Studies in Nonlinear Dynamics & Econometrics

Volume 16, *Issue* 1 2012 *Article* 2

Forecasting U.S. Output Growth with Non-Linear Models in the Presence of Data Uncertainty

Michael P. Clements*

*Warwick University, m.p.clements@warwick.ac.uk

DOI: 10.1515/1558-3708.1865

Forecasting U.S. Output Growth with Non-Linear Models in the Presence of Data Uncertainty*

Michael P. Clements

Abstract

We consider the impact of data revisions on the forecast performance of a SETAR regime-switching model of U.S. output growth. The impact of data uncertainty in real-time forecasting will affect a model's forecast performance via the effect on the model parameter estimates as well as via the forecast being conditioned on data measured with error. We find that benchmark revisions do affect the performance of the non-linear model of the growth rate, and that the performance relative to a linear comparator deteriorates in real-time compared to a pseudo out-of-sample forecasting exercise.

*I am grateful to Dick van Dijk for helpful comments on an earlier draft and to two anonymous referees for many insightful suggestions. Computations were performed using code written in the Gauss Programming Language.

1 Introduction

There has been much interest in the recent literature on the importance of allowing for the effects of data uncertainty. A number of related strands of this literature can be identified. Firstly, do key macroeconomic results or relationships established for one particular vintage of data remain relevant for other vintages of data? Croushore and Stark (2003) examine three major studies in macroeconomics, and find that of these three the results of the seminal paper by Hall (1978) on the rational expectations permanent income hypothesis appear to be dependent on the particular data vintage studied. Secondly, do data revisions affect the conduct of monetary policy and the evaluation of the efficacy of monetary policy, given that data revisions will typically affect the calculation of output gaps?[1] Thirdly, are data vintage effects important in forecasting? It has been suggested that the use of final-revised data may give a misleading impression of the usefulness of various predictor variables relative to appraisals based solely on the data vintages available at the time the forecasts are constructed (see, for example, Diebold and Rudebusch (1991), Faust, Rogers, and Wright (2003), and the recent review by Croushore (2006)). There has also been interest in how best to specify forecasting models when there are various data vintage estimates of the same observation.[2]

Our paper contributes to the literature on data uncertainty and forecasting, addressing the specific issue of whether data uncertainty disproportionately affects the forecast performance of the popular non-linear class of threshold autoregressive regime-switching models relative to their linear comparators. Our focus is on modelling and forecasting post WWII US output growth using self-exciting threshold autoregressive (SETAR) models, as this application has received a great deal of attention in the literature. Comparisons of the relative forecast performance of linear and non-linear models in this context often suggest that any advantages in favour of non-linear models are typically modest, and that whether or not gains are realized at any point in time may depend on the regime in operation at the forecast origin. However, forecast comparisons of linear and non-linear models have typically been based on pseudo out-of-sample forecasting exercises, rather than real-time forecasting schemes. Real-time exercises specify and estimate the model, and calculate the forecast, using only observations available at the time the forecast is made. There are two dimensions to this: the first relates to the time period to which the observations refer, whereas the second is less obvious but has been the focus of much of the

[1] See, e.g., Runkle (1998), Orphanides (2001) and Orphanides and van Norden (2005).

[2] See, e.g., Howrey (1978, 1984), Sargent (1989), Koenig, Dolmas, and Piger (2003), Harrison, Kapetanios, and Yates (2005), Kishor and Koenig (2010), Jacobs and van Norden (2007), Cunningham, Eklund, Jeffery, Kapetanios, and Labhard (2009), and Clements and Galvão (2008, 2009, 2010).

Published by De Gruyter, 2012

recent literature, and is that the observations are only drawn from the vintages of data that would have been available at that time. A pseudo exercise uses only time-period observations which are in the relative past compared to the forecast origin, but draws these from a vintage that would not have been available at the forecast origin. In the pseudo exercise the models are estimated, and the forecasts are conditioned on, data that have been heavily revised. In a real-time forecasting exercise, the model will be estimated on a mixture of heavily-revised and lightly revised data, and the forecasts are conditioned on early estimates (of the recent time periods up to the forecast origin). Of interest is whether the relatively modest gains to the SETAR model observed in pseudo-forecasting exercises would also have been realized in real-time forecasting exercises when the data uncertainty at the forecast origin is taken into account.

The plan of the remainder of the paper is as follows. Section 2 briefly reviews the SETAR model and the evidence concerning the value of this model for forecasting US output growth compared to simple (linear) autoregressions. Section 3 then describes the empirical forecasting exercises, and how these are devised to estimate the impact of certain factors of interest, such as how data uncertainty affects parameter estimation uncertainty, and the impact of uncertainty about the values that the forecasts are conditioned on. Section 4 reports the results of the empirical forecast comparisons. We consider the constancy of the SETAR model over different data vintages, and the extent to which the relative forecast accuracy of the SETAR depends on the data vintage in pseudo out-of-sample comparisons. We then evaluate the forecast accuracy in a real-time setting. Section 5 presents a Monte Carlo study to supplement the empirical findings concerning the impact of data revisions on the forecast performance of the SETAR model. Section 6 offers some concluding remarks.

2 SETAR models and US output growth

The threshold autoregressive (TAR) model was first proposed by Tong (1978), Tong and Lim (1980) and Tong (1983) (see also Tong (1995)), and marks a simple departure from the linear time-series models popularized by Box and Jenkins (1970). Let y_t denote the variable of interest (in our case, the quarterly percentage growth rate of output). y_t is assumed to be determined by one of a small number of linear autoregressions. Which autoregression is in force depends upon the value of some past lag of the process relative to a threshold (or set of thresholds), or alternatively it may depend on the value of an extraneous variable. In contrast to the Hamilton (1989) model, where the regime-switching process is an unobservable discrete first-order Markov process, the regimes are observable. When the threshold variable is a lag of y_t, say, y_{t-d}, so that d is the length of the delay, then the model is 'self-exciting', giving rise to the acronym SETAR. When there are two regimes, then the process is in regime $i = 1$ at period t when $y_{t-d} \leq r$, and otherwise ($y_{t-d} > r$) in regime $i = 2$:

$$y_t = \phi_0^{\{i\}} + \phi_1^{\{i\}} y_{t-1} + \ldots + \phi_p^{\{i\}} y_{t-p} + \varepsilon_t^{\{i\}}, \quad \varepsilon_t^{\{i\}} \sim iid\left(0, \sigma^{2\{i\}}\right), \quad i = 1,2 \quad (1)$$

where the super-scripts $\{i\}$ indicate parameters that may vary across regime. As written, the model allows the variance of the disturbances to depend upon the regime. Stationarity and ergodicity conditions are discussed in Tong (1995) and the references therein. The above model can be written as a SETAR$(2; p, p)$ to denote two regimes and p autoregressive lags in each of the two regimes.

SETAR models have been used to model biological and physical processes, e.g., the Canadian lynx data and Wolf's sunspot numbers are classic examples (see, for example Tong (1995), chapter 7). They have also been applied to the modelling of economic and financial variables, with a prime application being US output growth, see, *inter alia*, Tiao and Tsay (1994), Potter (1995), Clements and Smith (1999, 1997), Clements and Krolzig (1998, 2003).

In terms of forecasting US output growth, any forecasting gains to the SETAR model have generally been found to be small.[3] Our interest is in whether the use of real-time data might further detract from any advantage that the SETAR model has in forecasting.

