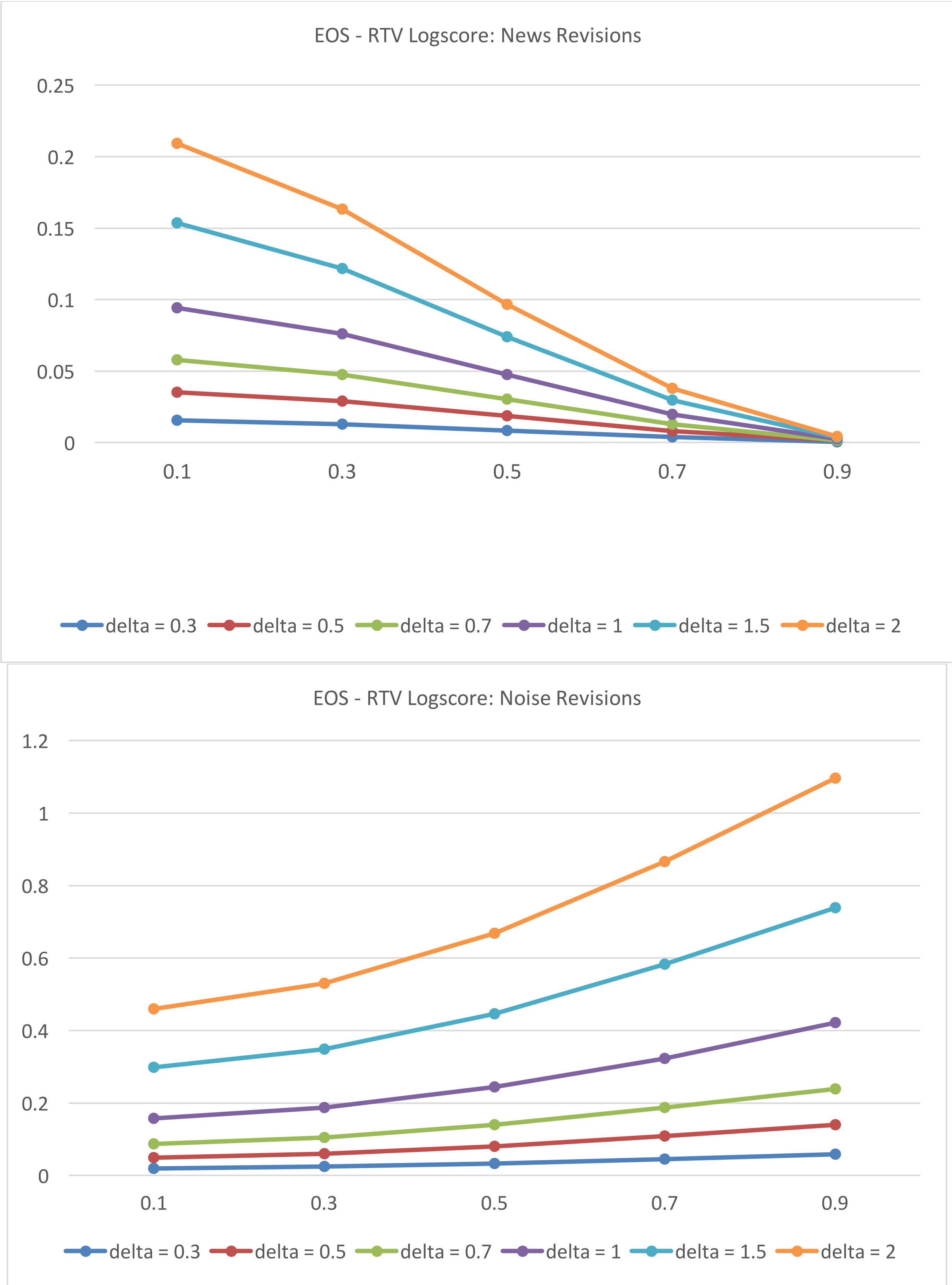
Note: Entries in blue indicate that null hypothesis of correct calibration and independence of PITs is not rejected at 10% level. Sample period as in Table 6.

