COMPARING SMOOTH TRANSITION AND MARKOV SWITCHING AUTOREGRESSIVE MODELS OF US UNEMPLOYMENT

PHILIPPE J. DESCHAMPS

Université de Fribourg

Final revision, June 2007

ABSTRACT. Logistic smooth transition and Markov switching autoregressive models of a logistic transform of the monthly US unemployment rate are estimated by Markov chain Monte Carlo methods. The Markov switching model is identified by constraining the first autoregression coefficient to differ across regimes. The transition variable in the LSTAR model is the lagged seasonal difference of the unemployment rate. Out of sample forecasts are obtained from Bayesian predictive densities. Although both models provide very similar descriptions, Bayes factors and predictive efficiency tests (both Bayesian and classical) favor the smooth transition model.

SÉMINAIRE D'ÉCONOMÉTRIE, BOULEVARD DE PÉROLLES 90, CH-1700 FRIBOURG, SWITZERLAND. TELEPHONE: +41-26-300-8252. TELEFAX: +41-26-481-3845.

E-mail address: philippe.deschamps@unifr.ch

1. Introduction

US unemployment is characterized by relatively brief periods of rapid economic contraction (rising unemployment) followed by relatively extended periods of slow economic expansion (falling unemployment). In the recent literature, several studies (Rothman, 1998; Montgomery *et al.*, 1998; Koop and Potter, 1999; Van Dijk *et al.*, 2002) have attempted to capture this salient feature by means of well-known nonlinear models such as the threshold autoregression (TAR), the closely related logistic smooth transition autoregression (LSTAR), and the Markov switching autoregression (MSAR). The first two papers attempt to base model choice on a comparison of forecasting performance.

There are two important conceptual differences between the MSAR and the TAR or LSTAR models. First, MSAR incorporates less prior information than TAR and LSTAR. Indeed, a filtered or smoothed regime probability in an MSAR model can be interpreted as a transition function which is estimated flexibly from the data. By contrast, specifying the transition function in a TAR or LSTAR model necessitates the choice of a transition variable (a difficult problem). Secondly, regime changes are predetermined in a TAR or LSTAR model, but exogenous in the MSAR: in the latter, even if the model parameters were known, these changes could not be predicted with certainty from past data due to the presence of additional disturbances (in the Markov evolution equation). It is of interest to investigate whether the added flexibility and complexity of the MSAR model results in a superior predictive ability.

For the reasons given by West and McCracken (1998), it may be important to base such an investigation on small-sample predictive densities that take parameter uncertainty into account. In the context of maximum likelihood estimation, this requires using the bootstrap, and involves the repeated estimation of nonlinear models by local optimization algorithms. Convergence difficulties may make this impractical: see, e.g., Chan and McAleer (2002, 2003). These difficulties do not arise if Bayesian methods are used: Markov chain Monte Carlo (MCMC) is used for simulating the joint posterior, and the resulting parameter replications are used for dynamic simulations of future observations.

In a Bayesian context, a prior identification constraint on the parameters of Markov switching models should be imposed; if this is not done, a multimodal posterior is obtained, except by spurious accident when mixing in the MCMC sampler is poor. This constraint can take the form $\theta_1 < \theta_2 < \cdots < \theta_K$, where θ_i is a particular population parameter in regime i and K is the number of regimes. The permutation sampler proposed by Frühwirth-Schnatter (2001) provides an effective procedure both for choosing an appropriate prior identification constraint, and for subsequently imposing this constraint. An MCMC posterior simulator for LSTAR models has

been proposed by Lopes and Salazar (2006).

Among the authors mentioned in the first paragraph, only Montgomery *et al.* (1998) investigate the forecasting performance of an MSAR model; and only Koop and Potter (1999) fully rely on Bayesian methods. Even though the MSAR model in Montgomery *et al.* (1998) is estimated by the Gibbs sampler, the presentation is frequentist: the authors do not present their prior specification (including those aspects of the prior that are relevant for model identification) and only discuss point estimates and point forecasts. Koop and Potter (1999) provide a thorough Bayesian treatment of a TAR model of US unemployment; however, they do not compare its forecasting performance with that of an MSAR model.