3 Design of forecasting exercise

We focus the comparisons on the impact of data revisions (including benchmark revisions) on the relative forecasting performance on non-linear regime-switching models compared to linear autoregressive models. It will be useful to briefly describe the nature of the revisions in the data to which we have access. The first estimate of output (GDP after 1991, GNP before 1991) in the Real-Time Data Set for Macroeconomists (RTDSM, www.philadelphiafed.org, see Croushore and

[3] For example, Tiao and Tsay (1994) find a maximum gain to the SETAR of only 6% (relative to an AR(2)), and that at 3-steps ahead. However, dividing up the forecast errors into two groups depending upon the regime at the forecast origin, and then assessing forecast accuracy for each regime separately, the SETAR records gains of up to 15% in the first regime. Because a clear majority of the data points (around three quarters) fall in the second, expansionary regime, the linear model will largely be determined by these points and will match the second-regime of the SETAR model. Thus the forecast performance of the two models is broadly similar for data points in the second regime. However, data points in the first regime of the SETAR model are characterised by different dynamics, so it is here that the SETAR model can gain relative to the linear model.

Stark (2001)) contains the Bureau of Economic Analysis (BEA) 'advance' estimate, which we denote y_t^{t+1}, and the second estimate y_t^{t+2} is the BEA 'final' estimate. The superscript denotes the quarter when the vintage estimate becomes available. The data of the BEA are then subject to three annual revisions which occur in the July of each year, as described by, e.g., Fixler and Grimm (2005, 2008) and Landefeld, Seskin, and Fraumeni (2008). This means that y_t^{t+14} - the estimate some three and a half years after the observations period - will have been subject to the three annual rounds of revision irrespective of the quarter of the year to which period t belongs. In addition to the regular rounds of revisions, every five or ten years the BEA data are typically subject to 'benchmark revisions', reflecting methodological changes in measurement or collection procedures (including base year changes).[4]

Because the revisions process comprises a relatively small number of revisions, for estimation samples consisting of a reasonable number of observations all but a small proportion at the end of the sample period will be 'fully revised'. Hence one might expect that the estimated parameters of a model will not be much affected by data revisions. Although regular revisions might only have a minor impact, benchmark revisions entail that the whole of the data series is revised. Nevertheless, benchmark revisions might be thought to have little effect on growth rates, and therefore on the models we consider in this paper. However, the use of a fixed-weighting method prior to the introduction of chain-weighting in 1996 means that the growth rates of real variables that are distant to the base year will be affected. Consequently, it remains an empirical issue whether benchmark revisions, and other revisions more generally, affect the relative forecast performance of our linear and non-linear models of US output growth. The other channel through which data uncertainty might affect relative forecast performance is via the variable values that the forecasts are conditioned on.

We use recursive forecasting exercises, and we now explain precisely how these are run. Recall that y_t^{t+i} denotes the percentage growth rate between quarters t and $t-1$ as recorded in the data vintage of period $t+i$, where $i = 1, 2, \ldots$, so that data are only available with a lag. We focus on 1-step ahead forecasts. Hence, at time $t+1$ we wish to forecast the period $t+1$ value of output growth using the $t+1$ data vintage. Therefore the available data comprises $\left\{y_{t-i}^{t+1}\right\}_{i=0,1,2,\ldots}$. Of course earlier vintage estimates will be available for some observations, specifically,

[4] For example, Siklos (2008) identifies eight benchmark revisions in 1966, 1971, 1976, 1981, 1986, 1992, 1996 and 2001, all occurring in the data vintage of the first quarter of the year. So, for example, the 1981:1 vintage has data up to 1980:4, which is calculated on a different basis or definition to the data through 1980:3 in the 1980:4 vintage. The way which the national accounts data are calculated then remains unchanged until the 1986:1 vintage. Base year changes occurred in 1976, 1985 and 1991.

$\left\{y_{t-i}^{t-j}\right\}_{j=0,1,2,...;i>j}$, but we assume the latest estimates of each time period are used at each forecast origin. This is known as the use of end-of-sample (EOS) data by Koenig et al. (2003) and Clements and Galvão (2008, 2010), and is the standard way of using real-time data.

Following standard notational practice (e.g., West (2006)), we suppose that there are a total of $T = R + P$ time periods, and that the initial forecast origin is $t = R$, with a final forecast origin of $t = R + P - 1$, so that R is the initial estimation size, which increases to $R + P - 1$ as the forecast origin moves through the sample, generating in total P one-step ahead forecasts.

Our interest is in the MSFEs - the average of the squared forecast errors. Subsequent discussion will be facilitated by the following notation that makes explicit the channels by which data uncertainty may affect forecast accuracy. For an AR(p) model, and the pseudo out-of-sample forecasting exercise, we write the regression we estimate as:

$$y_{s+1}^{f} = \mathbf{x}_s^{f\prime}\beta^{f} + u_{s+1}^{f} \tag{2}$$

where $s = 1, 2, \ldots t-1$, for a forecast origin $t \in [R, \ldots, R+P-1]$, where $\mathbf{x}_s^{f\prime} = \left[1, y_s^f, y_{s-1}^f, \ldots, y_{s-p+1}^f\right]$ for an AR(p), and where 1 denotes the intercept. The f-superscript denotes a 'final vintage' value, where $f > T$. All the P forecasts are generated from models estimated on the same vintage of data. The superscript on β allows that the model may depend on the 'final-vintage' of data that is used. Parameter instability is found to be a feature of many macroeconomic relationships (see, e.g., Stock and Watson (1996)) and its implications for forecasting have been discussed extensively by Clements and Hendry (1999, 2006), *inter alia*: here we focus on data vintage effects and ignore possible instability over time (for a given vintage). We design our empirical investigation to minimise the potential impact of instability on the forecast comparisons. Hence, we compare forecast accuracy for different final vintages keeping the time periods covered by the estimation and forecasting periods fixed.

The forecasts are based on the estimated version of (2), and are given by:

$$\hat{y}_{t+1}^{f} = \mathbf{x}_t^{f\prime}\hat{\beta}_t^{f} \tag{3}$$

The scripts on the estimated model parameter indicate the data vintage (super) and the end of the estimation period (sub), not that the estimator allows for a time-varying parameter vector, as estimation is by OLS. The resulting MSFE is calculated as:

$$MSFE_{AR,f} = P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^{f} - \mathbf{x}_t^{f\prime}\hat{\beta}_t^{f}\right)^2.$$

The real-time EOS model estimates are calculated by OLS applied to:

$$y_{s+1}^{t+1} = \mathbf{x}_s^{t+1\prime}\beta^{t+1} + u_{s+1}^{t+1}$$

for $s = 1, 2, \ldots t-1$, for a forecast origin t, where as above $t \in [R, \ldots, R+P-1]$. Here, $\mathbf{x}_s^{t+1\prime} = \left[1, y_s^{t+1}, y_{s-1}^{t+1}, \ldots, y_{s-p+1}^{t+1}\right]$. The forecasts are given by:

$$\hat{y}_{t+1}^{EOS} = \mathbf{x}_t^{t+1\prime}\hat{\beta}_t^{t+1}, \tag{4}$$

and these are again evaluated using final-vintage actuals, so that:

$$MSFE_{AR,EOS} = P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^{f} - \mathbf{x}_t^{t+1\prime}\hat{\beta}_t^{t+1}\right)^2$$

A variant that isolates the effect on the forecasts of the impact of data uncertainty on the parameter estimates is:

$$MSFE_{AR,EOS,x^f} = P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^{f} - \mathbf{x}_t^{f\prime}\hat{\beta}_t^{t+1}\right)^2. \tag{5}$$

The only difference between $MSFE_{AR,f}$ and $MSFE_{AR,EOS,x^f}$ is that the latter estimates the model on the latest data vintage available at each forecast origin. Both measures condition the forecast on the final-vintage values of the observations. Finally, we may use real-time data to condition the forecasts on, but final-vintage model estimates:

$$MSFE_{AR,f,x^{t+1}} = P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^{f} - \mathbf{x}_t^{t+1\prime}\hat{\beta}_t^{f}\right)^2. \tag{6}$$

to help isolate the way in which the use of real-time data affects the forecasts.

