Forecasting the U.S. Unemployment Rate

Tommaso Proietti*
Dipartimento di Scienze Statistiche
Università di Udine

6th April 2001

Abstract

The primary interest of this paper is in out-of-sample forecasting for the U.S. monthly unemployment rate. Several linear unobserved components models are fitted and their comparative forecasting accuracy is assessed by means of and extensive rolling-origin procedure using a test period that covers the last two decades. An attempt is made to link forecasting performance to the time domain properties of the models and the evidence is that highly persistent models perform better. Deletion diagnostics and normality tests, along with documenting possible departures from linearity and Gaussianity attributable to business cycle and turning point asymmetries, foster the conclusion that these are mostly concentrated in the fit period (1948-1980). It is also argued that seasonal adjustment is not neutral with respect to these findings. A search is made for plausible non linear extensions capable of accounting for dynamic asymmetries in unemployment rates, leading to the specification of a cyclical trend model with smooth transition in the underlying parameters that improves forecast accuracy at short lead times and at the end of the sample period. Though significant, the gains are not exceptionally large, confirming our expectations. The generalised impulse response function casts some light on the interpretation of the results.

Keywords: Structural time series models; Nonlinearity; Forecasting performance; Persistence; Seasonal adjustment; Leave-k-out diagnostics; Generalised impulse response function.

*Dipartimento di Scienze Statistiche, Via Treppo 18, 33100, Udine, Italy. E-mail: proietti@dss.uniud.it.
This research was supported by Cofin. MURST 2000 - prot. MM13035581_003 as part of the project *Linearity and Nonlinearity in Time Series Dynamics* (Coord. Prof. E.B. Dagum).

1 Introduction

The U.S. unemployment rate represents a case study in nonlinear dynamics. Asymmetric behaviour over the course of the business cycle has been documented in a variety of papers including Neftçi (1984), DeLong and Summers (1986) and Rothman (1991), who deal with the type of asymmetry named *steepness*, taking place when contractions are steeper than expansions: this implies that unemployment rates rise faster than they decrease. Sichel (1993) found evidence for *deepness*, which occurs when contractions are deeper than expansions, so that the amplitude of peaks in unemployment rates exceeds that of troughs; McQueen and Thorley (1993) detect turning point asymmetry (*sharpness*), such that peaks are sharp and troughs are more rounded.

Needless to say, the series represents a testbed for nonlinear time series models; within the class of regime switching models, threshold autoregressive models (TAR) are prominent; references include Hansen (1997), Koop and Porter (1999), who focus on the monthly unemployment rate for males aged 20 and over, and Montgomery et al. (1998). The latter conduct a rolling forecast experiment for the rates, which shows that TAR and Markov-switching models outperform the linear benchmark model during periods of rapidly increasing unemployment, but not globally; moreover, they find that using monthly data for forecasting quarterly rates improves the forecasting accuracy only in the short term.

Skalin and Terasvirta (1999) use a logistic smooth transition autoregressive model (LSTAR) for the first differences of the unemployment rate of OECD countries including a lagged level term and were unable to reject linearity for the U.S. quarterly seasonally unadjusted series. Their specification assumes that the series is globally stationary, but possibly nonlinear and locally nonstationary. Van Dijk *et al.* (2000) apply the model to the seasonally unadjusted monthly series for males aged 20 and perform an out-of-sample forecast accuracy analysis showing that the LSTAR model outperforms the linear AR counterpart at long run forecast horizons during downturns and at short run horizons during expansions.

Rothman (1998) compares the out-of-sample forecasting accuracy of six nonlinear models and finds that results are sensitive to the detrending issue. Parker and Rothman (1998) model the quarterly adjusted rate by an AR(2) process including an explanatory variable measuring the *current depth of recession*, and show with a rolling forecast experiment that significant reductions of the forecast MSE are achieved.

The primary interest of this paper is in out-of-sample forecasting for the U.S. unemployment rate. The series considered is the monthly seasonally adjusted series, defined as the ratio of the seasonally adjusted unemployment level and the civilian labour force level provided by the Bureau of Labor Statistics (BLS). This is plotted in figure 1, along with the unadjusted series, that will also be dealt with at some stage. *Prima facie* the plot confirms the presence of dynamic asymmetries in the form of steepness, which appears to be the dominant feature of the cycle dynamics during the seventies and the mid-fifties; the last two decades and some cyclical patterns at the beginning of the sample period and around 1962 seem to be more characterised by deepness and sharpness, in combination with moderate steepness, since a period of fastly increasing rates is followed by fast decreases. Perhaps a three phases characterisation is enforced here with the third phase representing a more prolonged and moderate decline in unemployment rates. Another feature of interest is the tendency of the series to remain on a level it has reached, with no apparent tendency to return to a stable underlying level; this is referred to as hysteresis or persistence.

Our approach is much in the same spirit of Montgomery *et al.* (1998) in that it focuses on an in-depth comparison of forecasting models (with an emphasis on short term forecasting), aimed at providing an understanding of the strengths and weaknesses of each. With respect to the above article we concentrate on monthly rather than quarterly data, extend the out-of-sample forecast comparison and adopt an unobserved components modelling approach.

The outline of the paper is as follows: section 2 introduces a few linear unobserved components models for forecasting the U.S. unemployment rate and illustrates their basic properties, along with the implied impulse response function. Forecasting accuracy is validated in section 3 by a rolling forecast experiment conducted over the period 1980.1-2000.12. A comparison is also made with the ARIMA benchmark adopted by Montgomery *et al.* (1998). The best performance is provided by a trend plus irregular model with the trend specified as an highly persistent ARIMA(1,1,0) process.

We then document departures from linearity and normality (section 4) using various residuals and deletion diagnostic resulting from the linear model fit. The conclusion is that these are most prominent in the test period. Section 5 raises two issues; the first is whether the appropriate transformation (logistic) affects the results and whether seasonal adjustment affects the findings about persistence and nonlinearity.

In section 6 nonlinear alternatives are specified to account for dynamic asymmetries in unemployment rates. This is done by imposing smooth transition on the parameters of the model. A new transition variable that is better behaved is introduced and the results are discussed. In section 7 we deal with forecasting with nonlinear structural models and report the outcome of the rolling forecast experiment; these show that a nonlinear cyclical trend model outperforms the linear benchmark at short run forecast horizons. Interestingly, the improvement is greater at the end of the sample period. To interpret this finding and to gain a better understanding of the dynamic properties of the model we find the generalised impulse response function (section 8) quite helpful. Section 9 concludes the paper.

2 Linear Structural Models

In this section we consider seven forecasting models for the levels of the unemployment rate; the models entertained assume that hysteresis arises from the presence of a unit root in the reduced form; stationarity tests and unit root tests, although the latter are not directly relevant here due to the presence of moving average features, support this assumption. Moreover, the argument that unemployment rates are stationary due to the bounded nature of the measurement scale is flawed, since we can always view the series as a nonlinear transformation of a nonstationary series (an appropriate scale for testing stationarity and integratedness would be the logistic scale, see also section 5).

Although we have in mind a set of alternative that are more plausible than others, the uncertainty lies in the characterisation of the persistence and in the modelling of the short run dynamics; so the approach we take in this section and the next is that of exploring the strengths and deficiencies of each forecasting model.

2.1 Specification

The seven models are hereby described; for more details on their time and frequency domain properties the reader is referred to Harvey (1989).

1. Local level model (LLM):

$$\begin{array}{rcll} y_t &=& \mu_t + \epsilon_t, \quad t = 1, 2, \dots, T, & \epsilon_t \sim \mathsf{NID}(0, \sigma^2_\epsilon) \\ \mu_{t+1} &=& \mu_t + \eta_t, & \eta_t \sim \mathsf{NID}(0, \sigma^2_\eta) \end{array} \tag{1}$$

   The series is decomposed into a trend component, μ_t, represented by a random walk with IID disturbances, and an irregular component, ϵ_t; the reduced form is IMA(1,1), with negative MA coefficient, and thus the model implies that persistence is not greater that one.

2. Local linear trend model (LLTM):

$$\begin{array}{rcllll} y_t &=& \mu_t &+ \epsilon_t, & & \epsilon_t \sim \mathsf{NID}(0, \sigma^2_\epsilon), \quad t = 1, 2, \dots, T, \\ \mu_{t+1} &=& \mu_t &+ \beta_t &+ \eta_t, & \eta_t \sim \mathsf{NID}(0, \sigma^2_\eta), \\ \beta_{t+1} &=& & \beta_t &+ \zeta_t, & \zeta_t \sim \mathsf{NID}(0, \sigma^2_\zeta), \end{array} \tag{2}$$

   where η_t, ζ_t, ϵ_t are independent of one another. The trend is now given by an IMA(2,1) process, with the Hodrick and Prescott (1997) trend arising as a special case ($\sigma^2_\eta = 0$, smoothing parameter $\lambda = \sigma^2_\epsilon / \sigma^2_\zeta$). The reduced form is a restricted IMA(2,2) model.

3. Trend plus cycle model (TpCM):

$$y_t = \mu_t + \psi_t + \epsilon_t, \quad \mu_{t+1} = \mu_t + \eta_t;$$

   the cyclical component, ψ_t, is specified by the stochastic difference equation:

$$\begin{bmatrix} \psi_{t+1} \\ \psi^*_{t+1} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi^*_t \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix}, \tag{3}$$

   where $\kappa_t \sim \mathsf{NID}(0, \sigma^2_\kappa)$ and $\kappa^*_t \sim \mathsf{NID}(0, \sigma^2_\kappa)$ are mutually independent and independent of η_t, ϵ_t at all times; $\rho \in [0, 1)$ is the damping factor and $\lambda \in [0, \pi]$ is the cycle frequency; ψ_t has univariate ARMA(2,1) representation such that the roots of the AR polynomial are a pair of complex conjugates with modulus ρ^{-1} and phase λ_c; correspondingly, the spectral density displays a peak at λ_c. The model thus assumes that changes in unemployment can be decomposed into permanent changes and transitory changes.

4. Cyclical trend model (CTM):

$$y_t = \mu_t + \epsilon_t, \quad \mu_{t+1} = \mu_t + \psi_t + \eta_t,$$

   where ψ_t is given in (3). The trend is an ARIMA(2,1,2) process with two sources of variation, one represented by IID disturbances, η_t, and the other induced by ψ_t. Thus, cyclical movements are integrated within the trend; the model is consistent with an alternative definition of business cycle fluctuations, that are not considered in terms of deviations from the trend component. The reduced form is, as for TpCM, an ARIMA(2,1,3) model, but the implications are quite different: TpCM implies that the spectral density of the first differences is a minimum at zero, whereas this restriction is not enforced by the CTM, which allows the innovations to be more persistent than a random walk.

5. Cyclical trend model (CTM_2):

$$y_t = \mu_t + \epsilon_t, \quad \mu_{t+1} = \mu_t + \psi_t.$$

This is an alternative, more parsimonious, specification of the cyclical trend model, with $\sigma^2_\eta = 0$. With respect to LLM, the trend disturbances are serially correlated (a stationary ARMA(2,1) process), which allows the clustering of positive and negative disturbances during upswings and downswings.

6. Cyclical trend model with seasonal feature ($\mathsf{CTM}_2\mathsf{s}$): the model differs from the previous specification (CTM_2) for the irregular component, which is formulated as the seasonal ARMA(1,1) process: $(1 - \Phi L^{12})\epsilon_t = (1 + \Theta L^{12})\xi_t$. This is introduced in order to account for a nuisance feature that is a likely consequence of overadjustment produced by X-11-ARIMA (see also section 5). As a matter of fact, the changes in the seasonally adjusted unemployment rate display significant negative serial correlation at seasonal lags.

7. Autoregressive trend model (ARTM):

$$y_t = \mu_t + \epsilon_t, \quad \mu_{t+1} = \mu_t + \psi_t, \quad \psi_{t+1} = \rho\psi_t + \kappa_t$$

This is the same as CTM_2 with $\lambda_c = 0$; the recursion for ψ^*_t becomes redundant and is dropped; the model is also referred to as a damped slope trend model.