Table 11: Forecasting with BVAR-SV

	Logscore			90% Coverage				PITs (Berkowitz)	
	EOS	RTV	t-stat	EOS	RTV	EOS_CC	RTV_CC	EOS	RTV
Y	0.690	-3.3	0.410	0.93	0.90	0.49	0.87	0.054	0.047
P	0.295	11.2	-0.731	0.83	0.88	0.26	0.86	0.585	0.753
C	0.540	-7.6	1.677	0.93	0.97	0.49	0.01	0.091	0.015
R (fed rate)	0.290	-7.2	1.466	0.93	0.92	0.03	0.10	0.007	0.002

Note: p=4; overall prior tightness = 0.2; See notes of tables 8, 9 and 10. Sample period as in table 6.

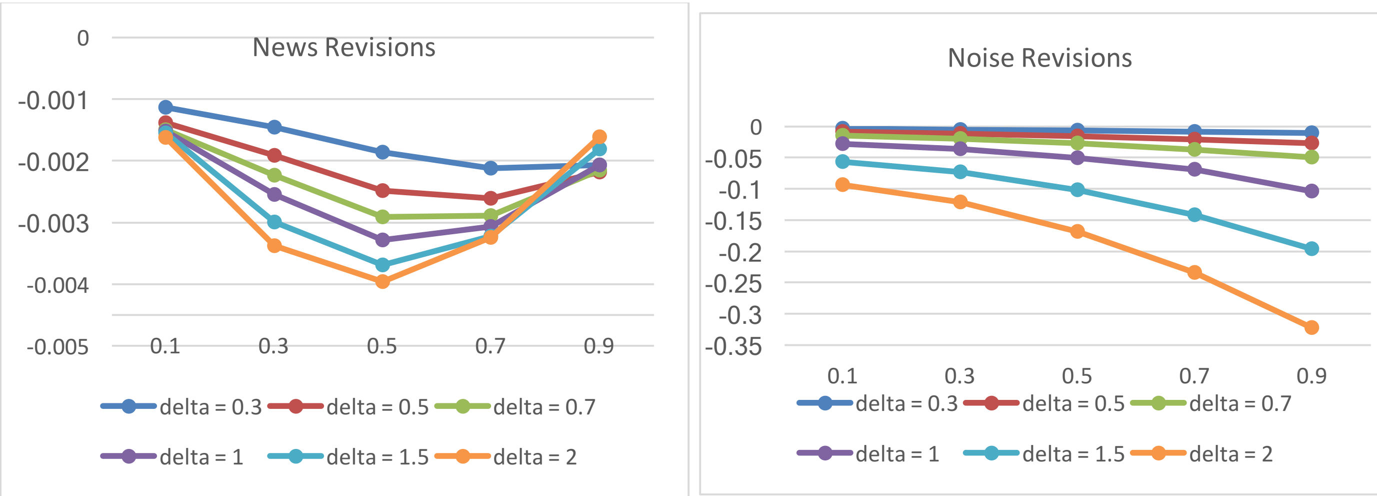
Figure 1: Analytical Results for $\Delta score^{News}$ and $\Delta score^{Noise}$



Note: Values for the AR(1) parameter ϕ on the horizontal axis. Values for $\phi=0.9$ and news revisions are small, near zero, but always positive.

Figure 2: Results of a Simulation Exercise with T=1000.

Figure 2A: EOS – RTV Logscore: Difference between ARCH values and AR.



Note: ARCH(0.1) data generating process. Values for the AR(1) parameter ϕ on the horizontal axis.

Figure 2B: EOS-RTV logscore: values for δ= 1 for different specifications.