So, the forecasts (3) are obtained using fully-revised data. These forecasts can be contrasted with those based on real-time forecasting (4), which use the latest vintage of data available at the time the forecast is made - both to estimate the model, and to condition the forecasts on. The two hybrid cases (defined implicitly in (5) and (6)) should help to isolate the impacts of parameter estimation and of the data the forecasts are conditioned on. Comparing (5) and (3) will highlight the impact of parameter estimation uncertainty by conditioning the forecasts on the final data in each case, and similarly a comparison of (6) and (3) will bring out the impact of data uncertainty at the forecast origin (i.e., that the data the forecast is conditioned on will subsequently be revised).

For the 2-regime SETAR model we calculate the same four measures. If we adopt the shorthand notation that

$$\begin{aligned} g(\mathbf{x}_{t-1};\theta) &= 1(y_{t-d} \le r)\left[\phi_0^{\{1\}} + \phi_1^{\{1\}} y_{t-1} + \ldots + \phi_p^{\{1\}} y_{t-p} + \varepsilon_t^{\{1\}}\right] \\ &\quad +1(y_{t-d} > r)\left[\phi_0^{\{2\}} + \phi_1^{\{2\}} y_{t-1} + \ldots + \phi_p^{\{2\}} y_{t-p} + \varepsilon_t^{\{2\}}\right] \end{aligned}$$

where $\theta = \left(\phi_0^{\{1\}}, \ldots \phi_p^{\{1\}}; \phi_0^{\{2\}}, \ldots \phi_p^{\{2\}}; r, d\right)$, then the corresponding accuracy measures are given by:

$$\begin{aligned} MSFE_{SETAR,f} &= P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^f - g\left(\mathbf{x}_t^f; \hat{\theta}_t^f\right)\right)^2 \\ MSFE_{SETAR,EOS} &= P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^f - g\left(\mathbf{x}_t^{t+1}; \hat{\theta}_t^{t+1}\right)\right)^2 \\ MSFE_{SETAR,EOS,x^f} &= P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^f - g\left(\mathbf{x}_t^f; \hat{\theta}_t^{t+1}\right)\right)^2 \\ MSFE_{SETAR,f,x^{t+1}} &= P^{-1}\sum_{t=R}^{T-1}\left(y_{t+1}^f - g\left(\mathbf{x}_t^{t+1}; \hat{\theta}_t^f\right)\right)^2. \end{aligned}$$

In the penultimate case, the model is estimated on vintage $t+1$ observations up to t, but the forecast is conditioned on $\mathbf{x}_t^f$, both in terms of regime determination at the origin, and in terms of evaluating the resulting autoregression. The last measure conditions the forecast (including the regime determination) on the real-time estimates of the data at the forecast origin, but uses estimates of the model's parameters from the final data.

4 Forecast comparison results

We use data on the quarterly vintages of real output (GDP after 1991, GNP before 1991) from the RTDSM. The RTDSM used in our study contains data vintages up to that of 2008:4. We focus on the two-regime SETAR model of US output growth of Potter (1995). The model was originally estimated over the period 1947:2 to 1990:4 (less observations lost from calculating lagged values). The model was a fifth-order autoregression in both regimes, with the third and fourth lags set to zero, a threshold of zero, and a delay lag of two periods. Following Tiao and Tsay (1994) we omit the lag five terms. The resulting model is essentially that of Potter (1995), in that (estimated on 1947:2 to 1990:4 from the 1991:1 data vintage) we find a threshold value that is approximately zero ($r = -0.008$), the delay is two ($d = 2$), and the 'recession' regime when $y_{t-2} \leq 0$ is characterised by a large negative autoregressive coefficient on the second lag, such that this regime tends to be short-lived relative to the second regime ($y_{t-2} > 0$) which more closely resembles the (single-regime) linear autoregressive model of the data. Specifically, the estimates of the lower and upper regime AR parameters (and estimated standard errors) are: $\rho_0^1 = -0.450$, $\rho_1^1 = 0.386$, $\rho_2^1 = -0.849$, $\sigma_1 = 1.252$; and $\rho_0^2 = 0.420$, $\rho_1^2 = 0.316$, $\rho_2^2 = 0.178$, $\sigma_2 = 0.888$.

4.1 Whole sample estimates and data vintage

In order to see whether these features were specific to the 1991:1 data vintage, we re-estimated the same specification for the sample period 1959:4 to 1990:4 on all the data vintages from 1991:1 to 2008:4, inclusive. The specification was kept as a second-order autoregression in both regimes, but the threshold value, the delay parameter and the regime-specific coefficients were estimated on each vintage of data using standard techniques.[5] The reason for truncating the beginning of the estimation sample, relative to Potter (1995), was to obtain a period for which data were available for the 72 vintages we consider. Table 1 records the SETAR(2;2,2) model estimates, and shows some marked changes in the estimated SETAR models across data vintages.[6] The changes across vintages generally coincide with benchmark changes: the estimates are little changed on all vintages up to (and including) 1991:4, were similar on vintages 1992:1 to 1995:4; on vintages 1996:1 to 1999:3; 1999:4 to 2003:4; and 2004:1 onwards. This suggests that benchmark revisions do affect the estimated SETAR model.

To provide further evidence of data vintage effects, we moved the end of the estimation period forward by a decade, so that the estimation period becomes 1957:2 to 2000:4, and then estimated the model on all vintages of data from 2001:1 to 2008:4. For this period, the model showed practically no variation over data vintage: see table 2. But the model is markedly different from that of Potter (1995), with the threshold estimated at around a growth rate of nearly half a percentage point, such that the two regimes now correspond to below and above average growth (rather than contraction and expansion). There is also clear evidence of non-constancy of the SETAR model over time for a given data vintage, as noted

[5] See e.g., Franses and van Dijk (2000).

[6] Holding the vintage constant at the 1991:1 data release, and setting the start date to 1959:4 resulted in changes relative to using the 1947:2 - 1990:4 sample for estimation. Compare the model estimates reported in the text above to those in the first line of table 1. This is an example of non-constancy over time, whereas table 1 (and table 2) highlight the non-constancy of the model estimates over data vintage.

by Clements and Smith (1997). This is apparent from a comparison of the model estimates in tables 1 and 2 for the same data vintage.

Hence there is non-constancy in the SETAR model estimates across data vintages for the sample period 1959–1990, mainly in response to benchmark revisions, but the model estimates change little across vintages 2000:1 to 2008:4 for an estimation sample period of 1957–2000.

Table 1: Model constancy over vintage. Estimation sample 1959:4 to 1990:4.

	AR(2)				SETAR(2;2,2), Regime 1				SETAR(2;2,2), Regime 2					
Vintage	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	r	d
1991.1	.43	.23	.18	.86	-0.29	.14	-0.50	1.74	0.61	.27	0.05	.62	-0.01	2
1991.2	.43	.23	.18	.85	-0.29	.14	-0.50	1.74	0.61	.27	0.05	.61	-0.01	2
1991.3	.43	.23	.18	.85	-0.29	.14	-0.50	1.74	0.61	.27	0.05	.61	-0.01	2
1991.4	.43	.23	.18	.85	-0.29	.14	-0.50	1.74	0.61	.27	0.05	.61	-0.01	2
1992.1	.44	.24	.15	.82	0.25	.25	-0.19	1.21	0.89	.14	-0.02	.49	0.57	2
1992.2	.43	.25	.15	.81	0.26	.27	-0.16	1.18	0.88	.13	0.00	.49	0.61	2
1992.3	.43	.25	.15	.82	0.36	.26	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1992.4	.43	.25	.15	.82	0.36	.26	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1993.1	.43	.25	.15	.82	0.36	.26	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1993.2	.43	.25	.15	.82	0.36	.26	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1993.3	.43	.25	.15	.82	0.36	.26	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1993.4	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1994.1	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1994.2	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1994.3	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1994.4	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1995.1	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1995.2	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1995.3	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1995.4	.44	.25	.14	.81	0.37	.25	-0.04	0.95	1.41	.20	-0.31	.53	1.04	2
1996.1	.50	.25	.12	.96	0.29	.21	-0.12	1.36	0.97	.27	-0.14	.67	0.59	2
1996.2	.50	.25	.12	.96	0.29	.21	-0.12	1.36	0.97	.27	-0.14	.67	0.59	2
1996.3	.50	.25	.12	.96	0.29	.21	-0.12	1.36	0.97	.27	-0.14	.67	0.59	2
1996.4	.50	.25	.12	.96	0.29	.21	-0.12	1.36	0.97	.27	-0.14	.67	0.59	2
1997.1	.50	.25	.12	.96	0.29	.21	-0.12	1.36	0.97	.27	-0.14	.67	0.59	2
1997.2	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1997.3	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1997.4	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1998.1	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1998.2	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1998.3	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1998.4	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1999.1	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1999.2	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1999.3	.48	.25	.13	.94	0.32	.24	-0.05	1.22	0.98	.27	-0.13	.70	0.65	2
1999.4	.51	.24	.14	.88	0.31	.21	-0.12	1.48	0.78	.27	-0.02	.59	0.46	2