The eventual forecast function is horizontal except for LLTM, for which it is a straight line. Some of the representations are nested and some are not (for instance TpCM and CTM); model selection and hypothesis testing constitute non standard issues and the reader is referred to Harvey (1989, ch. 5) and Harvey (2000) for these topics. As stated before the main interest of the paper is selecting a forecasting model according to its post-sample performance.

2.2 Estimation

The reference framework for statistical treatment is the linear Gaussian state space representation and likelihood inference is carried out by means of the Kalman filter (see appendix A.1). Parameters estimates for the full sample (1948.1-2000.12), along with some basic diagnostics and goodness of fit statistics computed on the KF standardised innovations, are reported in table 1[1].

[Table 1 about here]

[1] All the computations were performed using the library of state space functions SsfPack 2.3 by Koopman *et al.* (1999), linked to the object oriented matrix programming language Ox 2.1 of Doornik (1998). Computer programmes are available from the author upon request.

For the LLM the irregular variance was estimated to be zero, and therefore the model reduces to a simple random walk and provides naïve "no change" forecasts; in the rolling forecast experiment of the next section it will thus be taken as the benchmark against which the performance of the other models will be assessed. It suffers from various forms of misspecification as highlighted by the Box-Ljung portmanteau statistics, $Q(12), Q(24)$, and by the normality test recommended by Doornik and Hansen (1994) based on Bowman and Shenton (1975).

The additive cycle estimated by TpCM has a period of about 4 years and a half (54 months) and is characterised by a value of the damping factor which is close to one. Among the various specifications of the cyclical trend model, the usual information criteria prefer $\mathsf{CTM}_2\mathsf{s}$. Also, the Box-Ljung statistics using 12 and 24 autocorrelations show that the model quite effectively picks up the seasonal feature in the data. The issue that will be investigated in the next section is whether this improved in-sample fit is associated to an increased forecast accuracy. All models suffer from serious departure from normality[2].

2.3 Impulse response function

The impulse response function (IRF) is a standard tool in illustrating the dynamic properties of the time series models we have entertained; namely, it examines the effect of an innovation occurring at time t, $\nu_t = y_t - \mathsf{E}[y_t|Y_{t-1}]$, on the future pattern, $y_{t+l}, l \geq 0$, by looking at the sequence of dynamic multipliers:

$$\frac{\partial y_{t+l}}{\partial \nu_t}, \quad l = 0, 1, \ldots.$$

For the linear and time invariant models of this section the IRF is a function of l alone (and the model parameters), and it is readily computed from the steady state KF (see appendix A.1), being provided by the sequence $\{1, ZK, ZTK, ZT^2K, \ldots, ZT^{l-1}K, \ldots\}$.

Figure 2 shows the IRF of a selection of models; for LLMT it diverges to infinity at a linear rate. It should be noticed that for TpCM the IRF describes a damped wave converging rather slowly to a value (persistence) below one; this is a consequence of the orthogonal trend-cycle decomposition. Cyclical trend models imply high persistence, with the ARTM yielding the highest profile. The figure also reports the IRF pattern for the ARIMA model considered by Montgomery *et al.* (1998); this will be discussed in section 3.

[Figure 2 about here]

3 Comparitive Performance of Rolling Forecasts for Linear Models

We use a rolling forecast experiment as an out-of-sample test of forecast accuracy. As pointed out in Tashman (2000) a most crucial issue is how to split the series between the

[2]Two other models were entertained: the first is the LLTM with fixed slope, that is (2) with $\beta_t = \beta$; the second is the CTM_2 with a fixed slope, that is with trend component $\mu_{t+1} = \mu_t + \beta + \psi_t$. In both case the estimated slope was not significantly different from zero and so the models reduce respectively to LLM and CTM_2.

fit and the test period. Assuming that our interest lies in short run forecasting so that the greatest lead time is 12 months, we have decided to use the sample period 1948.1-1979.12 as the fit period and to leave the last 21 years of monthly observations for evaluating and comparing the out-of-sample forecast performance of the various alternative models. The test period includes two full cycles in the unemployment rate and the long lasting expansion of the last decade, so the different phases of the business cycle are well represented. Of course, we will be mostly concerned with the performance at the end of the sample, but we are also interested in assessing the sensitivity of the results to the state of the business cycle.

Hence, starting from January 1980, each of the models of the previous section is estimated and 1 to 12 step-ahead forecasts are computed. Then, the forecast origin is moved one step forward and the process is repeated until the end of sample is reached; notice that we reestimate the model each time the forecast origin is updated, and so parameter estimation will contribute as an additional source of forecast variability. The experiment provides in total 252 one step ahead forecasts and 240 12-step-ahead forecasts.

As hinted before, our benchmark model will be LLM; as the irregular variance was always estimated equal to zero, y_t is simply a random walk and thus its forecasts will be used as a reference for comparison. Furthermore, we also included in our comparison the monthly ARIMA model considered by Montgomery *et al.* (1998); the latter is a seasonal model with orders $(2,0,1)_\times(1,0,1)_{12}$ and its estimation with the full sample gave the following results:

$$(1-\underset{(.05)}{1.88}L+\underset{(.05)}{0.88}L^2)(1-\underset{(.07)}{0.53}L^{12})y_t=\underset{(.003)}{0.001}+(1-\underset{(.07)}{0.74}L)(1-\underset{(.05)}{0.80}L^{12})\xi_t$$

where the sample variance of the innovation series ξ_t is 0.0364. This is virtually identical to that reported in the referenced paper, apart from the constant term, which is not significant; notice that the non seasonal AR polynomial contains a unit root. The IRF is reported in figure 2; its pattern suggests that the model is capturing not solely the residual seasonal effect, but it is boosting the impact of innovations at short lead times; that the model is picking up something else is suggested by Montgomerty *et al.* in their comment to the quarterly version at the beginning of section 3.1 of their paper. The implied persistence is 0.92, but the IRF makes a long and abrupt swing before converging to it.

Table 2 reports a few basic statistics upon which forecasting accuracy will be assessed. Denoting the l-step-ahead forecast for model (j) by $\tilde{y}^{(j)}_{t+l|t}$, we present the average of the forecast errors (mean error), $y_{t+l}-\tilde{y}^{(j)}_{t+l|t}$; the symmetric mean absolute percentage error (sMAPE), given by the average of $100|y_{t+l}-\tilde{y}^{(j)}_{t+l|t}|/[0.5(y_{t+l}+\tilde{y}^{(j)}_{t+l|t})]$, a measure which is featured in the M3-Competition (Makridakis and Hibon, 2000) and aims at providing a symmetric treatment of underforecasts and overforecasts; the median relative absolute error (mRAE, see Armstrong and Collopy, 1992), a robust comparative measure of performance computing the median of the distribution of the ratios $|y_{t+l}-\tilde{y}^{(j)}_{t+l|t}|/|y_{t+l}-\tilde{y}^{(1)}_{t+l|t}|$, where (1) indexes the benchmark model; finally, we report the mean square forecast error (MSFE). For simplicity these statistics are reported only for 1, 3 (one quarter), 6 (two quarters), 9 (three quarters), and 12 (one year) step ahead.

[Table 2 about here]

In terms of MSFE the greatest accuracy is provided by the ARIMA model for horizons up to two quarters; then its performance rapidly deteriorates and the best forecasting model turns out to be ARTM. The latter is ranked best in terms of mRAE and sMAPE, immediately followed by CTM_2. Therefore, on aggregate, a simple structural time series model, such as ARTM, is capable of outperforming the benchmark ARIMA model in Montgomery *al.* (1998).

Among the cyclical trend models, having $\sigma^2_\eta > 0$, as in the original CTM formulation, yields a poorer forecasting model; the performances of CTM_2 and $\mathsf{CTM}_2\mathsf{s}$ are almost indistinguishable, but the former is preferred in terms of mRAE and is more parsimonious. This confirms that modelling the negative serial correlation at seasonal lags improves the accuracy of the forecasts only at very short lead times. We also report in passing that this is not a stable feature of the series: for instance, the autocorrelation at lag 12 of Δy_t estimated using the full sample is -0.18 (with asymptotic standard error 0.04); in the test period the estimate resulted -0.03, which is not significantly different from zero.

A piece of evidence that does not emerge from the table concerns the role of the irregular term; when the forecasting experiment is conducted imposing $\sigma^2_\epsilon = 0$ (no irregular) on the cyclical and autoregressive trend models, this produces worse forecasts.

The empirical evidence speaks strongly against TpCM, which is outperformed by the naïve model at all lead times. The table also tells that the forecasts are negatively biased; the largest biases are found for the ARIMA model. LLTM presents a distinctive trade-off between bias and forecast error variance; as the forecast function is highly adaptive, assigning a large weight to the most recent observations, the bias is very small, but on the other hand the forecasts' variability is quite high (letting the changes in the trend be non stationary inflates overly the variance).

The upper right panel of figure 3 displays the MSFE of the linear models relative to that of the benchmark model for every lead time. Among the structural models, the greatest accuracy is provided by ARTM; yielding a proportional reduction in MSFE above 20% for lead times greater than two months ahead. $\mathsf{CTM}_2\mathsf{s}$ improves over ARTM solely for $l = 1, 2$, and overall CTM_2 can be ranked as second best. In conclusion, our preferred linear structural model is ARTM; this will also represent the benchmark against which non linear models, dealt with in section 6, are evaluated.

In order to assess the cyclical sensitivity of its forecasting performance, we present in the second panel the one-step-ahead relative MSFE for a rolling window of 60 consecutive observations. The plot reveals that the forecasting accuracy deteriorates in periods of slow decline in the unemployment rate and scores better in periods of fastly rising and declining unemployment rates, during which the persistence of the innovations should be greater. Similar behaviour characterises the relative MSFE of multistep forecasts. We mention in closing this section that the end of sample performance of the ARIMA model is worse than the naïve model, with a 10% increase in MSFE with respect to the benchmark.

4 Leave-k-out Diagnostics and Nonlinearity

Departure from normality was detected for the innovations of the ARTM; the breakdown into the contribution of the two terms skewness and kurtosis reveal that the latter is mostly responsible for the recorded high value.

Asymmetric behaviour with respect to the business cycle phase would show up in a skewed distribution for the auxiliary residuals (see Harvey and Koopman, 1992). These are estimators of the disturbances associated with the components, conditional on the entire information set, e.g. $\mathsf{E}(\kappa_t|Y_T)$, and are computed from the output of the smoothing filter, reviewed in appendix A.2, as indicated in Koopman (1992). Unlike the innovations, they are autocorrelated even if the model is correctly specified, and Harvey and Koopman (1992) show how they can be employed to form appropriate tests of normality, correcting for serial correlation. The skewness test conducted on $\mathsf{E}[\kappa_t|Y_T]$ is highly significant; significant kurtosis was also detected.

Additional evidence can be gathered using leave-k-out diagnostics, arising from the deletion of groups of observations (see e.g. Bruce and Martin, 1989); these aim at spotting patches of observations that are not adequately fitted by a linear model; for linear state space models they are easily computed using the KF based algorithm proposed in Proietti (2000) and outlined in appendix A.2. The case for computing them is quite strong in our application since dynamic asymmetries affect clusters of consecutive observations.

Figure 4 is a plot of the indicator variable $I(\tau_{(I)} > c)$, $\tau_{(I)}$ being the statistic (11) in appendix A.2 and c the 5% critical value of the reference distribution $F(k, T-k)$. The dimension of the set of deleted observations is 1, 5, and 11; when k=1 the statistic is a test for the presence of an additive outlier at time t; for $k = 5, 11$, the indicator values refer to the midpoint of the deletion interval.