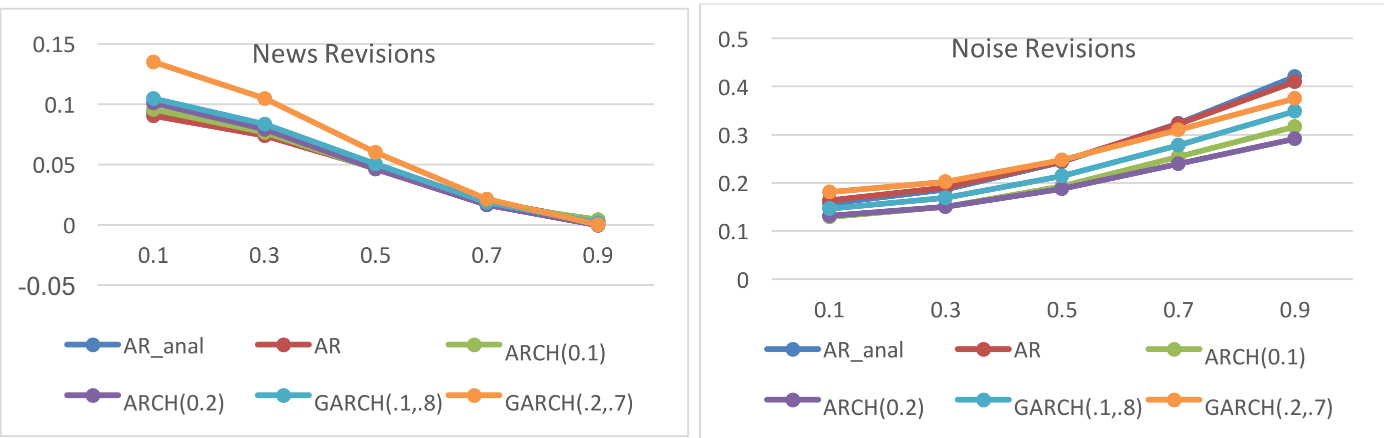


Figure 2C: EOS Excess Coverage for 90% nominal intervals with δ=

ϕ	AR	ARCH(0.1)	ARCH(0.2)	GARCH(.1,.8)	GARCH(.2,.7)
News					
0.1	0.08	0.08	0.08	0.08	0.08
0.3	0.07	0.07	0.07	0.07	0.07
0.5	0.06	0.06	0.06	0.06	0.06
0.7	0.04	0.04	0.04	0.04	0.04
0.9	0.02	0.02	0.02	0.02	0.02
Noise					
0.1	-0.14	-0.13	-0.12	-0.13	-0.13
0.3	-0.15	-0.14	-0.13	-0.14	-0.14
0.5	-0.17	-0.15	-0.14	-0.16	-0.16
0.7	-0.20	-0.18	-0.17	-0.18	-0.18
0.9	-0.23	-0.21	-0.20	-0.21	-0.21

Note: All based on 5000 replications. ARCH(0.1) is ARCH of order 1 with w1=0.1, while ARCH(0.2) has w2=0.1. GARCH (0.1,0.8) is a GARCH(1,1) with w1=0.1 and w2=0.8, while GARCH (0.2,0.7) is the same model but with w1=0.2 and w2=0.7. All forecasting models (EOS and RTV) are estimated with an AR(1) for the conditional mean.