Table 1 continued

	AR(2)				SETAR(2;2,2), Regime 1				SETAR(2;2,2), Regime 2					
Vintage	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	r	d
2000.1	.51	.24	.14	.88	.31	.21	-0.12	1.48	.78	.27	-0.02	.59	.46	2
2000.2	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2000.3	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2000.4	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2001.1	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2001.2	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2001.3	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2001.4	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2002.1	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2002.2	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2002.3	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2002.4	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2003.1	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2003.2	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2003.3	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2003.4	.49	.25	.15	.89	.28	.22	-0.13	1.53	.77	.28	-0.02	.60	.43	2
2004.1	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2004.2	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2004.3	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2004.4	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2005.1	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2005.2	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2005.3	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2005.4	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2006.1	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2006.2	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2006.3	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2006.4	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2007.1	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2007.2	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2007.3	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2007.4	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2008.1	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2008.2	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2008.3	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2
2008.4	.51	.23	.15	.90	.34	.23	-0.06	1.31	.91	.24	-0.07	.64	.66	2

Table 2: Model constancy over vintage. Estimation sample 1957:2 to 2000:4.

	AR(2)				SETAR(2;2,2), Regime 1				SETAR(2;2,2), Regime 2					
Vintage	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	Int.	y_{t-1}	y_{t-2}	$\hat{\sigma}^2$	r	d
2001.1	.53	.29	.08	.83	.33	.19	-0.25	1.37	.63	.35	.01	.60	.43	2
2001.2	.53	.29	.08	.83	.33	.19	-0.25	1.37	.63	.35	.01	.60	.43	2
2001.3	.54	.28	.09	.84	.33	.19	-0.25	1.37	.63	.34	.02	.61	.43	2
2001.4	.54	.28	.09	.84	.33	.19	-0.25	1.37	.63	.34	.02	.61	.43	2
2002.1	.54	.28	.09	.84	.33	.19	-0.25	1.37	.63	.34	.02	.61	.43	2
2002.2	.54	.28	.09	.84	.33	.19	-0.25	1.37	.63	.34	.02	.61	.43	2
2002.3	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2002.4	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2003.1	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2003.2	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2003.3	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2003.4	.53	.28	.08	.83	.33	.19	-0.25	1.37	.61	.35	.02	.60	.43	2
2004.1	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2004.2	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2004.3	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2004.4	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2005.1	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2005.2	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2005.3	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2005.4	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2006.1	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2006.2	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2006.3	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2006.4	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2007.1	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2007.2	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2007.3	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2007.4	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2008.1	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2008.2	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2008.3	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2
2008.4	.55	.26	.09	.84	.32	.17	-0.30	1.40	.63	.34	.02	.59	.42	2

4.2 Pseudo out-of-sample forecasting and the data vintage

The fact that a model exhibits parameter non-constancy over time or over data vintage (as here) does not necessarily imply that the forecasts will be much affected.[7] For this reason, we calculate pseudo out-of-sample forecasts for all data vintages for each of the two samples. Our first forecasting exercise is based on the period 1959–1990. For a given data vintage, say the 1991:1 vintage, we initially estimated the SETAR model on the observations 1959 to 1980:4, and then generated a one-step ahead forecast of 1981:1. The estimation sample was then extended to include the 1981:1 observation, and the model was re-estimated and a forecast generated of 1981:2, and so on up to the final estimation period that contained observations through 1990:3. This generated a sequence of $P = 40$ one-step ahead forecasts for

[7] For example, the changes in individual parameters may be largely off-setting so that the conditional model of the forecast is little changed: see e.g., Clements and Hendry (2006).

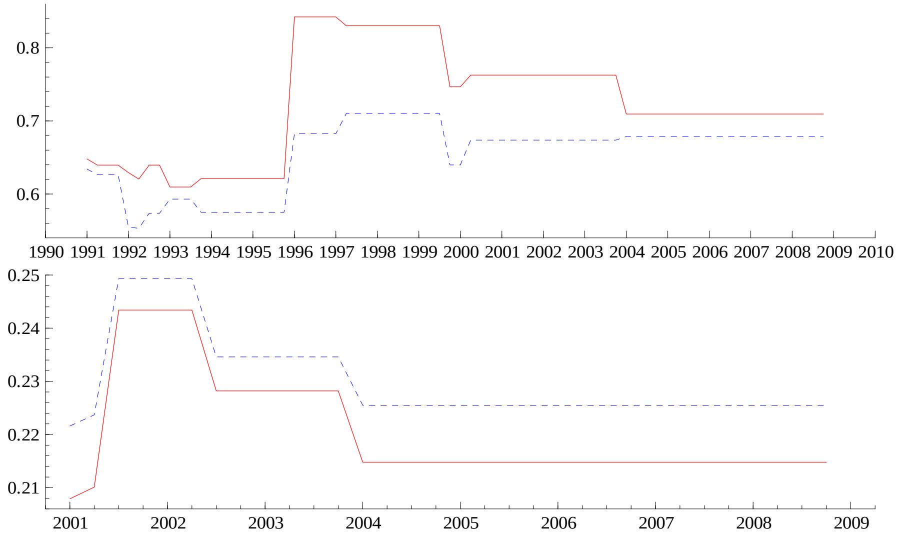

Figure 1: Pseudo out-of-sample MSFEs for recursive 1-step forecasts by data vintage. Top panel based on forecasts for the period 1981:1 to 1990:4, for the data vintages specified on the horizontal axis. (R=1980:4, T=1990:4). Bottom panel based on forecasts for the period 1991:1 to 2000:4, for the data vintages specified on the horizontal axis. (R=1990:4, T=2000:4). In both panels, the SETAR MSFEs are the solid line, and the AR MSFEs the dotted line. Note that key benchmark revisions occurred in 1996:1, 1999:4 and 2004:1 (top panel).

the 1991:1 data vintage. The accuracy of this set of forecasts was evaluated using the 1991:1 vintage 'actuals' to calculate the MSFE: in terms of the notation in section 3, for the AR and SETAR these corresponds to $MSFE_{AR,f}$ and $MSFE_{SETAR,f}$, where f =1991:1. We then repeated the exercise for f =1991:2, and so on, up to f =2008:4. Figure 1 (top panel) plots $MSFE_{AR,f}$ and $MSFE_{SETAR,f}$ for each of the seventy two data vintages 1991:1 to 2008:4. It is apparent that the AR model is always more accurate than the SETAR, but that the relative degree of inaccuracy of the SETAR varies considerably by data vintage, from around 2% to as much as 25% (the ratio of the SETAR to AR expressed as a percentage). It is readily apparent that for both models the main changes in accuracy occur at the times of benchmark revisions: the breaks in the figure occur for the vintages of 1996:1, 1999:4, and 2004:1. This suggests that benchmark revisions are not benign for forecast accuracy comparisons. For pseudo out-of-sample comparisons of forecast accuracy it matters which data vintage is employed. This is true both for assessing the relative accuracy of the non-linear model against the linear benchmark, and for assessing the absolute accuracy of the models for this period. After a benchmark revision, $MSFE_{AR,f}$ and $MSFE_{SETAR,f}$ remain largely unchanged across vintages until the next benchmark revision, as was expected given that only a small proportion of recent data observations change in response to regular revisions from one data vintage to the next.