[Figure 4 about here]

As concerns leave-1-out diagnostics (upper panel), quite a few isolated observations are flagged in the first years of the sample period up to 1961, and in the years 1974-1985; there is some clustering around turning points (especially noticeable around Oct. 1949), in periods of fast rising unemployment (end of 1953, 1974). We suspect that some masking has taken place, in that the effect of adjacent observations could be clouded, and that joint deletion of consecutive observations can bring to the surface the masked outliers.

As a matter of fact, the central and lower panels point out more clearly that violation of linearity and Gaussianity arises in economic downturns, that is in periods of rapidly increasing unemployment, and around turning points. The other relevant piece of evidence is these departures are concentrated in the fit period; only the beginning of the test period, i.e. the slowdown at the beginning of the eighties appears problematic. We might therefore anticipate that the gains of non linear models need not to be great in this situation.

It must be acknowledged that in interpreting these plots a balance has to made between unmasking (which is likely to have taken place during the downturns taking place in the second half of 1974 and in the first half of 1980) and smearing (probably occurring in 1949 and 1959) the effect of outliers on adjacent time points: moreover, use of critical values from the F distribution would be correct only if we knew the exact timing of the outlying effect. Nevertheless we find these plots quite informative as a descriptive device for detecting groups of suspect observations.

5 Transformations and Seasonal Adjustment

Linear Gaussian models of the untransformed unemployment rates are not strictly adequate, since the series is bounded between 0 and 100. Theoretically, this implies a nonzero probability, however small, of predicting negative unemployment rates; moreover, some nonlinear features can be ascribed to the measurement scale itself, which imposes floors and ceilings.

Wallis (1987) advocated for this case the logistic transformation, $y_t^* = \ln[y_t/(100-y_t)]$, and proposed an approximate method for forecasting y_t; among the papers quoted in the introduction, only Koop and Potter (1999) adopt it, whereas the most common strategy (especially in forecast accuracy experiments) is to leave the data untransformed or to use logarithms.

When applied to our series the logistic trasformation has basically the effect of amplifying the troughs in 1951-53, 1968-69 (corresponding to the smallest rates observed) and downscaling the data around the peak in 1982 (maximum rate), leaving the last decade practically unaffected. Furthermore, ARTM applied to y_t^* yielded very close parameter estimates leaving persistence and the signal-noise ratio unaffected, reduced kurtosis and skewness coefficients for the innovations, which remain highly significant (for the AR disturbances these coefficients are virtually the same), and increased Box-Ljung statistics.

In conclusion, it appears that the nonlinear and nongaussian effects are more a dynamic feature of the series than only a scale property of the data and we feel that the transformation issue is unlikely to affect our main findings.

Seasonal adjustment has left a clear mark on the series, although we have argued that extending the model to account for seasonal features does not add much in terms of forecast accuracy in our test period. The availability of unadjusted data (plotted in figure 1) raises two issues: firstly, does seasonal adjustment affect the stylised facts concerning U.S. unemployment rates (nonlinearity, propagation of innovations)? Secondly, why do the seasonally adjusted data display seasonal features?

To address the first we set off fitting the basic structural model with a cyclical trend (BsmCT):

$$y_t = \mu_t + \gamma_t + \epsilon_t,$$

where $\mu_{t+1} = \mu_t + \psi_t$, ψ_t is given by (3), and the seasonal component, γ_t, is modelled by the sum of six nonstationary cycles defined at the seasonal frequencies $\lambda_j = 2\pi j/12$, $j = 1, \dots, 6$:

$$\gamma_t = \sum_{j=1}^{6} \gamma_{jt}, \qquad \begin{bmatrix} \gamma_{j,t+1} \\ \gamma_{j,t+1}^* \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{jt} \\ \gamma_{jt}^* \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega_{jt}^* \end{bmatrix},$$

for $j = 1, \dots, 5$ and $\gamma_{6,t+1} = -\gamma_{6,t} + \omega_{6,t}$. The disturbances ω_{jt} and ω_{jt}^* are assumed to be normally and independently distributed with common variance $\sigma^2_{\omega_j}$.

Information criteria led to constrain $\sigma^2_{\omega_j}$ to be constant across j for $j = 2, \dots, 6$. The corresponding parameter estimates using the full sample were: $\sigma^2_\epsilon = 0.0091$; $\sigma^2_\kappa = 0.0201$; $\rho = 0.62, \lambda_c = 0.01$ $\sigma^2_{\omega 1} = 12.64 \times 10^{-5}$; $\sigma^2_{\omega j} = 1.78 \times 10^{-5}$. The Box-Ljung portmanteau statistics resulted $Q(12) = 13.65$ and $Q(24) = 28.50$, which are not significant, and the DH normality test took the value 44.08, which is significant, but quite small, in comparison to the value reported in table 1 for the models adapted to the seasonally adjusted series. Hence, we find less evidence for departure from the maintained assumptions.

Given the estimated $\lambda_c \simeq 0$, the model for the trend reduces to an ARIMA(1,1,0), with autoregressive parameter 0.62. Correspondingly, the persistence of the innovations is reduced with respect to ARTM. That seasonal adjustment increases persistence is documented in Jaeger and Kunst (1990) and Pesaran and Samiei (1991). Additionally, that seasonal adjustment may induce nonlinearity is documented in Ghysels *et al.* (1996).

To answer the second question, it should be noticed that the variance of the seasonal cycle defined at the fundamental frequency is about seven times as great as that associated to the remaining frequencies. One likely occurrence is that the seasonal adjustment procedure is unable to account for this kind of heteroscedastity, underadjusting the fundamental frequency and overadjusting the others. This point, however, has some caveats and would require further investigation, since the BLS does not perform the adjustment directly on the unemployment rate series, but proceeds to adjust the single elementary series of unemployment and civilian labour force disaggregated for population groups, and then computes the seasonally adjusted rate by taking the total seasonally adjusted unemployment level as a percent of the total seasonally adjusted civilian labour force.

6 Nonlinear Cyclical Trend Models

In this section we consider four nonlinear trend models, derived from ARTM and CTM_2, that can account for dynamic asymmetries in unemployment rates. These belong to the framework of smooth transition structural time series models (see Proietti, 1999), according to which we let the fundamental parameters vary according to the state of the system, as described by an appropriate transition variable.

6.1 Specification

In the ARTM case, for which the trend has the representation $\mu_{t+1} = \mu_t + \psi_t$, with $\psi_{t+1} = \rho\psi_t + \kappa_t$, it appears a sensible option to allow the autoregressive coefficient ρ and the variance of the disturbances κ_t to evolve over time; thus for a given measure of the state of the economy, S_t, bounded between 0 and 1, we set

$$\rho_t = \rho_0(1 - S_t) + \rho_1 S_t, \;\; \sigma^2_{\kappa t} = \sigma^2_{\kappa 0}(1 - S_t) + \sigma^2_{\kappa 1} S_t. \tag{4}$$

The regime S_t is a function of a transition variable, z_t, which is observable at time t; its choice will be discussed in a moment. We will adopt a logistic transition mechanism (see e.g. Granger and Teräsvirta, 1993), defining $S_t = [1 + \exp(-\tau(z_t - c))]^{-1}$ where τ is a smoothness parameter determining the speed of the transition, and c is a threshold. The resulting model is labelled ARTMSt.

In the CTM_2 case it makes sense to allow variation also in the frequency of the cycle, and we will write

$$\lambda_{ct} = \lambda_{c0}(1 - S_t) + \lambda_{c1} S_t, \tag{5}$$

and denote the corresponding model $\mathsf{CTM}_2\mathsf{St}$.

Although including a drift term in the trend was not significant globally, it might be the case that this plays a role locally. so we will also consider the trend model:

$$\mu_{t+1} = \mu_t + \beta_t + \psi_t, \;\; \beta_t = \beta_0(1 - S_t) + \beta_1 S_t,$$

where the AR and cycle parameters follow (4) and (5). This will give rise respectively to the ARTMStD and $\mathsf{CTM_2StD}$ models.

The transition variable is usually based upon the differences $\Delta_r y_t = y_t - y_{t-r}$, an unweighted sum of current and past one-step changes in the unemployment rates; this has two drawbacks: the transition is not necessarily smooth unless r is large; the differencing filter induces a phase shift, affecting the timing of turning points. Another possibility is to define z_t in terms of the underlying trend, e.g. $z_t = \mathsf{E}[\Delta_r \mu_t | Y_{t-1}] = \mathsf{E}[\psi_t + \cdots + \psi_{t-r+1} | Y_{t-1}]$; however, on the one hand we do not expect great gains for a relatively smooth series such as the one considered here, and, on the other, inference would be complicated computationally by the need to support the KF equations with a fixed interval smoother in order to construct the transition variable; if the analysis were conducted on seasonally unadjusted data, this would nevertheless be our preferred strategy.

Our definition of z_t builds upon the truncated version of the current-depth-of-recession variable, proposed originally Beaudry and Koop (1993) for U.S. GNP, exploited by Parker and Rothman (1997) in a similar exercise referred to forecasting the U.S. unemployment rate. The latter include (lagged values of) $\min_{j=0,\dots,r} y_{t-j} - y_t$ as an explanatory variable in an ARMA framework, measuring the distance of the current rate from its historical local minimum; on the other hand $\max_{j=0,\dots,r} y_{t-j} - y_t$ would measure the depth of the decrease in unemployment rates (an expansion in the economy).

We propose to combine the two into the variable:

$$z_t = 2y_t - \min_{j=0,\dots,r} y_{t-j} - \max_{j=0,\dots,r} y_{t-j}.$$

For monotonic patterns z_t is coincident with $\Delta_r y_t$, since either $y_t = \min_{j=0,\dots,r} y_{t-j}$ or $y_t = \max_{j=0,\dots,r} y_{t-j}$. When y_{t-1} is a turning point, $z_t = \Delta_r y_t + \Delta y_t = \Delta(1 + S_r(L))y_t$, so for instance after a peak has taken place, z_t will be smaller than $\Delta_r y_t$, as we add the negative change from time $t-1$ to time t; this results in an earlier recognition of the change in regime and lessens the problem of phase shifts. The following table illustrates the comparison for the case $r = 5$ with reference to the pattern of observations around the turning point at Dec. 1982 ($y_t = 10.8$):

y_t	9.6	9.8	9.8	10.1	10.4	10.7	$\boxed{10.8}$	10.4	10.4
$\Delta_5 y_t$	1.0	1.0	0.8	0.8	1.1	1.1	1.0	0.6	0.3
z_t	1.0	1.0	0.8	0.8	1.1	1.1	1.0	0.1	-0.1

The plot of z_t with $r = 5$ for the full sample is available at the bottom of the upper panel of figure 5.

When z_t is coupled with the logistic transition mechanism it yields the regime variable $S_t = [1 + \exp(-\tau(z_t - c))]^{-1}$; this framework is suitable for modelling business cycle asymmetries such as those arising when unemployment rates are characterised by steep increases during recessions (S_t close or equal to one) and slower declines during expansions (S_t close or equal to zero).

6.2 Estimation and testing

A score test of linearity against smooth transition alternatives can be carried out for unobserved components models using the same idea proposed by Luukkonen *et al.* (1988) to circumvent the lack of identifiability under the alternative, which amounts to replacing S_t by a first order Taylor approximation around $\tau = 0$ (see Proietti, 1999). As suggested by Teräsvirta (1994), r is chosen by computing the score test for a set of values and selecting the one yielding the smallest p-value. When applied to our data the test is highly significant and suggests $r = 5$; it must be stressed that, contrary to expectations, the likelihood is also very informative on the choice of this parameter, displaying a distinctive peak at $r = 5$.

As far as estimation is concerned, the resulting model is conditionally Gaussian (see Harvey, 1989, sec. 3.7.1., and Lipster and Shiryaev, 1978, ch. 11), since the system matrices T_t, d_t, H_t, in the transition equation (see appendix A.1), depend solely on the information available at time t; likelihood inference is thus made via the prediction error decomposition. Parameter estimates, along with standard errors and diagnostics are reported in table 3.