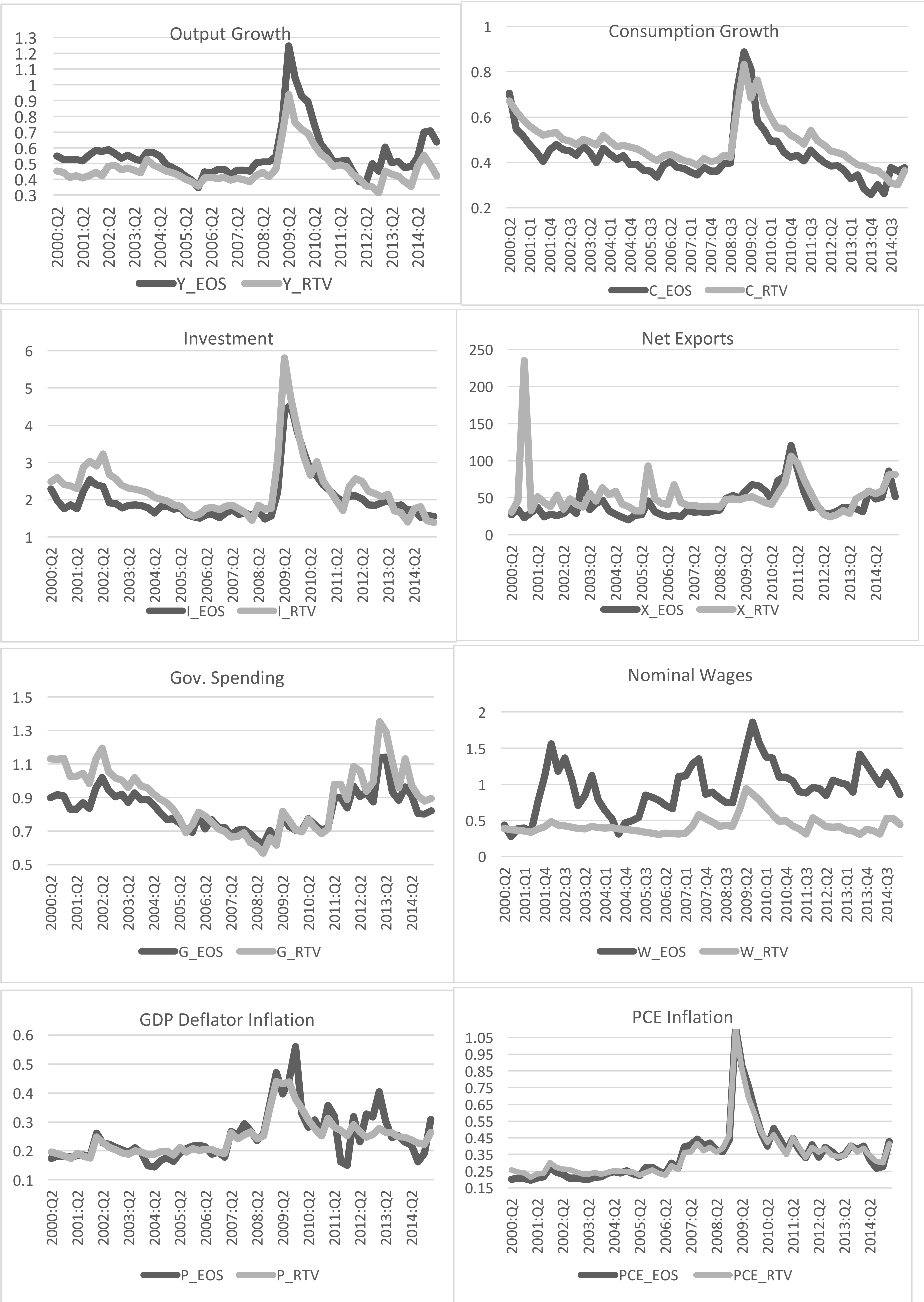
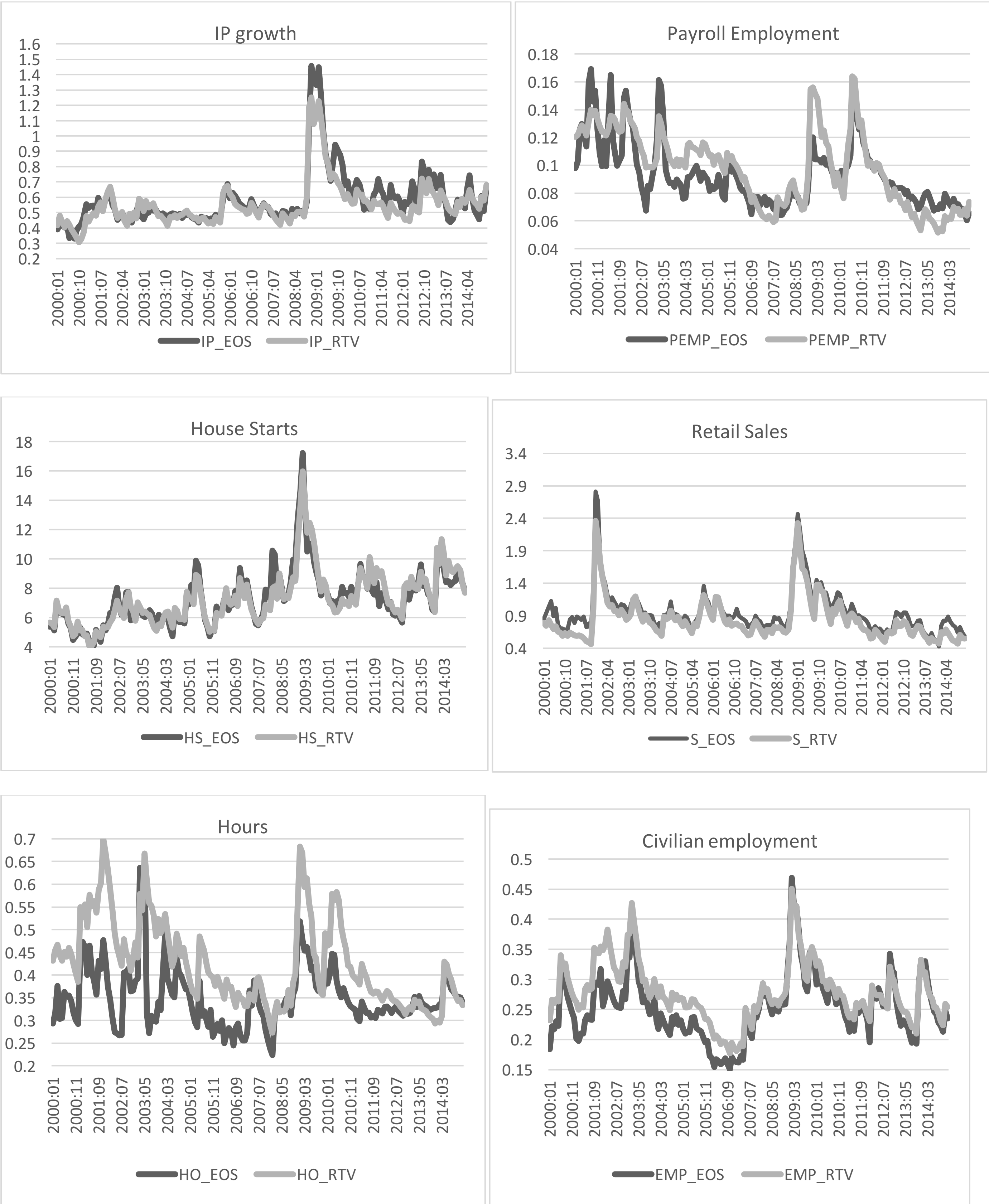