An LSTAR formulation can approximate the TAR models used in Montgomery *et al.* (1998) and Koop and Potter (1999), but can be estimated with standard econometric software (contrary to the TAR and MSAR models), and is therefore particularly convenient. Van Dijk *et al.* (2002) have been the only authors to investigate the forecasting accuracy of an LSTAR model of the US unemployment rate. However, they do not update the parameter estimates as new observations become available, presumably for the reasons given in our third paragraph.

As pointed out by Koop and Potter (1999), there are important benefits in using a logistic transformation of the unemployment rate. This transformation not only guarantees that predictions are restricted to the unit interval (an important consideration if the emphasis is on predictive densities), but also removes the strong residual leptokurticity which plagues a model estimated from untransformed data. Among the four contributions mentioned in the first paragraph, only the paper by Koop and Potter (1999) uses such a transformation.

On these grounds, and since the permutation sampler has only recently become available, it may be argued that the potential of the LSTAR and MSAR models for predicting the US unemployment rate should be examined in more detail, and that true Bayesian predictive densities should be used in the investigation. This is the twofold objective of this paper.

An outline follows. Section 2 presents an MCMC posterior simulator for LSTAR models. It differs from the previous one in two respects. First, an independence Metropolis-Hastings chain is used, rather than the random walk chain used by Lopes and Salazar (2006). Secondly, the autoregressive order p and transition delay parameter d are assumed to be fixed (whereas one of the algorithms proposed by Lopes and Salazar is defined on a space that includes p and d). In our approach, we propose to choose p and the transition variable (or function) according to the criterion of highest marginal likelihood; some potential advantages are discussed. Section 2 therefore also describes our application of the bridge sampling method of Meng and Wong (1996) to the estimation of marginal likelihoods in a STAR model.

Section 3 briefly describes the MCMC estimation of the MSAR model and the bridge sampling

estimation of marginal likelihoods for this model.

Section 4 presents estimated marginal likelihoods for 54 possible LSTAR, MSAR, and autoregressive (AR) models, where the dependent variable is a logistic transformation of the monthly US unemployment rate; a sensitivity analysis with respect to the prior parameters is done.

Section 5 discusses Bayesian misspecification diagnostics for the LSTAR and MSAR models that were found, in Section 4, to have the highest marginal likelihoods. The diagnostics are based on posterior predictive p-values for three relevant misspecification indicators.

Section 6 presents the MCMC estimates of the chosen LSTAR model and of its MSAR counterpart; some economic implications of the estimates are discussed.

Section 7 presents, for comparison purposes, maximum likelihood estimates of the models in Section 6, and diagnostics based on generalized residuals.

Finally, Section 8 attempts to discriminate between the MSAR, the LSTAR, and a benchmark AR model by means of simulated out of sample prediction exercises. For each model, Bayesian predictive densities are estimated from expanding windows of observations and for horizons of 1 to 6 months. Diagnostics based on probability integral transforms (Diebold *et al.*, 1998; Berkowitz, 2001), on one of the test statistics proposed by Diebold and Mariano (1995), and on efficiency tests based on regressions of observations on point predictions are reported; versions of the efficiency tests are analyzed from both classical and Bayesian standpoints. Section 9 concludes.

2. A POSTERIOR SIMULATOR FOR SMOOTH TRANSITION AUTOREGRESSIONS

The two-state LSTAR model introduced by Teräsvirta (1994) may be written as:

$$y_t = \alpha_1 + \sum_{j=1}^{p} \phi_{1j} y_{t-j} + G(s_t, \gamma, c)\Big(\alpha_2 + \sum_{j=1}^{p} \phi_{2j} y_{t-j}\Big) + u_t, \quad \text{with:} \tag{2.1}$$

$$G(s_t, \gamma, c) = \frac{1}{1 + \exp[-(\frac{\gamma^2}{r})(s_t - c)]} \tag{2.2}$$

where r is a scaling constant that can be set equal to the sample standard deviation of the observable variable s_t, and where, conditional on s_t and $y_{t-1}, \ldots, y_{t-p}$, u_t is a random disturbance with distribution $N(0, \sigma^2)$. $G(s_t, \gamma, c)$ is called the transition function; s_t the transition variable; γ the shape parameter; and c the location parameter. The model implies transitions between the two regimes where $G(s_t, \gamma, c) = 0$ (which tends to occur when $s_t < c$) and $G(s_t, \gamma, c) = 1$ (which tends to occur when $s_t > c$). When γ becomes large, $G(s_t, \gamma, c)$ tends to the step function postulated by a two-state TAR model. In (2.2), the division of γ^2 by r ensures that γ has a comparable order of magnitude across competing models.