For the second period, we initially estimated the models on the observations 1957:2 to 1990:4 from the 2001:1 data vintage, and then generated a one-step ahead forecast of 1991:1, and then extended the sample to include the 1991:1 observation, and generated a forecast of 1991:2, and so on up to the final estimation period that contained observations through 2000:3, so that again $P = 40$. We calculated the MSFE of these forecasts using the 2001:1 data vintage actual values. This process was then repeated for each of the other 31 data vintages from 2001:2 to 2008:4. Contrary to the first exercise, there was little variation in forecast accuracy of the SETAR across data vintage, either in absolute terms or relative to the accuracy of the AR: see figure 1 (bottom panel) . The SETAR forecasts are more accurate than the AR forecasts, and this is true irrespective of the data vintage. From the 2004:1 data vintage onwards the ratio to the SETAR to the AR MSFE was 0.95. Note the smaller range of the y-axis in the bottom panel, and the lower level of the MSFEs for both models: the period 1991-2000 was easier to predict than 1981-1990.

In the absence of benchmark revisions during the period 2001:1 to 2008:4, this lack of variation in MSFE across vintage is not surprising. Consider the first data vintage, 2001:1. The initial forecast of 1991:1 using data through 1990:4 will be based on 'mature' data, i.e., data that are fully-revised. The last forecast (of 2000:4 based on data through 2000:3) will contain some observations towards the end of the estimation sample which have not been subject to the standard three annual revisions. However, the effect of this data measurement error on the forecasts will be small as the MSFE is calculated by averaging across all the forecast errors for the period 1991:1 to 2000:4. For later data vintages, not even the recent forecasts will have been generated by models with observations still subject to the regular annual revisions.

4.3 Real-time forecasting exercise

Clearly, *pseudo* out-of-sample forecast exercises are not a suitable vehicle for gauging the impact of data uncertainty on forecasting models, even though much of the evidence on the value of non-linear models for forecasting is based on such exercises. In order to assess the effects of data uncertainty on forecasting in practice we carry out a real-time forecasting exercise that uses only the data vintage that would have been available at that time. This exercise is based on the second period, where we generated forecasts of 1991:1 to 2000:4. Recall that using data vintages 2001:1 to 2008:4, we found the MSFEs to be largely independent of the data vintage. Hence it does not matter which pseudo 'final-vintage' exercise we compare the real-time findings against, as the ratio of the pseudo-exercise MSFEs are unchanged for all final vintages from 2004:1 onwards.

We use exactly the same time periods for estimation and forecasting as in the pseudo-forecasting exercise. Consequently, the first forecast of 1991:1 is based on data through 1990:4 but taken from the vintage that was available at the time the forecast was made - this is the 1991:1 data vintage. The final forecast of 2000:4 is based on data through 2000:3 from the 2000:4 data vintage. Hence the forecasts will be conditioned on data which are first (or early) estimates, and will be generated using model estimates obtained from heavily revised data (older observations) and more recent observations which have not been subject to the regular rounds of revisions. In order to compare the real-time forecast exercise to the infeasible (in real time) exercise that uses 'final data', we take the exercise based on the 2008:4 vintage as our exemplar of the latter, and calculate forecast errors for the real-time exercise using actual values taken from the 2008:4 vintage. Doing so implies that the aim is to forecast the 'true values', which are best proxied by the latest-available vintage of data, rather than forecasting an early release. This seems a reasonable aim given that we are interested in forecasts of growth rates which were largely unaffected by benchmark revisions over this period. Using final-vintage actuals also allows us to directly compare the forecasts from models estimated on final (2008:4 vintage) data with the real-time forecasts for a *common* target. It seems reasonable to suppose that the real-time forecaster might wish to forecast final values, but it does not seem sensible to forecast early-release values using models estimated on - and forecasts conditioned on - final data. However, we do compare the real-time forecasts from the SETAR and AR models in terms of both first-release and first-revised actual values.

Our results are shown in table 3, for both 1 and 2-step ahead forecasts. Consider firstly the 1-step ahead forecasts. The results in rows 1 to 4 are for forecasting the 2008:4 vintage of data. In real-time, we find a small deterioration in the forecast performance of the linear AR model compared to the pseudo exercise (MSFE increases from 0.23 to 0.25) but a large worsening in accuracy of the SETAR (from 0.21 to 0.33: compare rows 1 and 2). In order to investigate the worsening forecast performance of the SETAR, we re-run the real-time exercise but at each forecast origin we condition the forecast on the final values (the 2008:4 data vintage values) of the relevant observations, but as before the model is estimated on the vintage available at that forecast origin (this is the $MSFE_{.,EOS,x^f}$ measure of section 3). We also calculate $MSFE_{.,f,x^{t+1}}$ for both models. In both these cases (rows 3 and 4 in the table), the SETAR MSFE is around 0.25, accounting for approximately one third of the deterioration in the MSFE in going from the pseudo exercise to the full real-time exercise (row 2), suggesting that the interaction of the two factors - the real-time estimates, and conditioning on the real-time data - has a greater impact than either factor alone. For the 2-step ahead horizon, the findings in broad terms are similar. Now the SETAR is more than 10% more accurate than the AR on MSFE in the pseudo out-of-sample exercise, but as in the case of the 1-step forecasts loses out in the real-time exercise (row 2). For the 2-step forecasts we find that conditioning the forecasts on the real-time data - with final data parameter estimates (row 4) is more harmful than the reverse combination (row 3: conditioning on final values with real-time parameter estimates).

Rows 5 and 6 of the table report results for early-release actual values. Consider firstly the 1-step forecasts in the top panel. Now, the forecasts of, say, 1991:1 are compared to the value of the 1991:1 observation in either the 1991:2 data vintage or the 1991:3 data vintage. As explained above, we only compare the real-time forecasts of the models in terms of their ability to predict early-release actual values. The real-time forecasting results using early-release actuals match those for forecasting the final data in real time, in that the SETAR model forecasts are less accurate than the AR model forecasts (although relatively less so). Both the AR and SETAR model forecasts are more accurate at predicting early-release data than the 2008:4 vintage data - the SETAR model MSFE approximately halves. The relative accuracy of the two models is similar whether we use first-release or first revised data. Hence our finding that the SETAR forecasts worsen in real time relative to the pseudo exercise (rows 1 and 2) is not dependent on the use of final-vintage data - the same is true for predicting early-release data. The 2-step ahead results using real-time actuals are in line with these findings for the 1-step ahead forecasts.

In the following section we report a simulation study of the effect of data revisions on forecast performance, in an attempt to shed further light on the empirical forecasting exercise.