[Table 3 about here]

The variance of the AR and cycle disturbances are much higher in an economic downturn; correspondingly the damping parameter is significantly lower for the models including a drift term ($\mathsf{ARTMStD}$ and $\mathsf{CTM_2StD}$); for the latter the drift term is positive and significantly different from zero. Moreover, for the cyclical trend models the estimated frequency is zero.

Thus, the parameter estimates associated with the different regimes imply that periods of increasing unemployment are characterised by greater uncertainty and lower persistence, although the trend disturbances are propagated with the addition of a positive drift. This provides only a very preliminary description of the models' dynamic properties, hiding some of essential features, as will be argued later in section 8, where we provide a more informative characterisation by means of the generalised impulse response function. Notice that the in-sample goodness of fit statistics are generally better than the linear case and that the normality statistic is drastically reduced.

The estimated transition function, S_t, for $\mathsf{CTM_2StD}$ and the transition variable z_t with $r = 5$ are represented in figure 5.

[Figure 5 about here]

7 Forecasting with Nonlinear Structural Models

One-step-ahead forecasts are immediately available from the KF output at the end of the fit period; multistep forecasts, conditional on the estimated parameters, are generated by the Monte Carlo method (see Granger and Terasvirta, 1993, ch. 8) as:

$$\tilde{y}_{t+l|t} = \frac{1}{M}\sum_{i=1}^{M}\tilde{y}^{(i)}_{t+l|t},$$

which, for $i = 1, \ldots, M$, requires the following steps:

$$
\begin{array}{lll}
1. & \text{draw} & y_{t+1}^{(i)} \sim f(y_{t+1}|Y_t) \\
2. & \text{draw} & y_{t+2}^{(i)} \sim f(y_{t+2}|Y_t, y_{t+1}^{(i)}) \\
\vdots & \vdots & \vdots \\
l-1. & \text{draw} & y_{t+l-1}^{(i)} \sim f(y_{t+l-1}|Y_t, y_{t+1}^{(i)}, \dots, y_{t+1}^{(i)}) \\
l. & \text{evaluate} & \tilde{y}_{t+l|t}^{(i)} = \mathsf{E}(y_{t+l}|Y_t, y_{t+1}^{(i)}, \dots, y_{t+l-1}^{(i)})
\end{array}
$$

These steps are easily carried out with the support of the Kalman filter. The first draw is made from the normal distribution $\mathsf{N}(\tilde{y}_{t+1|t}, \sigma^2 F_{T+1})$, where the moments are obtained from the last run of the KF within the sample; then, conditional on $z_{t+1}^{(i)}$ and $S_{t+1}^{(i)}$, one-step-ahead forecasts of the states and the observations are computed along with their covariance matrices, which are in turn used to compute the next draw, $y_{t+2}^{(i)}$, and so forth.

The predictive distribution, $f(y_{t+l}|Y_t)$, can be estimated by the method of composition (see e.g. Tanner, 1996, sec. 3.3), giving:

$$
\hat{f}(y_{t+l}|Y_t) = \frac{1}{M} \sum_{i=1}^{M} f(y_{t+l}|Y_t, y_{t+1}^{(i)}, \dots, y_{t+l-1}^{(i)}),
$$

where the densities on the right hand side are normal with mean $Z_{t+l} a_{t+l}^{(i)}$ and variances $\sigma^2 F_{t+l}$ delivered by the KF.

The rolling forecast exercise is repeated for the four nonlinear models using $M = 1000$; the results are reported in table 4, which for convenience reproduces those obtained for the linear benchmark (ARTM).

[Table 4 about here]

The evidence is rather mixed; in terms of MSFE the linear benchmark is clearly outperformed only by $\mathsf{CTM_2StD}$, especially at short horizons; the best performance in terms of mRAE is provided by ARTMSt, instead. The latter and $\mathsf{CTM_2St}$ tend to produce positively biased forecasts and their overall performance is quite similar, as expected. Somewhat unexpected is the poor performance of ARTMStD, compared to $\mathsf{CTM_2StD}$; as a matter of fact, the latter proves rather unstable at the beginning of the fit period, yielding negative estimates of ρ_1. This results in more volatile forecasts, especially at the beginning of the eighties, which are responsible for its poor comparative performance in terms of MSFE.

The last panel of the table reports the p-values of the Meese-Regoff test of comparative forecast accuracy (Meese-Regoff, 1988, see also Diebold and Mariano, 1995, and Clements and Hendry, 1999, ch. 13, for further discussion). In brief, its rationale is the following: the null hypothesis of equal forecast accuracy (in the MSFE sense) implies a covariance of zero between the sum (s_t) and the difference (d_t) of the forecast errors produced by the maintained and the benchmark models; under the null the estimated covariance is asymptotically normally distributed with zero mean and standard error that depends on the auto and cross covariances of s_t and d_t; the latter can be consistently estimated by e.g. a Bartlett window and used to build a statistic for $H_0 : \mathsf{Cov}(s_t, d_t) = 0$ with standard normal distribution. The results clearly indicate that only $\mathsf{CTM_2StD}$ shows significant improvements up until 6 step ahead.

The right panel of figure 6 plots the MSFE of the four nonlinear forecasting models relative to that of the benchmark against lead time; the percentage reduction achieved by $\mathsf{CTM_2StD}$ is around 8% for short leads and converges to zero as l increases. The panel on the left is a plot of the 3-step ahead relative MSFE for this model calculated over spans of 5 years, with the original series reproduced in the background; the plot is meant to provide some insight on how the comparative forecast accuracy of this model varies with time. It is noticeable that $\mathsf{CTM_2StD}$ yields the most significant improvements towards the end of the test period (20% MSFE reduction for the span including the last 5 years in the sample), during a phase of slowly declining unemployment rates. It is also remarkable that the performance is rather weak in the early nineties, in connection to the last upswing and successive decline, which corresponds to the period when the benchmark yields the best results.

[Figure 6 about here]

To corroborate the finding that $\mathsf{CTM_2StD}$ provides better forecasts at the end of the sample period, figure 7 presents the l-step-ahead ($l = 1, \ldots, 12$) forecasts from the origin 1999.12 using $M = 5000$ replicates, along with the 95% highest density region, $HDR_{0.05} = \{y_{t+l|t} : f(y_{t+l}|Y_t) \geq f_{0.05}\}$, where $f_{0.05}$ is chosen such that $p(y_{t+l} \in HDR_{0.05}|Y_t) = 0.95$. The latter is the smallest of all possible 95% forecast regions and is particularly helpful in revealing asymmetry and multimodality in the forecast densities; see Hindman (1996) for more details. Even though at the forecast origin the system is in an expansionary state, there is a possibility that it enters and gets stuck in a recessionary pattern; this shows up in the multistep forecast densities as a secondary mode. The figure also reports the upper and lower 95% confidence bounds for the ARTM forecasts, showing that the nonlinear model has reduced forecast uncertainty.

[Figure 7 about here]

8 Generalised Impulse Response Function

The notion of a generalised impulse response function (GIRF) provides further insight on the dynamic properties of $\mathsf{CTM_2StD}$. The underlying idea is that in a nonlinear system the impact of an innovation occurring at time t on the future path of the process depends crucially upon the history of the process and the size of the innovation.

According to the definition by Koop *et al.* (1996),

$$\text{GIRF}(l, \nu_t, Y_{t-1}) = \mathsf{E}[y_{t+l}|\nu_t, Y_{t-1}] - \mathsf{E}[y_{t+l}|Y_{t-1}],$$

where the notation stresses the dependence on the lead time l, the size of the innovation and the history of the process. The GIRF is evaluated by the same stochastic simulation techniques illustrated in the previous section; basically, for a given lead time, innovation ν_t, and history, we evaluate the two components $\mathsf{E}[y_{t+l}|Y_{t-1}]$ and $\mathsf{E}[y_{t+l}|\nu_t, Y_{t-1}]$, averaging out intermediate innovations and subtracting out the results; it is crucial that in the simulations we use common random numbers for drawing the intermediate shocks.

The GIRF was computed for the empirical innovations $y_t - \mathsf{E}(y_t|Y_{t-1})$ resulting from $\mathsf{CTM_2StD}$, and using the observed history for l equal to a large number, $l = 60$; $\text{GIRF}(60, \nu_t, Y_{t-1})$

can be taken as a measure of the long run impact, or persistence, of the observed shock. We notice in passing that the cumulation of the persistence values is an approximation of the Beveridge-Nelson (1981) trend in the series.

Figure 8 is a plot of GIRF$(60, \nu_t, Y_{t-1})$ versus the innovation ν_t conditional on the state of the economy as described by the variable S_t: during a period of increasing unemployment (left panel) innovations tend to be positive and highly persistent; there is also a small subset of very persistent large innovations. In periods of decreasing unemployment (right panel) we find a cluster of highly persistent shocks, which occur after a turning point, when unemployment rates are more rapidly declining. Another group of small shocks displays low persistence and is gathered around the unit line; these occur in periods of low decreasing unemployment. This fundamental asymmetry in the dynamics of the system could not be anticipated by a simple characterisation based on the parameter estimates in the two regimes, but is a quite natural consequence of nonlinearity and of the basic fact that the response of the system depends on its position at the time of a shock and the size of the latter.

[Figure 8 about here]

These dynamic features also cast some light on the interpretation of the forecasting performance of the $\mathsf{CTM_2StD}$ model. In this case we would be interested in estimating the GIRF at lead times not greater than 12; for this purpose we set $l = 3$, define positive and negative shock equal to twice the standard deviation of the innovations, and compute GIRF$(3, \pm 2\sigma, Y_{t-1})$ for all t starting from 1949; as before we condition on the observed histories Y_{t-1}. The upper panel of figure 9 plots the normalised GIRF for both positive and negative shocks, where the normalisation is made dividing for the initial shock. A noticeable feature is that positive shocks have a larger impact in a recession, a finding which is consistent with that of Koop and Potter (1999); both positive and negative shocks have larger impact immediately after a peak, when the system moves out of a recession. Moreover, during periods of slowly declining unemployment rates the impact is damped out and it is less than the original size.

A quick assessment on some of this properties can also be made from the innovation form of the model provided by the Kalman filter (see appendix A.1): writing $y_{t+l} = Z_{t+l}a_{t+l} + \nu_{t+l}$, $a_{t+l+1} = T_{t+l}a_{t+l} + d_{t+l} + K_{t+l}\nu_{t+l}$ and substituting recursively from the second equation, we have

$$\frac{\partial y_{t+l}}{\partial \nu_t} = Z_{t+l} \prod_{j=1}^{l+1} T_{t+l-j} K_t; \tag{6}$$

this represent a straightforward empirical measure of the impact of an unit innovation on the observed level of the series l step ahead, conditional also on the future state of the system. Naturally, the above discussion implies that (6) is not a GIRF; nevertheless, the plot of its values for $l = 3$, given in the bottom panel of figure 9, points out the same stylised facts: namely that the impact of shocks varies greatly in expansions, being far greater after a turning point has taken place and lately declining quite rapidly when unemployment decreases more slowly; then it jumps to a higher stable value as a downturn is entered. Finally, the model tend to behave like a random walk with a positive drift in recessions.

[Figure 9 about here]

The fact that $\mathsf{CTM_2StD}$ produces better forecasts at the end of the sample period can be ascribed to this feature of the model, which in periods of slowly declining rates has a reduced persistence (recall that ARTM, implying large persistence, did not out perform better than the random walk model at the end of the sample, see the second panel of figure 3).

9 Conclusive Remarks

This paper has investigated the out-of-sample performance of linear and nonlinear structural time series models of the U.S. unemployment rate by means of an extensive rolling forecast experiment using a test period made up of the last two decades.

One conclusion is that our study corroborates the hysteresis hypothesis, as we found that linear models characterised by higher persistence perform significantly better. Persistence is not a stable property, however, and a nonlinear specification, implying that this feature changes over the phases of the business cycle, outperforms the best linear models at least at short lead times. In particular, the model produces more accurate forecasts in periods of slowly decreasing unemployment rate, which is the regime prevailing at the end of the sample period.