The MCMC algorithm of this section iterates on the full conditional posteriors of the vector:

$$\beta' = (\, \alpha_1 \quad \phi_{11} \quad \ldots \quad \phi_{1p} \quad \alpha_2 \quad \phi_{21} \quad \ldots \quad \phi_{2p} \,), \tag{2.3}$$

of σ^2, and of (γ, c), using the most recently drawn conditioning values.

If γ and c are known, equation (2.1) collapses to the usual regression model $y = X\beta + u$, where row t of the $T \times (2p + 2)$ matrix X has the form:

$$(1 \quad y_{t-1} \quad \ldots \quad y_{t-p} \quad G_t \quad G_t y_{t-1} \quad \ldots \quad G_t y_{t-p})$$

with $G_t \equiv G(s_t, \gamma, c)$. A multinormal prior on β with expectation vector β_a and precision matrix V_a and an independent inverted Gamma prior on σ^2 with parameters a and b are assumed. It is then straightforward to show that the full conditional posteriors of β and σ^2 are respectively multinormal and inverted Gamma:

$$p(\beta \mid \sigma^2, \gamma, c, \text{data}) \propto \exp\left[-\frac{1}{2}(\beta - \beta_*)'V_*(\beta - \beta_*)\right] \tag{2.4}$$

$$p(\sigma^2 \mid \beta, \gamma, c, \text{data}) \propto (\sigma^2)^{-a^*-1} \exp\left(-\frac{b^*}{\sigma^2}\right) \tag{2.5}$$

with:

$$V_* = \frac{X'X}{\sigma^2} + V_a$$

$$\beta_* = V_*^{-1}\left(\frac{X'y}{\sigma^2} + V_a \beta_a\right)$$

$$a^* = a + \frac{T}{2}$$

$$b^* = b + \frac{(y - X\beta)'(y - X\beta)}{2}.$$

The full conditional posterior of the parameters of the transition function is nonstandard. Upon taking independent normal priors $N(\gamma_a, \sigma_\gamma^2)$ and $N(c_a, \sigma_c^2)$ for γ and c and noting $\vartheta = (\gamma, c)$, the kernel of the full conditional posterior $p(\vartheta \mid \beta, \sigma^2, \text{data})$ is:

$$k^*(\vartheta) = \exp\left[-\frac{(\gamma - \gamma_a)^2}{2\sigma_\gamma^2} - \frac{(c - c_a)^2}{2\sigma_c^2}\right.$$
$$\left. -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \alpha_1 - \sum_{j=1}^{p}\phi_{1j}y_{t-j} - G(s_t, \gamma, c)\left[\alpha_2 + \sum_{j=1}^{p}\phi_{2j}y_{t-j}\right]\right)^2\right]. \tag{2.6}$$

The algorithm for simulating ϑ is based on a Metropolis-Hastings independence chain (Tierney, 1994), using a multivariate Student candidate-generating density with location and scale parameters based on the following linearization, obtained by taking a first-order Taylor expansion of

(2.1)–(2.2) around (γ^*, c^*) and regrouping in the left-hand side those terms that do not depend on γ and c:

$$y_t^* = \gamma\, x_{1t}^* + c\, x_{2t}^* + v_t$$

where:

$$y_t^* = y_t - \alpha_1 - \sum_{j=1}^{p} \phi_{1j} y_{t-j} - \left(\alpha_2 + \sum_{j=1}^{p} \phi_{2j} y_{t-j}\right) \left(G(s_t, \gamma^*, c^*) - \left.\frac{\partial G_t}{\partial \gamma}\right|_{\gamma^*, c^*} \gamma^* - \left.\frac{\partial G_t}{\partial c}\right|_{\gamma^*, c^*} c^*\right)$$

$$x_{1t}^* = \left(\alpha_2 + \sum_{j=1}^{p} \phi_{2j} y_{t-j}\right) \left.\frac{\partial G_t}{\partial \gamma}\right|_{\gamma^*, c^*}$$

$$x_{2t}^* = \left(\alpha_2 + \sum_{j=1}^{p} \phi_{2j} y_{t-j}\right) \left.\frac{\partial G_t}{\partial c}\right|_{\gamma^*, c^*}.$$