Published by De Gruyter, 2012

Table 3: Pseudo out-of-sample and Real-time SETAR and AR Forecast Performance

		SETAR	AR	SETAR/AR
1-step ahead forecasts				
1.	Pseudo out-of-sample, $MSFE_{.,f}$	0.21	0.23	0.95
2.	Real-time, $MSFE_{.,EOS}$	0.33	0.25	1.34
3.	'Real-time', $MSFE_{.,EOS,x^f}$	0.25	0.23	1.12
4.	'Real-time', $MSFE_{.,f,x^{t+1}}$	0.25	0.24	1.06
5.	Real-time, $MSFE_{.,EOS}$	0.18	0.16	1.14
6.	Real-time, $MSFE_{.,EOS}$	0.23	0.20	1.13
2-step ahead forecasts				
1.	Pseudo out-of-sample, $MSFE_{.,f}$	0.17	0.20	0.87
2.	Real-time, $MSFE_{.,EOS}$	0.27	0.22	1.24
3.	'Real-time', $MSFE_{.,EOS,x^f}$	0.21	0.21	0.99
4.	'Real-time', $MSFE_{.,f,x^{t+1}}$	0.24	0.20	1.16
5.	Real-time, $MSFE_{.,EOS}$	0.16	0.13	1.16
6.	Real-time, $MSFE_{.,EOS}$	0.21	0.19	1.10

The forecast period is 1991:1 to 2000:4 for the 1-step forecasts, and 1991:2 to 2001:1 for the 2-step forecasts. The initial estimation period is 1957:2 to 1990:4 (from the 1991:1 vintage for the real-time exercise, from the 2008:4 vintage for the pseudo exercise), and the last is 1957:2 to 2000:3 (from the 2000:4 vintage for the real-time exercise, and from the 2008:4 for the pseudo exercise). The actual values are taken from the 2008:4 data vintage, for rows 1 to 4, are first-release for row 5, and first-revised for row 6.

5 Monte Carlo

We analyse the impact of data revisions on SETAR model forecast performance by Monte Carlo simulation. The advantage of using Monte Carlo is that we are able to focus on the issue of interest - data revisions - and abstract from other factors that influence empirical comparisons of forecast performance, such as the extent to which the SETAR model is a 'good' characterization of the data, and the constancy of the model over time, for example. However, the Monte Carlo requires a complete specification of the data generation process, including in our case the way in which different vintage estimates of the same observation are generated. We suppose that a specific data vintage estimate is equal to the true value plus an error, such that y_t^{t+s} consists of the true value $\widetilde{y}_t$, as well as (in the general case) news and noise components, v_t^{t+s} and ε_t^{t+s}, $y_t^{t+s} = \widetilde{y}_t + v_t^{t+s} + \varepsilon_t^{t+s}$. Data revisions are news when the earlier-released data are optimal forecasts of later data, so the revisions between the two releases are not correlated with the earlier data, $Cov\left(v_t^{t+s}, y_t^{t+s}\right) = 0$, for all s.[8] Data revisions are noise when the revisions are not correlated with the truth, $Cov\left(\varepsilon_t^{t+s}, \widetilde{y}_t\right) = 0$. We adopt the framework of Jacobs and van Norden (2007) which stacks the l different vintage estimates of y_t, namely, $y_t^{t+1}, \ldots, y_t^{t+l}$ in the vector $\mathbf{y}_t = \left(y_t^{t+1}, \ldots, y_t^{t+l}\right)'$, and similarly $\varepsilon_t = \left(\varepsilon_t^{t+1}, \ldots, \varepsilon_t^{t+l}\right)'$ and $\mathbf{v}_t = \left(v_t^{t+1}, \ldots, v_t^{t+l}\right)'$, so that:

$$\mathbf{y}_t = \mathbf{i}\widetilde{y}_t + \mathbf{v}_t + \varepsilon_t \tag{7}$$

where $\mathbf{i}$ is a l-vector of ones. We assume that $\widetilde{y}_t$ follows a two-regime SETAR process with regime-dependent disturbances η_{1t}, plus the 'news term' $-v_t^{t+1}$:

$$\widetilde{y}_t = 1_{\widetilde{y}_{t-d} \leq r}\left(\rho_0^1 + \sum_{i=1}^{2} \rho_i^1 \widetilde{y}_{t-i} + \sigma_1 \eta_{1t}\right) + 1_{\widetilde{y}_{t-d} > r}\left(\rho_0^2 + \sum_{i=1}^{2} \rho_i^2 \widetilde{y}_{t-i} + \sigma_2 \eta_{1t}\right) - v_t^{t+1}. \tag{8}$$

In the formulation of Jacobs and van Norden (2007):

$$\mathbf{v}_t = \begin{bmatrix} v_t^{t+1} \\ v_t^{t+2} \\ \vdots \\ v_t^{t+l} \end{bmatrix} = -\begin{bmatrix} \sum_{i=1}^{l} \sigma_{v_i} \eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v_l} \eta_{2t,l} \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} \varepsilon_t^{t+1} \\ \varepsilon_t^{t+2} \\ \vdots \\ \varepsilon_t^{t+l} \end{bmatrix} = \begin{bmatrix} \sigma_{\varepsilon_1} \eta_{3t,1} \\ \sigma_{\varepsilon_2} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon_l} \eta_{3t,l} \end{bmatrix}, \tag{9}$$

where $\eta_t = \left[\eta_{1t}, \eta_{2t}', \eta_{3t}'\right]$, $\eta_{2t}' = \left[\eta_{2t,1}, \ldots, \eta_{2t,l}\right]$, $\eta_{3t}' = \left[\eta_{3t,1}, \ldots, \eta_{3t,l}\right]$, and $E\left(\eta_t\right) = 0$, $E\left(\eta_t \eta_t'\right) = I$.

To see how this defines news and noise revisions, consider the case of news revisions, so that $\varepsilon_t = 0$. Then the first estimate is $y_t^{t+1} = \widetilde{y}_t + v_t^{t+1}$ is simply the first two terms of the RHS of (8) - the first estimate contains no news. The second estimate, $y_t^{t+2} = \widetilde{y}_t + v_t^{t+2}$ is the first two terms of the RHS of (8) plus

[8] This says that the difference between y_t^{t+s} and $\widetilde{y}_t$ is uncorrelated with y_t^{t+s} for all $1 \leq s \leq l$. The formulation also implies that all pairwise revisions are uncorrelated with the earlier value, etc., $Cov\left(y_t^{t+s+1} - y_t^{t+s}, y_t^{t+s}\right) = 0$, for all $1 \leq s \leq l-1$.

$v_t^{t+2} - v_t^{t+1} = \sigma_{v_1}\eta_{2t,1}$. Hence the revision between the first and second estimates is simply $\sigma_{v_1}\eta_{2t,1}$, which is uncorrelated with y_t^{t+1}. Successive estimates incorporate additional news terms, until $y_t^{t+l} = \widetilde{y}_t$ (if we assume $\sigma_{v_l} = 0$).

The formulation for noise is simpler - by design the noise terms do not appear in (8). If we assume that σ_{ε_s} decreases in s, then later estimates $y_t^{t+s} = \widetilde{y}_t + \sigma_{\varepsilon_s}\eta_{3t,s}$ are more accurate (less noisy) than earlier estimates.

In our Monte Carlo, we allow that data revisions to output growth are either news or noise.[9] The precise formulation we adopt in the Monte Carlo is a little more complicated than this, because we follow Clements and Galvão (2010) in specifying the revisions process to capture the way in which data are typically revised by the BEA. The first estimate is always revised so that $y_t^{t+2} \neq y_t^{t+1}$. However, the data for y_t is then left unchanged (ignoring benchmark revisions) until the annual revisions that occur in the 3rd quarters of the next three years. To see what this entails, suppose $t \in Q_1$. Then considering news revisions, we have that $y_t^{t+2} - y_t^{t+1} = v_t^{t+2} - v_t^{t+1} = \sigma_{v_1}\eta_{2t,1}$, where y_t^{t+2} is issued as part of the annual revision ($t+2 \in Q_3$), but thereafter $y_t^{t+i} - y_t^{t+i-1} = v_t^{t+i} - v_t^{t+i-1} = 0$ for $i = 3,4,5$, until $i = 6$, which corresponds to the second annual revision. Then, $y_t^{t+6} - y_t^{t+5} = v_t^{t+6} - v_t^{t+5} = \sigma_{v_1}\eta_{2t,1}$. Similar logic underlies the specification of the news terms for t falling in one of the other quarters of the year. The following specification of the news components captures the seasonality of the pattern of 'regular revisions':

$$\mathbf{V}_1 = \begin{bmatrix} \sigma_{v_1} & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{V}_2 = \begin{bmatrix} \sigma_{v_1} & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{V}_3 = \begin{bmatrix} \sigma_{v_1} & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{V}_4 = \begin{bmatrix} \sigma_{v_1} & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & \sigma_{v_2} & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & \sigma_{v_3} & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & \sigma_{v_4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so that $\mathbf{v}_{t,t\in Q_i} = -\mathbf{V}_i \times \left[\eta_{2t,1}, \eta_{2t,2}, \eta_{2t,3}, \eta_{2t,4}\right]'$.