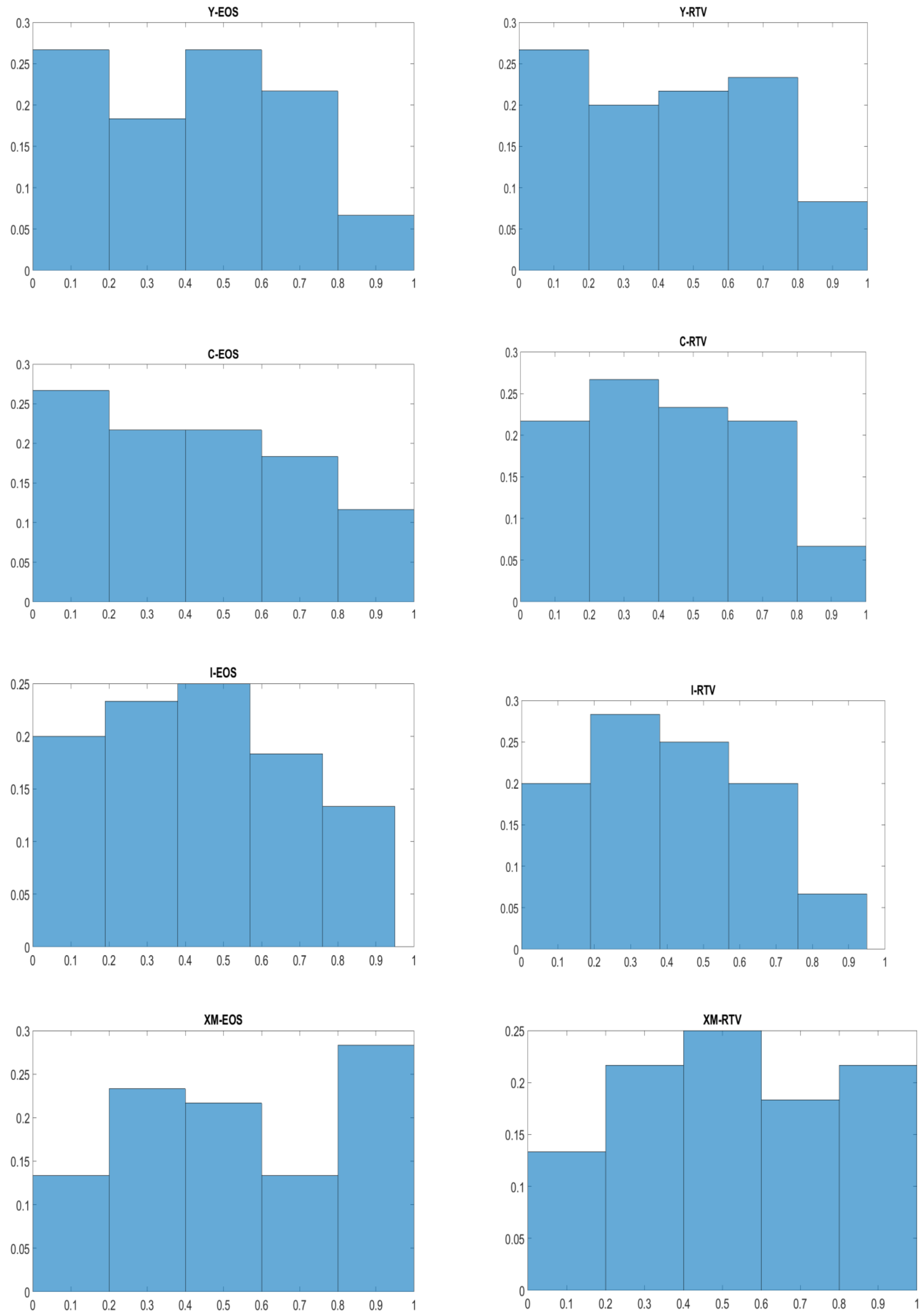
Figure 3: EOS and RTV one-step-ahead forecasting uncertainty computed with AR(1)+SV model for quarterly series.