The anchor point $\vartheta^* = (\gamma^*, c^*)$ is an approximate solution of the Bayesian update equations:

$$\vartheta^* = S_\vartheta^{-1} \left[\frac{X_*' y_*}{\sigma^2} + \begin{pmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_a \\ c_a \end{pmatrix}\right] \tag{2.7}$$

$$S_\vartheta = \frac{X_*' X_*}{\sigma^2} + \begin{pmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix}^{-1} \tag{2.8}$$

where X_* is the $T \times 2$ matrix with row t equal to $(x_{1t}^* \quad x_{2t}^*)$ and where y_* is the $T \times 1$ vector with elements y_t^*. This approximate solution is obtained from a few iterations on (2.7) and (2.8), with starting point given by the prior expectations. A candidate ϑ is drawn from a multivariate Student density with kernel:

$$k(\vartheta) = \left[1 + \frac{(\vartheta - \vartheta^*)' S_\vartheta (\vartheta - \vartheta^*)}{\nu}\right]^{-(\nu+2)/2} \tag{2.9}$$

and is accepted with probability:

$$\min\left[1, \frac{k(\vartheta_{\text{old}})}{k(\vartheta)} \frac{k^*(\vartheta)}{k^*(\vartheta_{\text{old}})}\right]$$

where ϑ_{old} is the most recently drawn vector. If the candidate is rejected, ϑ_{old} is retained. The number ν of degrees of freedom in (2.9) can be chosen by experimentation; in the empirical part of this paper, a value of $\nu = 3$ was chosen and led to acceptance rates of approximately 0.80.

Lopes and Salazar (2006) parameterize equation (2.2) in terms of γ rather that γ^2, and ensure the positivity of γ by specifying a prior with positive support (such as a Gamma distribution) for this parameter. They use a random walk Metropolis-Hastings chain, where two tuning parameters must be specified. An advantage of the method proposed in this section is that its implementation can be automatic: choosing $\nu = 3$ in (2.9) seems to give uniformly good results, with high

acceptance rates and well-mixing chains. This advantage will prove decisive in Section 8, where several thousand MCMC estimations will be needed.

Lopes and Salazar (2006) also present a reversible jump MCMC method where the autoregressive lag order and transition delay parameter are included in the parameter space. By contrast, our method treats p, and the transition variable s_t, as fixed; it is proposed to investigate the choice of p and s_t by estimating marginal likelihoods for a range of candidate models. Although less ambitious than reversible jump MCMC, this approach easily allows the comparative investigation of models where s_t is *any* transition variable, and $G(.)$ is *any* transition function (in particular, LSTAR and ESTAR models can easily be compared).

The rest of this section describes the method for estimating marginal likelihoods. It is based on the bridge sampling identity (Meng and Wong, 1996), which allows to estimate the ratio of the normalizing constants of two density kernels with overlapping support. For a model with prior $p(\theta)$ and likelihood $f(y \mid \theta)$, and given a bridge function $\alpha(\theta)$, the marginal likelihood:

$$p(y) = \int p(\theta) f(y \mid \theta) d\theta$$

is equal to:

$$p(y) = \frac{\int \left[p(\theta) f(y \mid \theta) \alpha(\theta) \right] q(\theta) d\theta}{\int \left[q(\theta) \alpha(\theta) \right] p(\theta \mid y) d\theta} \tag{2.10}$$

where $q(\theta)$ is a normalized importance density. The numerator in (2.10) can be estimated by an average of replications of $p(\theta) f(y \mid \theta) \alpha(\theta)$ where θ is drawn from $q(\theta)$, and the denominator by an average of replications of $q(\theta^*) \alpha(\theta^*)$, where θ^* is drawn from the posterior.

A good choice of $\alpha(\theta)$ is important for numerical efficiency. Meng and Wong (1996) recommend:

$$\alpha(\theta) = \frac{1}{\left[n q(\theta) + m \frac{p(\theta) f(y \mid \theta)}{p(y)} \right]} \tag{2.11}$$

where we have assumed that n replications from $q(\theta)$ and m replications from $p(\theta \mid y)$ are available. Since $p(y)$ is unknown, (2.11) must be obtained by an iterative procedure.

For the STAR model, the author chose an importance density $q(\theta)$ having the same parametric form as the prior, but with moments that match the empirical posterior moments obtained by MCMC. This choice is appropriate when the marginal posteriors are unimodal, and gave very good results in this case.