[9]Mankiw and Shapiro (1986) finds revisions to output growth are news, whereas more recently Aruoba (2008) suggests some predictability of revisions (i.e., noise), so that both possibilities are of interest.

For noise, we have:

$$\varepsilon_{t,t\in Q_1}=\begin{bmatrix}\sigma_{\varepsilon_1}\eta_{3t,1}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ 0\end{bmatrix},\ \varepsilon_{t,t\in Q_2}=\begin{bmatrix}\sigma_{\varepsilon_1}\eta_{3t,1}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ 0\\ 0\end{bmatrix},\ \varepsilon_{t,t\in Q_3}=\begin{bmatrix}\sigma_{\varepsilon_1}\eta_{3t,1}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ 0\\ 0\\ 0\end{bmatrix},\ \varepsilon_{t,t\in Q_4}=\begin{bmatrix}\sigma_{\varepsilon_1}\eta_{3t,1}\\ \sigma_{\varepsilon_2}\eta_{3t,2}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_3}\eta_{3t,3}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ \sigma_{\varepsilon_4}\eta_{3t,4}\\ 0\\ 0\\ 0\\ 0\end{bmatrix}.$$

The specification that we have described assumes that there are only four underlying variance terms to be calculated. One might allow the $\{\sigma_{v_1},\sigma_{v_2},\sigma_{v_3},\sigma_{v_4}\}$ to be indexed by quarter, so that the effect of (say) the first annual revision depends on the quarter to which t belongs, but this seems an unnecessary complication.

We chose as our DGP the SETAR model described in section 4: that is, the ρ parameters in (8) of the $\text{SETAR}(2;2,2)$ were set to the values obtained from estimating the model on the sample 1947-1990 from the 1991:1 data vintage. Recall that for this period the model exhibits quite different behaviour in the 'recession regime', characterised by negative output growth in period $t-2$, so that the data we simulate will have a distinct 'nonlinear imprint'.[10]

Assuming news revisions, we then estimated σ_{v_1} as the standard deviation of all first revisions (irrespective of the quarter t falls in); σ_{v_2} is the standard deviation of 5th revisions for $t\in Q1$, 4th revisions for $t\in Q2$, 3rd revisions for $t\in Q3$, 2nd revisions for $t\in Q4$; σ_{v_3} is the standard deviation of 9th revisions for $t\in Q1$, 8th revisions for $t\in Q2$, 7th revisions for $t\in Q3$, 6th revisions for $t\in Q4$; and σ_{v_4} is the standard deviation of the 13th revisions for $t\in Q1$, 12th revisions for $t\in Q2$, 11th revisions for $t\in Q3$, and 10th revisions for $t\in Q4$. Using the first to fourteenth estimates of data for the time periods 1965:3 to 1990:4 we obtained estimates of σ_{v_i}, $i=1,\ldots,4$. We also calculated σ_{ε_i}, $i=1,\ldots,4$ for noise revisions.

[10] Hence the parameters of (8) are estimated on the 1991:1 vintage, although this equation is for $\widetilde{y}_t$, which suggests it should refer to 'fully revised data' for all t. Further, our estimates of σ_1 and σ_2 in equation (8) will be an amalgam of the regime-specific disturbance terms and the measurement errors. We then calculate the news/noise disturbance terms separately as described in the text. Our interest is not in accurately estimating a SETAR model for underlying output growth when there are data revisions, but in determining the effects of revisions when the underlying process is clearly nonlinear.

Table 4: Monte Carlo results

			Pseudo out-of-sample			Real-time						
	Rev	Act.	SETAR	AR	SETAR/ AR	SETAR	AR	SETAR/ AR	SETAR, $_{EOS,x^f}$	SETAR, $_{EOS,x^f}/_f$	SETAR, $_{f,x^{t+1}}$	SETAR, $_{f,x^{t+1}}/_f$
1	.	Fin	1.165	1.132	1.029	.	.	.	.	.	.	.
2	.	Fin	1.052	1.106	0.951	.	.	.	.	.	.	.
3	Ne	Fin	1.212	1.268	0.955	1.274	1.297	0.982	1.212	1.000	1.274	1.052
4	Ne	2nd	1.212	1.268	0.955	1.177	1.200	0.981	1.115	0.920	1.176	0.971
5	Noi	Fin	1.052	1.106	0.951	1.073	1.109	0.967	1.054	1.001	1.072	1.019
6	Noi	2nd	1.052	1.106	0.951	1.101	1.138	0.967	1.082	1.028	1.101	1.046

Notes: The data generating process is the SETAR(2;2,2) as described in the text. The estimation sample is 200 for rows 2 to 6, and 100 for the first row. The second column 'Rev.' denotes whether revisions are news ('Ne'), or noise ('Noi') or zero, '.', and the third column ('Act.') whether the actual values used to calculated forecast errors are y_t^{t+14} ('Fin') or 2nd-estimates y_t^{t+2} ('2nd'). The remaining columns are, for the pseudo exercise, the Monte Carlo estimates (20,000 replications) of the estimated SETAR model MSFE, the AR model MSFE, the ratio of the two; and then these three are repeated for the real-time exercise. We then present the 'real-time' SETAR MSFE when the forecasts are conditioned on final data ($_{EOS,x^f}$), and the ratio of this to the pseudo out-of-sample SETAR MSFE ($_{EOS,x^f}/_f$). Finally, the last two columns are the 'real-time' SETAR MSFE when the model estimates are calculated on final data ($_{f,x^{t+1}}$), and the ratio of this to the pseudo out-of-sample SETAR MSFE ($_{f,x^{t+1}}/_f$).

Table 4 reports the results of a Monte Carlo based on 20,000 replications of a recursive 1-step ahead forecasting exercise for $T = 100$, 200 and $P = 8$. We consider six 'cases' or sets of design parameters, chosen to highlight aspects of interest. Cases 1 and 2 set all data revisions to zero, so that the data do not change across vintage. Hence the real-time exercise is the same as the pseudo out-of-sample (which accounts for the 'dots' in the tables, where unchanged entries are not repeated). The SETAR model is 5% more accurate than the AR when $T = 200$ (Case 2), but less accurate than the AR on the shorter sample, $T = 100$ (Case 1), indicating that a relatively large sample is required for the SETAR *model* to outperform the AR even when the data are generated from a SETAR. This is consistent with the general findings in the literature concerning the poor forecasting performance of the SETAR (see section 2), and suggests that parameter estimation uncertainty detracts from the forecast performance of the SETAR even when there are no data revisions. For the remaining cases 3–6, we set $T = 200$.

Consider news revisions (Cases 3 and 4). In the pseudo exercise, the absolute accuracy of both models worsens, but the SETAR remains around 5% more accurate than the AR, as when there are no revisions.[11] The relative accuracy of the SETAR worsens in the real-time exercise (although it remains more accurate than the AR). Conditioning the SETAR forecasts on the final data (but using real-time parameter estimates) restores the accuracy of the forecasts to that in the pseudo-exercise, when forecast errors are calculated using final-value actuals, whereas using real-time estimates alone has little effect. Hence the Monte Carlo suggests that it is the effect of conditioning forecasts on data measured with error that leads to the worsening of forecast performance in real-time.[12] The results for noise are broadly similar - the relative performance of the SETAR worsens less in real-time than when revisions are news, but as in the case of news, conditioning the forecasts on final data is sufficient to restore the forecast performance of the SETAR.