The accuracy gains, though significant, are not overwhelmingly great. This could be anticipated by simple diagnostics immediately available from the linear fit, revealing that nonlinear and nongaussian features are more prominent in the fit period (1948-1979).

The general conclusion is that structural time series models prove a very useful forecasting tool: they are parsimonious and are formulated in terms of core parameters and components with direct interpretation, which renders them easily extensible to account for nonlinearity and for dealing with seasonally unadjusted series.

References

Anderson, B.D.O., and Moore J.B. (1979), *Optimal Filtering*, Prentice Hall, Englewood Cliffs NJ.

Armstrong, J.S., and Collopy, F. (1992), "Error Measures for Generalizing about Forecasting Methods: Empirical Comparison", *International Journal of Forecasting*, 8, 69-80.

Beaudry, P., and Koop, G. (1993), "Do recessions permanently change output?". *Journal of Monetary Economics*, 31, 149-163.

Beveridge, S., Nelson, C.R. (1981), "A new Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of the 'Business Cycle'", *Journal of Monetary Economics*, 7, 151-174.

Bowman, K.O., and Shenton, L.R. (1975), "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2", *Biometrika*, 62, 243-250.

Clements, M.P., and Hendry D.F. (1998), *Forecasting Economic Time Series*, Cambridge University Press, Cambridge, UK.

De Jong, P. (1988), "A Cross-validation Filter for Time Series Models", *Biometrika*, 75, 594-600.

De Jong, P. (1989), "Smoothing and Interpolation with the State Space Model", *Journal of the American Statistical Association*, 84, 1085-1088.

DeLong, B.J., and Summers, L. (1986), "Are Business Cycles Symmetrical?", in R. Gordon (ed.), *The American Business Cycle: Continuity and Change*, NBER and University of Chicago Press, Chicago.

Diebold, F.X., and Mariano, R.S. (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

Doornik, J.A. (1998), *Object-Oriented Matrix Programming Using Ox 2.0*, Timberlake Consultants Press, London.

Doornik, J.A. and Hansen H. (1994), 'An omnibus test for univariate and multivariate normality', Discussion paper, Nuffield College, Oxford.

Ghysels, E., Granger, C.W.J., and Siklos, P.L. (1996), "Is Seasonal Adjustment a Linear or Nonlinear Data-Filtering Process?" *Journal of Business and Economic Statistics*, 14, 374-386.

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

Harvey, A.C. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A.C. (2000), "Testing in Unobserved Components Models", *Working Paper*, University of Cambridge, Faculty of Economics and Politics.

Harvey, A.C., and Koopman, S.J. (1992) "Diagnostic Checking of Unobserved Components Time Series Models", *Journal of Business and Economic Statistics*, 10, 377-389.

Hindman, R.J. (1996), “Highest-density Forecast Regions for Non-linear and Non-normal Time Series Models”, *Journal of Forecasting*, 14, 431-441.

Hodrick R.J., and Prescott, E.C. (1997), “Postwar U.S. Business Cycles: an Empirical Investigation”, *Journal of Money, Credit and Banking*, 29, 1-16.

Jaeger A., and R. Kunst (1990), “Seasonal Adjustment and Measuring Persistence in Output”, *Journal of Applied Econometrics*, 5, 47-58.

Kohn, R., and Ansley, C.F. (1989), “A Fast Algorithm for Signal Extraction, Infuence and Cross-validation in State Space Models”, *Biometrika*, 76, 65-79.

Koop, G., and Potter, S.M. (1999), “Dynamic Asymmetries in U.S. Unemployment” *Journal of Business and Economic Statistics*, 17, 3, 298-312.

Koop, G., Pesaran, M., and Potter, S. (1996), ”Impulse response analysis in nonlinear multivariate models”, *Journal of Econometrics*, 74, 119147.

Koopman, S.J. (1993), “Disturbance Smoother for State Space Models”, *Biometrika*, 80, 117-126.

Koopman, S.J. (1997), “Exact initial Kalman filtering and smoothing for non-stationary time series models”, *Journal of the American Statistical Association*, 92, 1630-1638.

Koopman S.J., Shepard, N., and Doornik, J.A. (1999), “Statistical algorithms for models in state space using SsfPack 2.2”, *Econometrics Journal*, 2, 113-166.

Lipster, R.S., and A.N. Shiryaev (1978): *Statistics of Ramdom Processes II.* New York: Springer.

Luukkonen, R., Saikkonen, P., and T. Teräsvirta (1988): “Testing Linearity against Smooth Transition Autoregressive Models”, *Biometrika*, 75, 491-499.

Makridakis, S., and Hibon, M. (2000), “The M3-Competition: results, conclusions and implications”, *International Journal of Forecasting*, 16, 437-450.

McQueen, G., and Thorley, S. (1993), “Asymmetric Business Cycle Turning Points”, *Journal of Monetary Economics*, 31, 341362.

Meese, R.A., and Rogoff, K. (1988), “What is Real? The Exchange Rate-Interest Differential Relation over the Modern Floating-Rate Period”, *Journal of Finance*, 43, 933-948.

Neftçi, S. N. (1984), “Are Economic Time Series Asymmetric over the Business Cycle?”. *Journal of Political Economy*,92, 307-328.

Parker, R.E., and Rothman, P. (1997), “The Current Depth-of-Recession and Unemployment Rate Forecasts”, *Studies in Nonlinear Dynamics and Econometrics*, 2, 151-158.

Pesaran, M. H. and H. Samiei (1991), ‘Persistence, Seasonality and Trend in the UK Egg Production.’ *Applied Economics*, 23, 479-484.

Proietti, T. (1999), “Characterising asymmetries in business cycles using smooth transition structural time series models”, *Studies in Nonlinear Dynamics and Econometrics*, 3.3, 141-156.

Proietti, T. (2000), “Leave-k-out diagnostics in state space models”, *Discussion paper 74*, Sonderforschungsbereich 373, Quantifikation und Simulation ökonomischer Prozesse, Humboldt-Universität, Berlin.

Rothman, P. (1991), "Further Evidence on the Asymmetric Behaviour of Unemployment Rates over the Business Cycle", *Journal of Macroeconomics*, 13, 291-298.

Rothman, P. (1991), "Forecasting Asymmetric Unemployment Rates", *Review of Economics and Statistics*, 80, 164-168.

Sichel, D. E. (1993), "Business Cycle Asymmetry: a Deeper Look", *Economic Inquiry*, 31, 224-236.

Skalin J., and Teräsvirta, T. (1999), "Modelling Asymmetries and Moving Equilibria in Unemployment Rates", *Working Paper Series in Economics and Finance*, No. 262, Stockholm School of Economics.

Tanner, M.A. (1996). *Tools for Statistical Inference*, 3rd edition, Springer-Verlag.

Tashman, L.J. (2000): "Out-of-sample tests of forecast accuracy: an analysis and review", *International Journal of Forecasting*, 16, 437-450.

Teräsvirta, T. (1994), "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", *Journal of American Statistical Association*, 89, 208-218.

van Dijk, D., Teräsvirta, T., and Franses, P.H. (2000), "Smooth Transition Autoregressive Models - A Survey of Recent Developments", *Working Paper Series in Economics and Finance*, No. 380, Stockholm School of Economics.

A Algorithms

A.1 The Kalman filter

The models considered in the paper admit the state space representation:

$$\begin{array}{rcll} y_t & = & Z_t\alpha_t + G_t\varepsilon_t, & t = 1, 2, \ldots, T, \\ \alpha_{t+1} & = & T_t\alpha_t + d_t + H_t\varepsilon_t, & \end{array} \tag{7}$$

with $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2 I)$ and $\alpha_1 \sim \mathsf{N}(a_1, \sigma^2 P_1)$ independent of $\varepsilon_t, \forall t$. The model is time invariant if the system matrices Z_t, G_t, T_t, d_t, H_t, do not depend on t.

The Kalman filter (KF) (Anderson and Moore, 1979), is a well-known recursive algorithm for computing the minimum mean square estimator of α_t and its mean square error (MSE) matrix conditional on $Y_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$. Defining $a_t = \mathsf{E}(\alpha_t|Y_{t-1})$, $\mathsf{MSE}(a_t) = \sigma^2 P_t = \mathsf{E}[(\alpha_t - a_t)(\alpha_t - a_t)'|Y_{t-1}]$, it is given by the set of recursions:

$$\begin{array}{rclrcl} \nu_t & = & y_t - Z_t a_t, & F_t & = & Z_t P_t Z_t' + G_t G_t' \\ q_t & = & q_{t-1} + \nu_t' F_t^{-1}\nu_t, & K_t & = & (T_t P_t Z_t' + H_t G_t')F_t^{-1} \\ a_{t+1} & = & T_t a_t + d_t + K_t\nu_t, & P_{t+1} & = & T_t P_t T_t' + H_t H_t' - K_t F_t K_t' \end{array} \tag{8}$$

with $q_0 = 0$, $\nu_t = y_t - \mathsf{E}(y_t|Y_{t-1})$ are the filter innovations or one-step-ahead prediction errors, with MSE matrix $\sigma^2 F_t$. Initialisation of the state vector when nonstationary state components are present is discussed in Koopman (1997). In SsfPack the diagonal elements of P_1 corresponding to nonstationary components are set equal to a large number (10^7).

The Kalman filter is said to have reached a steady state if, for some t, $P_t = P$. The conditions under which $\lim_{t\to\infty} P_t = P$ are given in Harvey (1989, sec. 3.3.3 and 3.3.4), and, when they are met, the matrix P is the solution of the Riccati equation $P = TPT' + HH' - KFK'$, with $K = (TPZ' + HG')F^{-1}$ and $F = ZPZ' + GG'$.

A.2 Leave-k-out diagnostics

Leave-k-out diagnostics are based upon the output of the smoothing filter (De Jong, 1988, 1989, Kohn and Ansley, 1989):

$$\begin{array}{rclrcl} u_t & = & F_t^{-1}v_t - K_t' r_t, & M_t & = & F_t^{-1} + K_t' N_t K_t, \\ r_{t-1} & = & Z_t' F_t^{-1}\nu_t + L_t' r_t, & N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \end{array} \tag{9}$$

$L_t = T_t - K_t Z_t$, started with $r_T = 0$ and $N_T = 0$; u_t is termed a *smoothing error.*

Assuming that observations $y_{i-k+1}, \ldots, y_i$ are deleted, let us denote the stack of the deleted observations by $y_{(I)}$ and the vector of deletion residuals by $d_{(I)} = y_{(I)} - \mathsf{E}[y_{(I)}|y_1, \ldots, y_{i-k}, y_{i+1}, \ldots, y_T]$; let also $y = (y_1, \ldots, y_T)' \sim \mathsf{N}(0, \sigma^2 V)$, so that $q_T = y'V^{-1}y$.

The statistic

$$\hat{\tau}_{(I)} = \frac{d_{(I)}'[\mathsf{Cov}(d_{(I)})]^{-1}d_{(I)}}{q_T - d_{(I)}'[\mathsf{Cov}(d_{(I)})]^{-1}d_{(I)}} \cdot \frac{T-k}{k}.$$

provides a test that the observations are jointly outlying. Under normality the exact distribution of $\hat{\tau}_{(I)}$ is $\mathrm{F}(k, T-k)$. It can be computed by running backwards the Kalman filter on the pseudo-model made up of the measurement equation $u_t = F_t^{-1}\nu_t - K_t' r_t$ and transition equation $r_{t-1} = Z_t' F_t^{-1}\nu_t + L_t' r_t$, where $F_t^{-1/2}\nu_t$ act as disturbances, u_t are the observations and r_t the states. This produces:

$$\begin{array}{rclrcl} u_t^* &=& u_t + K_t' r_t^*, & M_t^* &=& F_t^{-1} + K_t' N_t^* K_t, \\ q_{t-1}^* &=& q_t^* + u_t^{*\prime} M_t^{*-1} u_t^*, & K_t^* &=& (Z_t' F_t^{-1} - L_t' N_t^* K_t) M_t^{*-1}, \\ r_{t-1}^* &=& L_t' r_t^* + K_t^* u_t^*, & N_{t-1}^* &=& Z_t' F_t^{-1} Z_t + L_t' N_t^* L_t - K_t^* M_t^* K_t^{*\prime}, \end{array} \tag{10}$$

for $t = i, i-1, \ldots, i-k+1$. The filter is initialised by the unconditional mean and covariance matrix of r_i, that is $r_i^* = 0$ and $N_i^* = N_i$.