Bridge sampling has been shown by Frühwirth-Schnatter (2004) to include as special cases other well-known methods for estimating $p(y)$, such as the method of Gelfand and Dey (1994). The method proposed by Chib and Jeliazkov (2001) has also been shown by Mira and Nicholls (2004) to be a special case of bridge sampling.

6

3. Bayesian estimation of Markov switching autoregressions

The MSAR counterpart of (2.1)–(2.2) is:

$$y_t = \alpha_1 + \sum_{j=1}^{p} \phi_{1j} y_{t-j} + \left(\alpha_2 + \sum_{j=1}^{p} \phi_{2j} y_{t-j} \right) G_t + u_t$$

$$= \left(\gamma_1 + \sum_{j=1}^{p} \beta_{1j} y_{t-j} \right) (1 - G_t) + \left(\gamma_2 + \sum_{j=1}^{p} \beta_{2j} y_{t-j} \right) G_t + u_t \qquad (3.1)$$

where $\gamma_1 = \alpha_1$, $\beta_{1j} = \phi_{1j}$, $\gamma_2 = \alpha_1 + \alpha_2$, and $\beta_{2j} = \phi_{1j} + \phi_{2j}$ for $j = 1, \ldots, p$. Given $(G_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p})$, u_t has the distribution $N(0, \sigma^2)$. The parameters α_i and ϕ_{ij} have the same interpretation as in (2.1), but G_t is here a discrete random variable with a value of zero in the first regime and a value of unity in the second. The prior on $(G_1, \ldots, G_T)$ is first-order Markov, with $P[G_t = 0 \mid G_{t-1} = 0] = \mathbf{p}$ and $P[G_t = 1 \mid G_{t-1} = 1] = \mathbf{q}$; independent uniform hyperpriors on $\mathbf{p}$ and $\mathbf{q}$ are assumed.

The MCMC estimation of this model is now well established; see, e. g., Albert and Chib (1993), and Chib (1996). Assuming conjugate priors, the full conditional posterior of the regression coefficients is multivariate normal; that of σ^2 is inverted Gamma; those of $\mathbf{p}$ and $\mathbf{q}$ are Beta; and draws of $(G_1, \ldots, G_T)$ can be obtained by a simulation smoother. The identification of (3.1), however, has only recently been adequately discussed in the literature. Frühwirth-Schnatter (2001) proposes the permutation sampler, which comes in an unconstrained and a constrained version. In the unconstrained version, each pass of the Gibbs sampler is followed by a random permutation of the regime definitions; this guarantees a balanced sample from the unconstrained posterior. An examination of this sample is used to suggest an appropriate identification constraint; in model (3.1), one may choose a single constraint of the form $\alpha_2 > 0$, $\phi_{2j} > 0$ for one $j \in \{1, \ldots, p\}$, or $\mathbf{p} < \mathbf{q}$. This examination is followed by an application of the constrained version of the sampler, where the chosen identification constraint is imposed. More details can be found in Frühwirth-Schnatter (2006), where Bayesian and classical methods for estimation and specification search in finite mixture models such as (3.1) are extensively reviewed.

It is important to note that the permutation sampler requires a prior that is invariant with respect to relabeling. In the context of (3.1), this implies a prior with $p(\gamma_1, \gamma_2) = p(\gamma_2, \gamma_1)$ and $p(\beta_{1j}, \beta_{2j}) = p(\beta_{2j}, \beta_{1j})$ for all $j = 1, \ldots, p$. When the prior on the regression coefficients is normal, it is of course a simple matter to translate such a prior into an equivalent prior on the α_i and ϕ_{ij}.

Frühwirth-Schnatter (2004; 2006, chapter 5) recommends using bridge sampling for estimating the marginal likelihood of this model from the *unconstrained* permutation sampler output.

The unconstrained marginal likelihood is proportional to the constrained one, with a factor of proportionality equal to the factorial of the number of regimes; this implies that computing the constrained marginal likelihood is unnecessary, and also that marginal likelihoods cannot be used to investigate the adequacy of an identification constraint.