6 Conclusion

There is now an established set of explanations as to why non-linear models may not forecast more accurately than linear models in terms of out-of-sample forecasting. These include: the possibility that the out-of-sample period is not characterised by the 'non-linear features' that typify the in-sample data ((Granger and Teräsvirta, 1993, chapter 9) and Teräsvirta and Anderson (1992)), including that the 'minority-regime' is not realized (e.g., Tiao and Tsay (1994)) and/or the evaluation is not done conditional on the regime; the apparent non-linearities are due to outliers or structural breaks, but these offer no gain in improved out-of-sample performance; non-linearities are a feature of the DGP, but are not large enough to yield much of an improvement to forecasting; and finally, they are present and important but the wrong types of non-linear models have been used to try and capture them (Diebold and Nason (1990), *inter alia*). To these we add the impact of data uncertainty.

There are three main findings. In pseudo out-of-sample forecast comparisons, we find that benchmark revisions affect the forecast accuracy of the SETAR model of US output growth, and affect its relative accuracy compared to a linear comparator. Forecast performance depends on which vintage of data is taken as the 'final data'. Secondly, pseudo out-of-sample forecast comparisons may not be a good guide to real-time rankings of models in terms of forecast performance when one of the models is a non-linear regime-switching model. For the case of

[11] Note that the entries are the same for the pseudo-out-of-sample exercise for both 'final' and '2nd' actuals. This is because we always use final actuals for the pseudo-out-of-sample exercise. The two different sets of actuals only apply for the real-time exercises.

[12] Case 4 evaluates the forecasts in terms of an early release (y_t^{t+2}), and suggests that conditioning the 'real-time' SETAR forecasts on final data will improve upon the pseudo exercise. But this is an odd exercise, as explained in section 4.3.

US output growth, the forecasts of the SETAR model are far less accurate in real time. Thirdly, in the empirical exercise the worsening in performance of the SETAR model was due to the interaction of the effects of data mis-measurement on the parameter estimates and in the data the forecasts were conditioned on, in the case of 1-step forecasting, but was primarily due to errors in the data being conditioned on for the 2-step forecasts. Our Monte Carlo suggested that errors in the data being conditioned on in generating the forecast would have an adverse on the SETAR forecasts, but that the impact on accuracy of the effect of measurement errors on model estimation is small. The latter accords with the limited impact we would expect on parameter estimates for reasonable sample sizes given the relatively small number of data revisions. The Monte Carlo misses the 'interaction effect' found empirically for the 1-step forecasts - that it is the conjunction of the two effects that is mainly responsible for the worsening in accuracy. It is not surprising that the model underpinning the Monte Carlo misses some features that are important empirically, given the complicated nature of the empirical revisions process, and possible instabilities. The bottom line is that the Monte Carlo confirms the general finding that the SETAR model's forecast performance will look better using final data than in real-time.

These findings relate to a single macroeconomic time series and a particular non-linear model. It would be of interest to establish how general our findings are.

References

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

Box, G. E. P. and G. M. Jenkins (1970): *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Clements, M. P. and A. B. Galvão (2008): "Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States," *Journal of Business and Economic Statistics*, 26, 546–554, no. 4.

Clements, M. P. and A. B. Galvão (2009): "Forecasting US output growth using leading indicators: An appraisal using MIDAS models," *Journal of Applied Econometrics*, 24, 1187–1206.

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

Clements, M. P. and D. F. Hendry (1999): *Forecasting Non-Stationary Economic Time Series*, Cambridge, Mass.: MIT Press, the Zeuthen Lectures on Economic Forecasting.

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

Clements, M. P. and H.-M. Krolzig (1998): "A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP," *Econometrics Journal*, 1, C47–75.

Clements, M. P. and H.-M. Krolzig (2003): "Business cycle asymmetries: Characterisation and testing based on Markov-switching autoregressions," *Journal of Business and Economic Statistics*, 21, 196–211.

Clements, M. P. and J. Smith (1997): "The performance of alternative forecasting methods for SETAR models," *International Journal of Forecasting*, 13, 463–475.

Clements, M. P. and J. Smith (1999): "A Monte Carlo study of the forecasting performance of empirical SETAR models," *Journal of Applied Econometrics*, 14, 124–141.

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

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

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

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

Diebold, F. X. and J. A. Nason (1990): "Nonparametric exchange rate prediction," *Journal of International Economics*, 28, 315–332.

Diebold, F. X. and G. D. Rudebusch (1991): "Forecasting output with the composite leading index: A real-time analysis," *Journal of the American Statistical Association*, 86, 603–610.

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

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

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

Franses, P. H. and D. van Dijk (2000): *Non-linear time series models in empirical finance*, Cambridge: Cambridge University Press.

Granger, C. W. J. and T. Teräsvirta (1993): *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Hall, R. E. (1978): "Stochastic implications of the life cycle-permanent income hypothesis: Evidence," *Journal of Political Economy*, 86, 971–987.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.

Harrison, R., G. Kapetanios, and T. Yates (2005): "Forecasting with measurement errors in dynamic models," *International Journal of Forecasting*, 21, 595–607.

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

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

Jacobs, J. P. A. M. and S. van Norden (2007): "Modeling data revisions: Measurement error and dynamics of 'true' values," Technical report cref 07-09, HEC, Montreal.

Kishor, N. K. and E. F. Koenig (2010): "VAR estimation and forecasting when data are subject to revision," *Journal of Business and Economic Statistics*, forthcoming.

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

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

Mankiw, N. G. and M. D. Shapiro (1986): "News or noise: An analysis of GNP revisions," *Survey of Current Business (May 1986), US Department of Commerce, Bureau of Economic Analysis*, 20–25.

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

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

Potter, S. (1995): "A nonlinear approach to US GNP," *Journal of Applied Econometrics*, 10, 109–125.

Runkle, D. E. (1998): "Revisionist history: How data revisions distort economic policy research," *Federal Reserve Bank of Minneapolis Quarterly Review*, 22, 3–12, no. 4.

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

Siklos, P. L. (2008): "What can we learn from comprehensive data revisions for forecasting inflation: Some US evidence," in D. E. Rapach and M. E. Wohar, eds., *Forecasting in the Presence of Structural Breaks and Model Uncertainty. Frontiers of Economics and Globalization. Volume 3*, Emerald, 271–299.

Stock, J. H. and M. W. Watson (1996): "Evidence on structural instability in macroeconomic time series relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Teräsvirta, T. and H. M. Anderson (1992): "Characterizing nonlinearities in business cycles using smooth transition autoregressive models," *Journal of Applied Econometrics*, 7, 119–139.

Tiao, G. C. and R. S. Tsay (1994): "Some advances in non-linear and adaptive modelling in time-series," *Journal of Forecasting*, 13, 109–131.

Tong, H. (1978): "On a threshold model," in C. H. Chen, ed., *Pattern Recognition and Signal Processing*, Amsterdam: Sijhoff and Noordoff, 101–141.

Tong, H. (1983): *Threshold Models in Non-Linear Time Series Analysis*, Springer-Verlag, New York.

Tong, H. (1995): *Non-linear Time Series. A Dynamical System Approach*, Oxford: Clarendon Press, first published 1990.

Tong, H. and K. S. Lim (1980): "Threshold autoregression, limit cycles and cyclical data," *Journal of The Royal Statistical Society*, B 42, 245–292.

West, K. D. (2006): "Forecasting evaluation," in G. Elliott, C. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, Elsevier, Horth-Holland, 99–134.