In Proietti (2000) it is shown that $q_{i-k+1}^* = d_{(I)}' [\mathsf{Cov}(d_{(I)})]^{-1} d_{(I)}$ so the statistic can be computed as

$$\hat{\tau}_{(I)} = \frac{q_{i-k+1}^*}{q_T - q_{i-k+1}^*} \frac{T-k}{k}. \tag{11}$$

Notice that, once the filter (10 is run for the maximum k desired, leave-j-out diagnostics for $j < k$ are immediately available.

Table 1: Parameter estimates and diagnostics for linear structural models of US unemployment

	LLM	LLTM	TpCM	CTM	CTM_2	CTM_2s	ARTM
σ^2_η	0.0469	0.0110	0.0213	0.0088			
σ^2_ζ		0.0041					
σ^2_κ			0.0186	0.0029	0.0050	0.0042	0.0085
ρ			0.9825	0.9119	0.8723	0.8886	0.7853
λ_c			0.1159	0.1866	0.1846	0.1872	
σ^2_ϵ	0.0000	0.0093	0.0000	0.0103	0.0135	0.0128	0.0122
Φ						0.6209	
Θ						-0.8694	
loglik	69.38	89.42	94.86	121.25	119.47	134.36	115.75
$Q(12)$	170.74	42.25	74.30	19.58	23.03	8.78	30.18
$Q(24)$	205.28	65.93	90.14	44.88	48.34	29.87	51.56
Normality	348.63	417.24	499.91	405.72	359.57	346.55	436.07

Table 2: Linear models: comparison of forecast performance in the test period 1980.1-2000.12.

	LLM	LLTM	TpCM	CTM	CTM_2	CTM_2s	ARTM	ARIMA
Lead time				Mean Error				
1 month	-0.0091	-0.0014	-0.0117	-0.0061	-0.0054	-0.0055	-0.0046	-0.0144
1 quarter	-0.0279	-0.0054	-0.0371	-0.0206	-0.0184	-0.0188	-0.0161	-0.0497
2 quarters	-0.0684	-0.0229	-0.0904	-0.0611	-0.0565	-0.0569	-0.0500	-0.1239
3 quarters	-0.1146	-0.0472	-0.1515	-0.1129	-0.1050	-0.1065	-0.0932	-0.2147
1 year	-0.1579	-0.0698	-0.2101	-0.1638	-0.1518	-0.1545	-0.1351	-0.3090
Lead time				Symmetric Mean Absolute Percentage Error				
1 month	1.87	1.84	1.96	1.81	1.83	1.79	1.79	1.76
1 quarter	3.44	3.26	3.89	3.22	3.20	3.15	3.09	3.14
2 quarters	5.83	5.91	7.33	5.73	5.56	5.56	5.09	5.57
3 quarters	8.00	8.95	10.95	8.26	7.83	7.93	6.93	8.13
1 year	10.24	12.57	14.50	10.95	10.29	10.53	9.09	11.02
Lead time				Median Relative Absolute Error				
1 month	1.00	1.00	1.11	0.99	1.00	1.00	0.97	0.93
1 quarter	1.00	0.86	1.30	1.01	0.98	0.97	0.90	0.94
2 quarters	1.00	0.90	1.52	1.02	0.94	1.00	0.85	0.95
3 quarters	1.00	1.02	1.69	1.04	0.97	1.00	0.85	1.10
1 year	1.00	1.07	1.76	1.08	0.99	1.01	0.87	1.21
Lead time				Mean Square Forecast Error				
1 month	0.0265	0.0247	0.0264	0.0236	0.0237	0.0224	0.0233	0.0221
1 quarter	0.1062	0.0928	0.1105	0.0810	0.0811	0.0794	0.0779	0.0747
2 quarters	0.3040	0.3276	0.3736	0.2456	0.2416	0.2399	0.2265	0.2247
3 quarters	0.5694	0.7868	0.8060	0.5090	0.4887	0.4923	0.4553	0.4789
1 year	0.8943	1.5627	1.3825	0.8735	0.8266	0.8444	0.7710	0.8393

Table 3: Parameter estimates (Est.), standard errors (S.E.) and diagnostics for nonlinear structural models of US unemployment ("conc" denotes that the parameter is concentrated out of the likelihood function).

	ARTMSt		ARTMStD		CTM$_2$St		CTM$_2$StD	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
$\sigma^2_{\kappa 0}$	0.0029	0.10	0.0026	0.10	0.0009	0.04	0.0004	0.02
$\sigma^2_{\kappa 1}$	0.0643	2.24	0.0668	1.59	0.0550	1.91	0.0639	1.32
ρ_0	0.8148	0.01	0.7867	0.01	0.8981	0.06	0.8957	0.03
ρ_1	0.4978	0.07	0.2289	0.03	0.5494	0.08	0.2095	0.05
λ_{c0}					0.1419	0.02	0.1571	0.02
λ_{c1}					0.0000	-	0.0000	-
β_0			-0.0160	0.03			-0.0193	0.01
β_1			0.1025	0.02			0.1172	0.02
τ	33.85	30.61	50.85	34.18	39.94	30.61	52.67	25.94
c	0.3436	0.05	0.3204	0.02	0.3122	0.05	0.3127	0.02
σ^2_ϵ	0.0097	conc	0.0099	conc	0.0109	conc	0.0116	conc
loglik	165.98		172.31		172.13		181.26	
$Q(12)$	25.51		23.59		19.70		21.36	
$Q(24)$	44.53		42.10		36.92		37.26	
Normality	125.78		86.17		122.02		62.48	

Table 4: Nonlinear models: comparison of forecast performance in the test period 1980.1-2000.12.

	ARTM	ARTMSt	ARTMStD	CTM_2St	CTM_2StD
Lead time			Mean Error		
1 month	-0.0046	0.0018	-0.0074	-0.0018	-0.0104
1 quarter	-0.0161	0.0146	-0.0208	0.0082	-0.0254
2 quarters	-0.0500	0.0346	-0.0565	0.0240	-0.0614
3 quarters	-0.0932	0.0618	-0.0961	0.0469	-0.1027
1 year	-0.1351	0.1028	-0.1270	0.0820	-0.1408
Lead time		Symmetric Mean Absolute Percentage Error			
1 month	1.79	1.79	1.77	1.78	1.76
1 quarter	3.09	3.07	3.05	3.02	2.91
2 quarters	5.09	5.15	5.11	5.11	4.91
3 quarters	6.93	6.90	6.92	6.92	6.78
1 year	9.09	8.82	9.03	8.82	8.98
Lead time		Median Relative Absolute Error			
1 month	1.00	1.00	0.97	1.00	0.99
1 quarter	1.00	0.97	0.94	0.95	0.91
2 quarters	1.00	0.97	1.00	0.99	0.97
3 quarters	1.00	0.94	0.99	0.96	1.00
1 year	1.00	0.87	0.99	0.89	1.01
Lead time		Mean Square Forecast Error			
1 month	0.0233	0.0228	0.0226	0.0223	0.0217
1 quarter	0.0779	0.0821	0.0794	0.0793	0.0715
2 quarters	0.2265	0.2544	0.2385	0.2476	0.2134
3 quarters	0.4553	0.5045	0.4780	0.5005	0.4454
1 year	0.7710	0.8198	0.7931	0.8233	0.7675
Lead time		Meese-Rogoff Test (p-values)			
1 month		0.24	0.15	0.15	0.05
1 quarter		0.82	0.70	0.62	0.03
2 quarters		0.94	0.95	0.86	0.08
3 quarters		0.91	0.94	0.85	0.29
1 year		0.82	0.85	0.78	0.44

Figure 1: U.S. unemployment rate, Jan. 1948 - Dec. 2000; seasonally adjusted and unadjusted series.

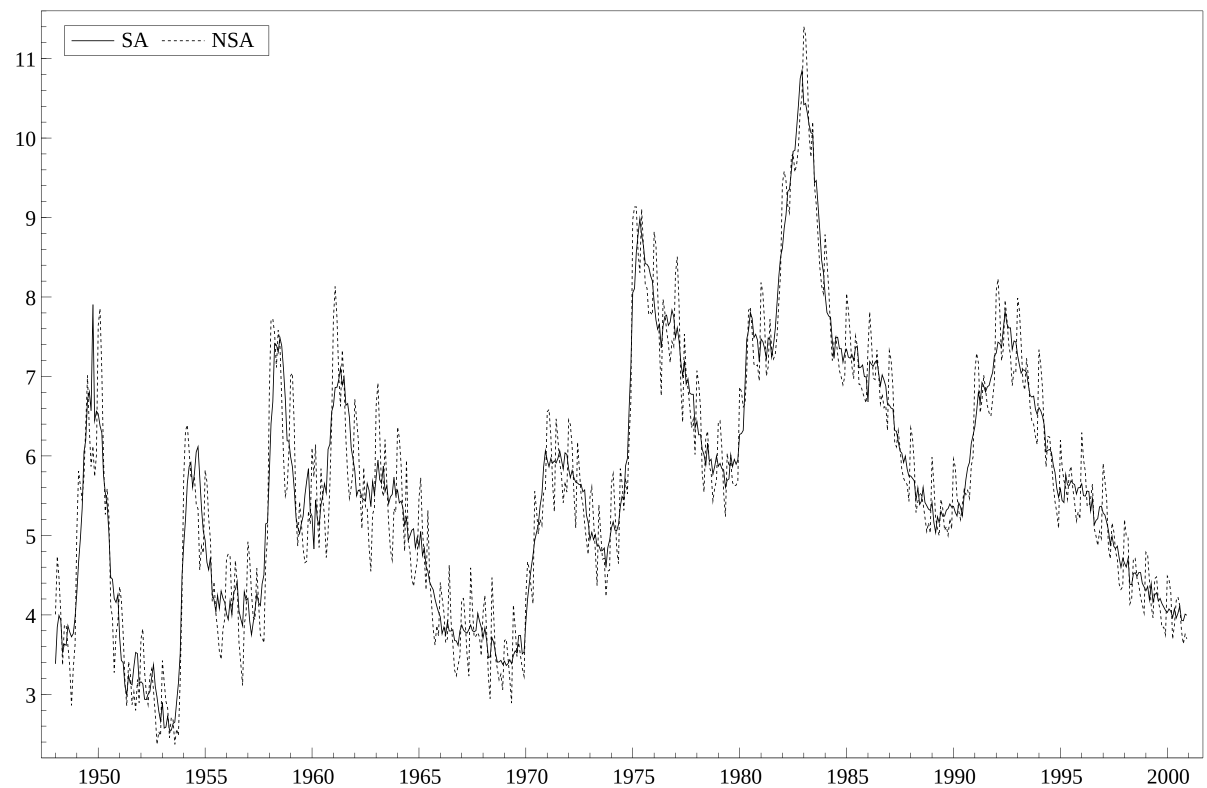

Figure 2: Impulse response function for linear models of US unemployment rate (ARIMA refers to the model fitted by Montgomery *et al.* (1998), see section 3).