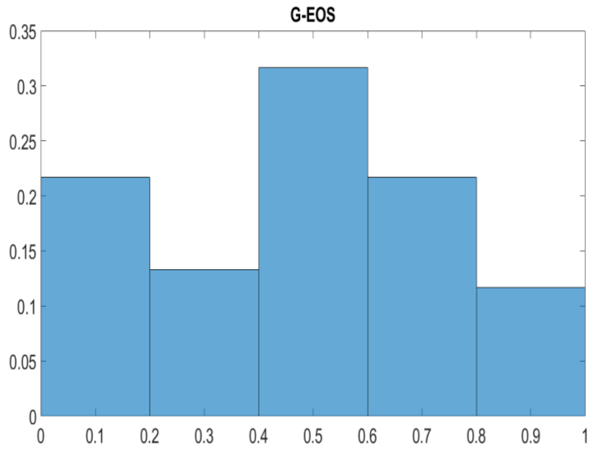
Figure 4: EOS and RTV one-step-ahead forecasting uncertainty computed with AR(1)+SV model for monthly series.

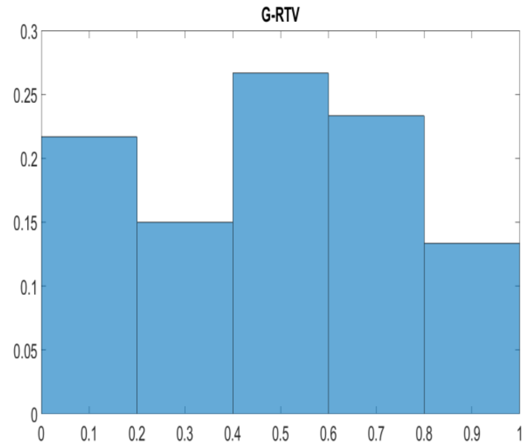


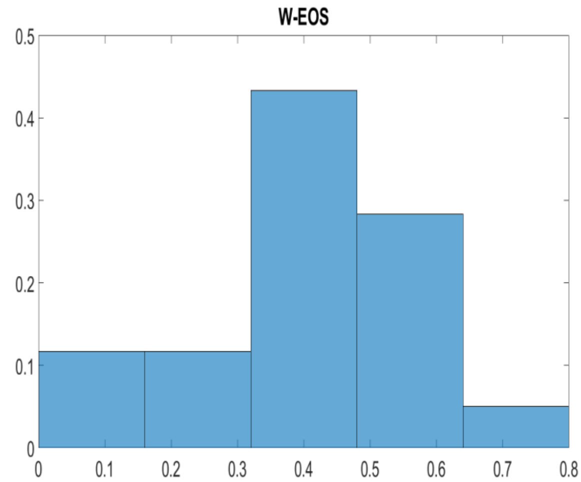
Note: These predicted volatility values take into account parameter uncertainty (variance of the 12000 predictive density draws)

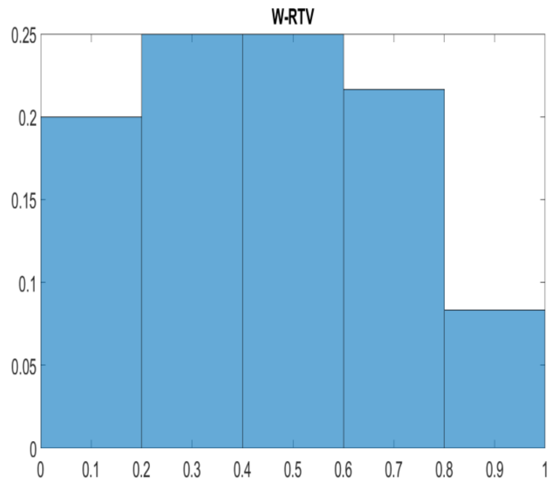
Figure 5: EOS and RTV PITs for the AR(1)+SV model for quarterly series (P=60)

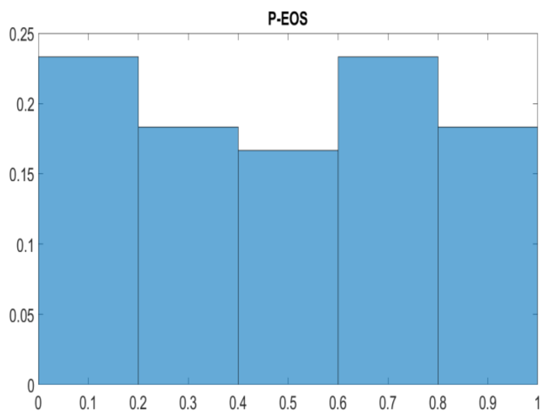


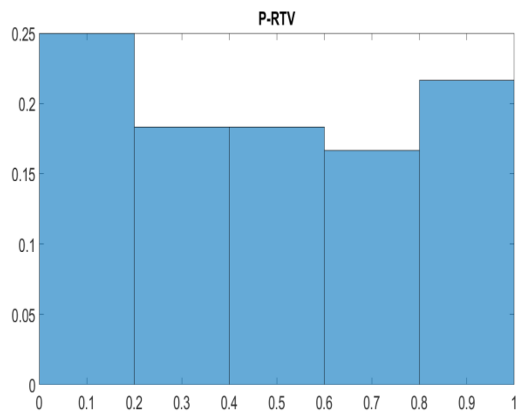

G-EOS
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

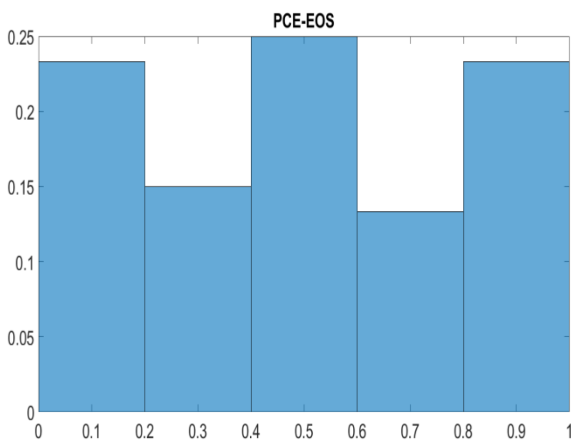

G-RTV
0 0.05 0.1 0.15 0.2 0.25 0.3
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

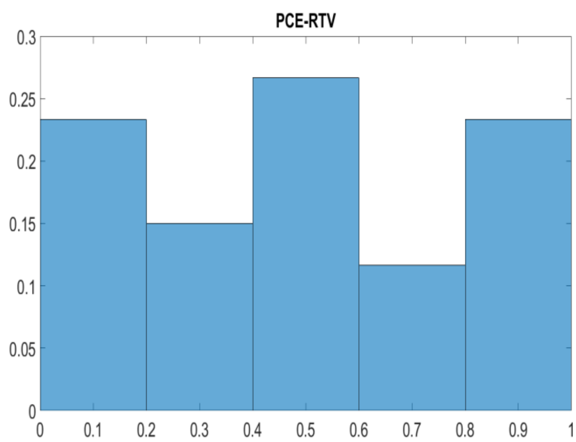

W-EOS
0 0.1 0.2 0.3 0.4 0.5
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8


W-RTV
0 0.05 0.1 0.15 0.2 0.25
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


P-EOS
0 0.05 0.1 0.15 0.2 0.25
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


P-RTV
0 0.05 0.1 0.15 0.2 0.25
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


PCE-EOS
0 0.05 0.1 0.15 0.2 0.25
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


PCE-RTV
0 0.05 0.1 0.15 0.2 0.25 0.3
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1