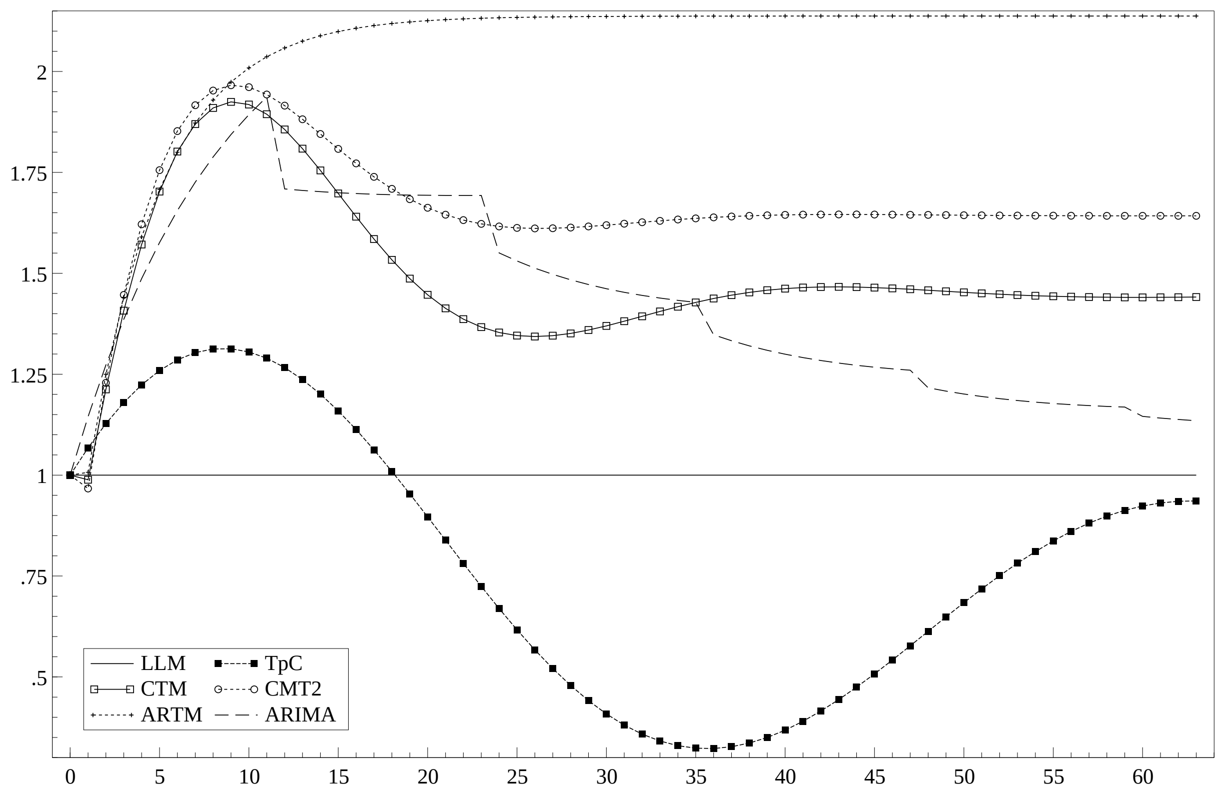

Figure 3: Rolling forecast comparison of linear models: mean square forecast error relative to benchmark (LLM).

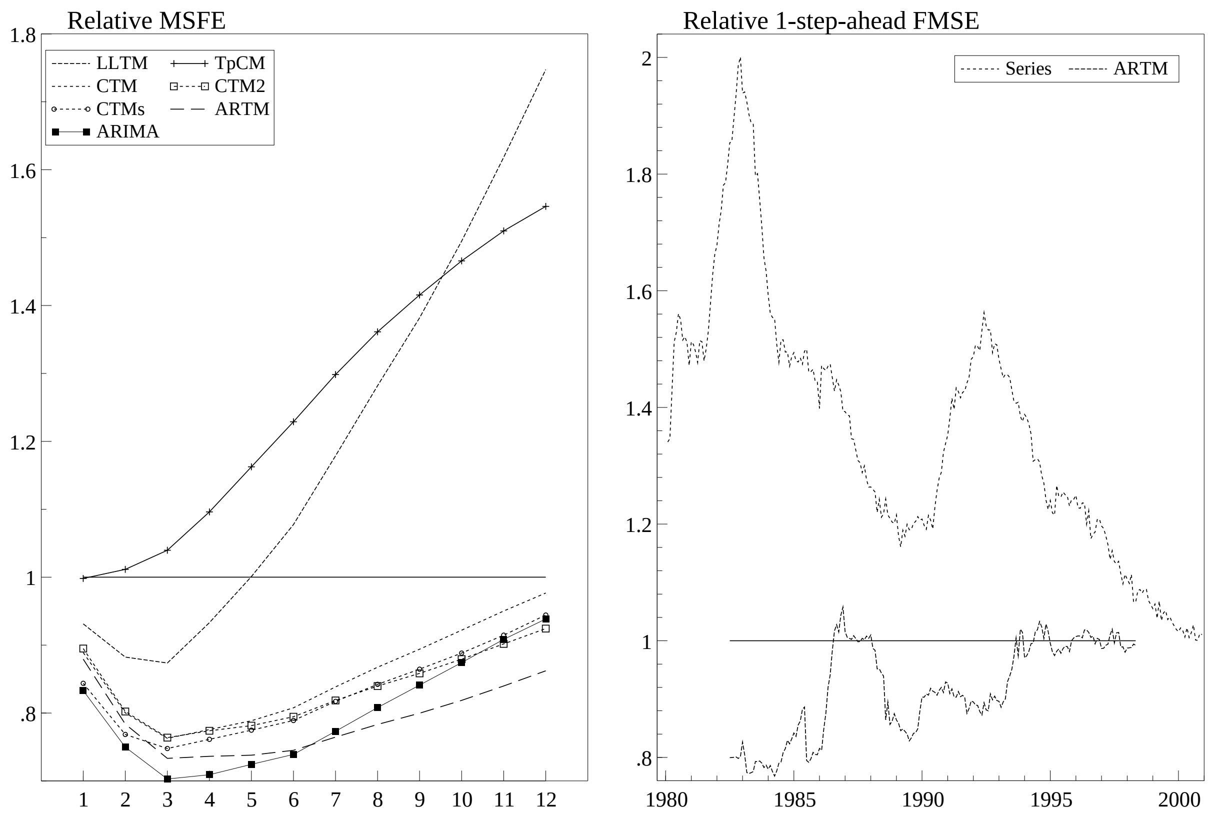

Figure 4: Leave-k-out diagnostics for ARTM.

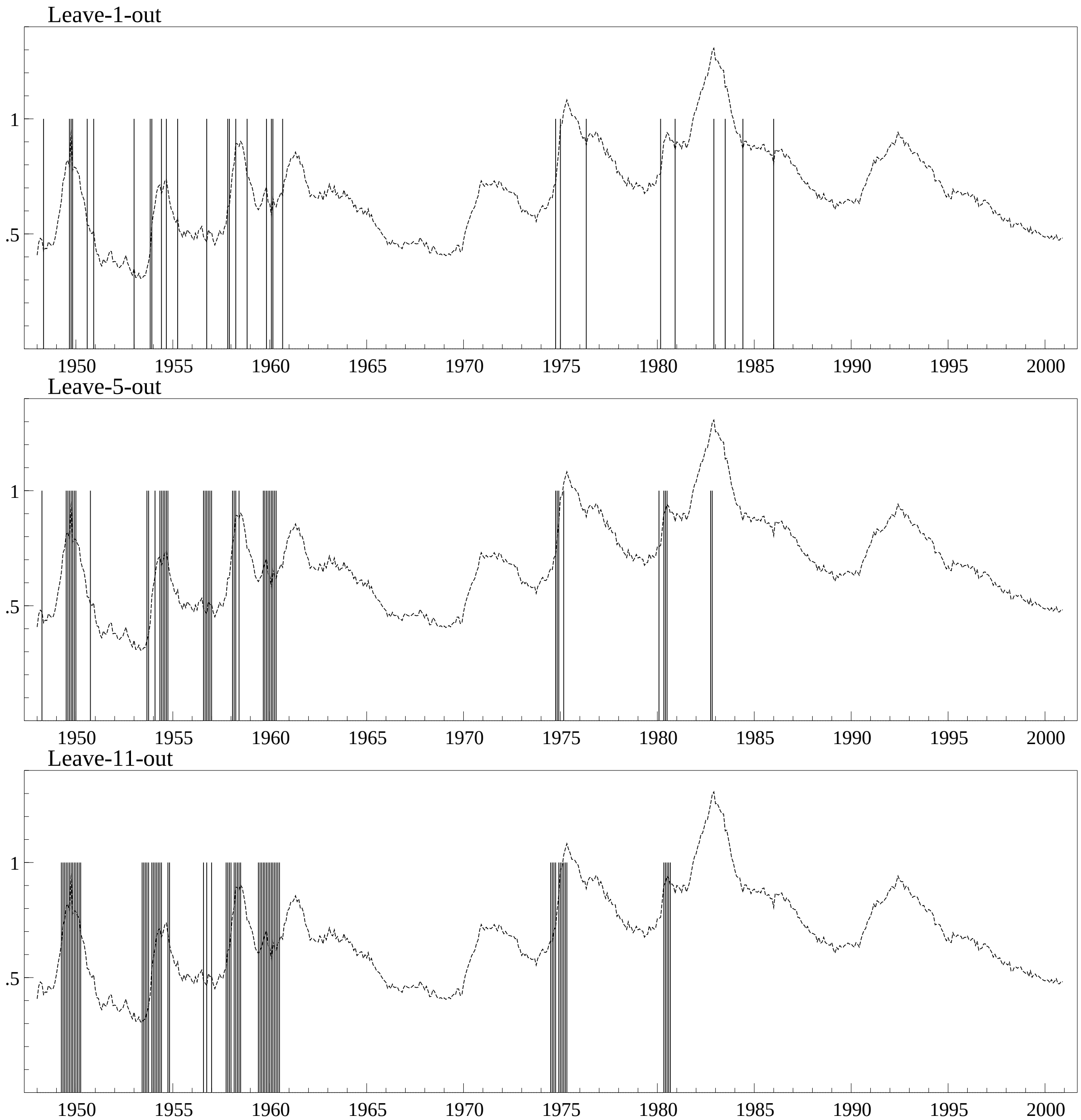

Figure 5: Transition variable z_t and estimated transition function, $S_t = [1+\exp(-\tau(z_t - c))]^{-1}$ for the model $\mathsf{CTM_2StD}$.

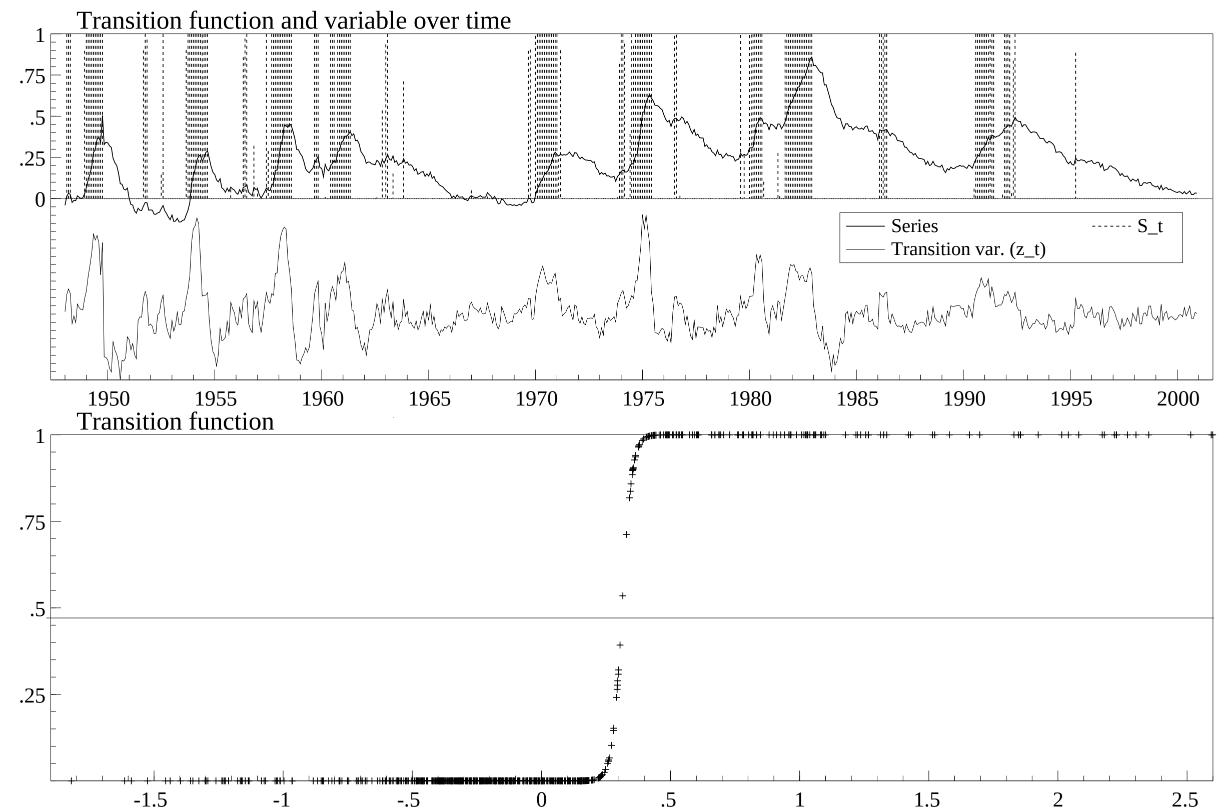

Figure 6: Rolling forecast comparison of nonlinear models: mean square forecast error relative to benchmark (ARTM).

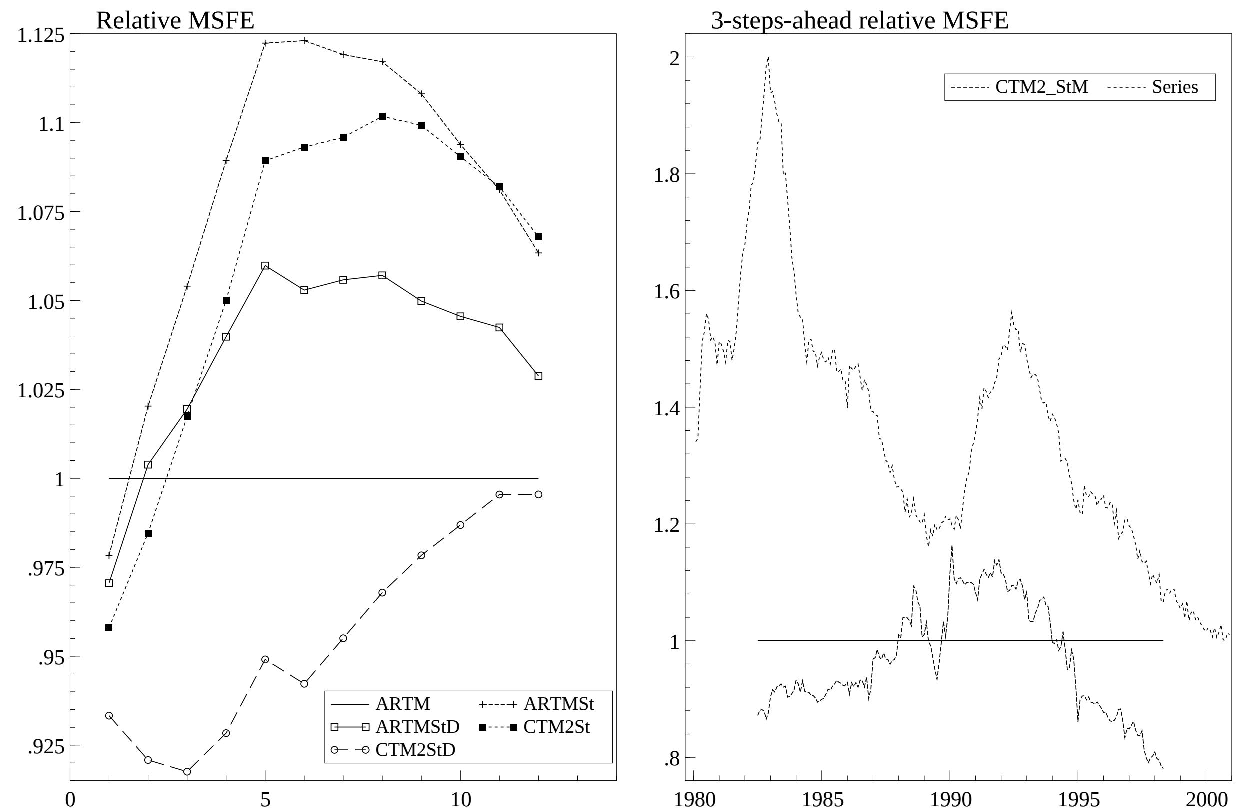

Figure 7: $\mathsf{CTM_2StD}$ model forecasts, 2000.1-2000.12, highest density regions, observed values and upper and lower 95% confidence limits for the linear benchmark ARTM.

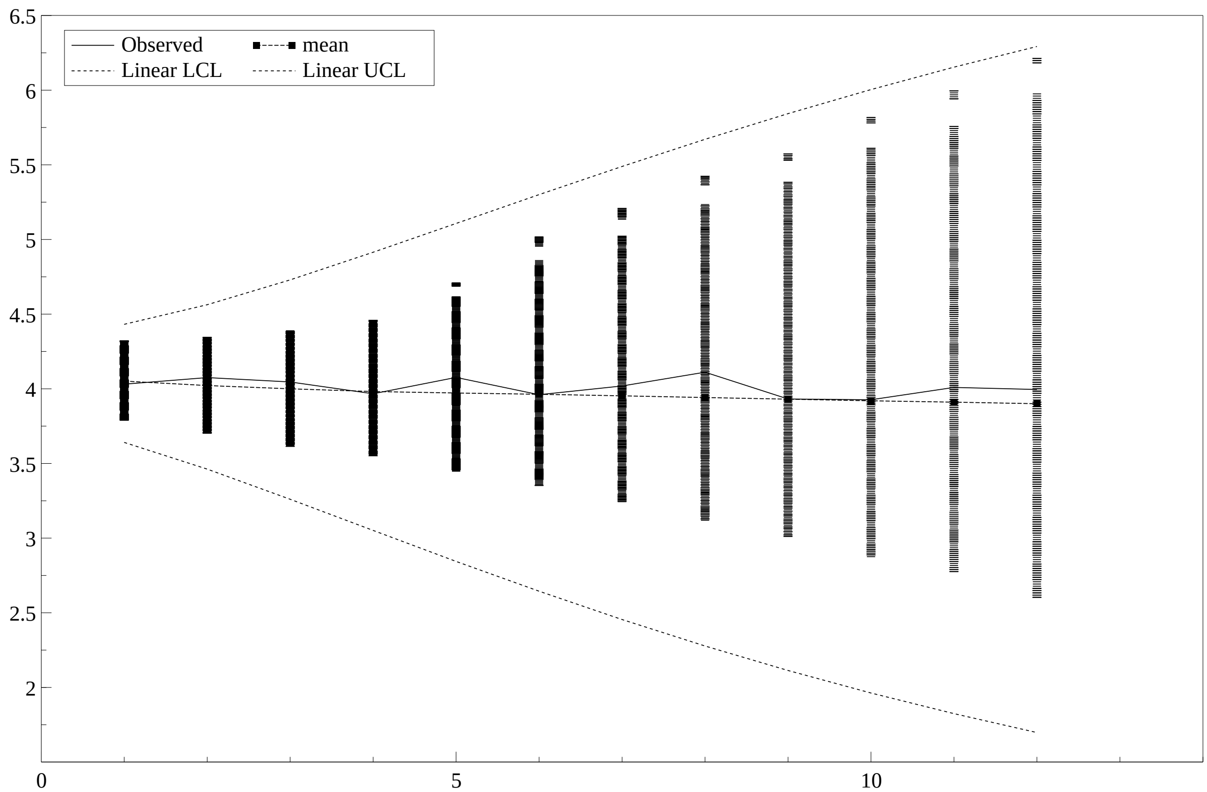

Figure 8: Generalised impulse response function. Plot of $\mathrm{GIRF}(60, \nu_t, Y_{t-1})$ (persistence) versus ν_t (innovation) for the model $\mathsf{CTM_2StD}$ during an economic downturn (increasing unemployment rates, left panel) and during expansions (decreasing unemployment rates, right panel).

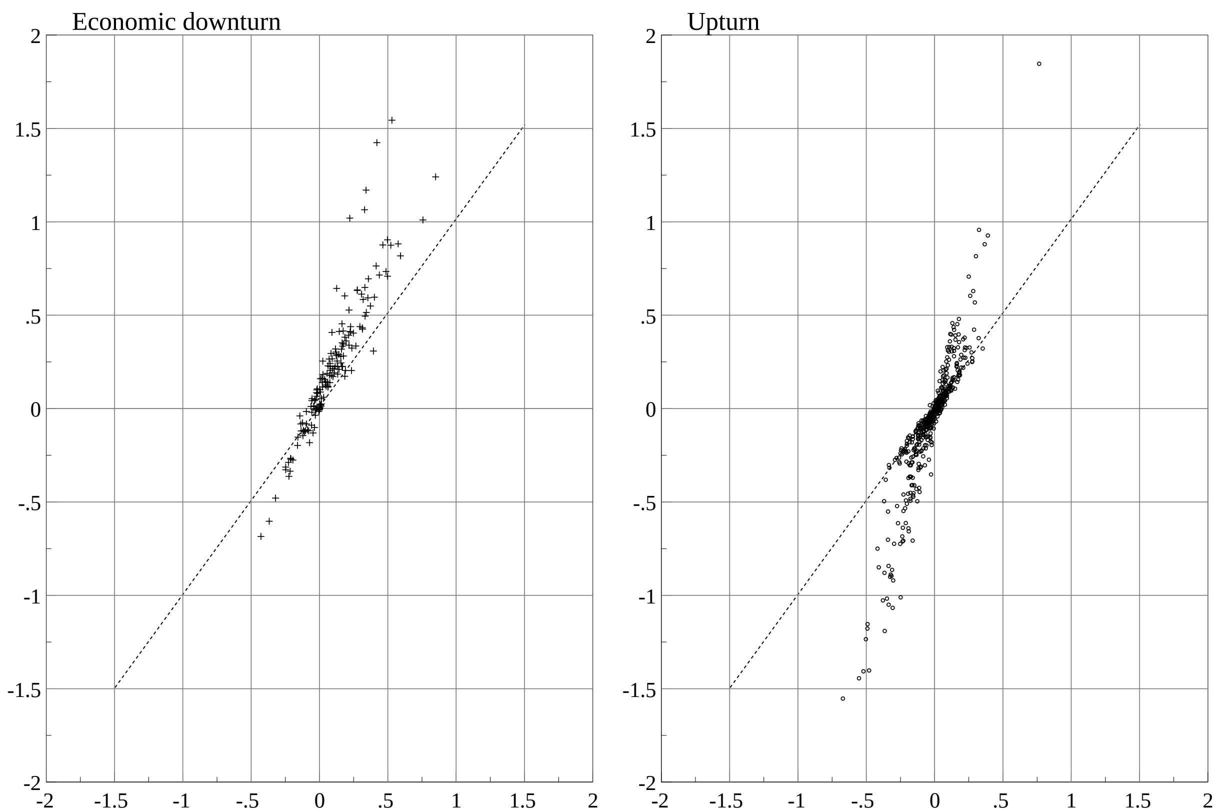

Figure 9: Normalised GIRF$(3, \pm 2\sigma, Y_{t-1})$ (upper panel) and $\partial y_{t+3}/\partial \nu_t$ (lower panel) for the non linear model $\mathsf{CTM_2StD}$; the index plot in the background displays the values of the transition function S_t.