In the bridge sampling identity (2.10), the likelihood $f(y \mid \theta)$ is needed. It can easily be computed by integrating out the latent variables $(G_1, \ldots, G_T)$, as follows:

$$f(y \mid \theta) = \prod_{t=1}^{T} \sum_{k=0}^{1} f_t(y_t \mid y^{t-1}, \theta, G_t = k) P[G_t = k \mid y^{t-1}, \theta] \qquad (3.2)$$

where:

$$\theta = (\gamma_1, \gamma_2, \beta_{11}, \ldots, \beta_{2p}, \mathbf{p}, \mathbf{q}, \sigma^2),$$

where y^{t-1} contains all observations on y_s up to $t-1$, and where the conditional probabilities of the regimes can be evaluated by the filter described in Hamilton (1994, pp. 692–693).

The construction of the importance density $q(\theta)$ in (2.10) requires care, due to the multimodality of the unconstrained posterior. The author adopts the suggestion of Frühwirth-Schnatter (2004) to construct $q(\theta)$ from a discrete mixture of transition kernels. Specifically, a sample $(G^1, \ldots, G^S)$ with $G^i = (G_1^i, \ldots, G_T^i)$ is drawn from the unconstrained posterior. The importance density is:

$$q(\theta) = \left[\frac{1}{S} \sum_{i=1}^{S} k_1(\mathbf{p}, \mathbf{q} \mid G^i) k_2(\gamma_1, \gamma_2, \beta_{11}, \ldots, \beta_{2p} \mid y, G^i, \sigma_0^2) \right] k_3(\sigma^2) \qquad (3.3)$$

where $k_1(.)$ is the full conditional posterior of the transition probabilities, $k_2(.)$ is the full conditional posterior of the regression coefficients, $k_3(\sigma^2)$ is an inverted gamma whose moments match the empirical moments of the MCMC output for σ^2, and σ_0^2 is the previous draw from the full conditional posterior of σ^2. In practice, for a two-state model, setting $S = 20$ will suffice.

4. Bayesian specification search

We will now apply the methods of the two preceding sections to LSTAR and MSAR models of the unemployment rate. The dependent variable is $y_t = \ln[0.01 U_t / (1 - 0.01 U_t)]$, where U_t is the civilian male (over 20 years old) deseasonalized monthly US unemployment rate (in percentage points) from 1960:1 to 2004:12, taken from the LRMT20 series in the Haver Analytics USECON database.

Van Dijk *et al.* (2002) suggest using as transition variable the lagged seasonal difference of the unemployment rate. Indeed, Figure 1 reveals that the series $s_t = U_{t-1} - U_{t-13}$ closely reproduces the business cycle, with low values of s_t corresponding to expansions and high values of s_t to contractions. However, Van Dijk *et al.* (2002) do not use the logistic transformation of U_t, and

their estimates (equations 45 and 46 in their paper) imply an extreme lack of normality in the residuals, with a p-value of 5.9×10^{-5} for the Bera-Jarque statistic.

Table 1 reports the logarithms of the marginal likelihoods for AR, LSTAR, and two- and three-state MSAR models with dependent variable y_t, for various values of the autoregressive order p and of the transition delay parameter d in $s_t = U_{t-d} - U_{t-d-12}$. The prior parameters were as follows. For the MSAR models, the autoregression coefficients (β_{ij} in (3.1)) are independent standard normal, and the intercepts γ_i are independent $N(0, 0.01)$. For the LSTAR models, the prior on the intercepts (α_1, α_2) is bivariate normal with null expectation vector, variances of $V(\alpha_1) = 0.01$ and $V(\alpha_2) = 0.02$, and correlation coefficient $-1/\sqrt{2}$; and the priors on the autoregression coefficients (ϕ_{1j}, ϕ_{2j}) are bivariate normal with null expectation vector, variances of $V(\phi_{1j}) = 1$ and $V(\phi_{2j}) = 2$, and correlation coefficient $-1/\sqrt{2}$. This choice ensures that the priors on the regression coefficients are identical for the MSAR and the LSTAR models. The prior parameters on γ and c are $\gamma_a = 3$, $\sigma_\gamma^2 = 0.1$, $c_a = 0$, and $\sigma_c^2 = 1$. For the AR models, the prior on the intercept is $N(0, 0.01)$, and the priors on the autoregression coefficients are $N(0, 1)$. In all models, the inverted Gamma prior on σ^2 is almost improper, with $a = b = 10^{-6}$, and prior coefficient independence across covariates is assumed.

An examination of Table 1 confirms that bridge sampling is very efficient, especially for the AR and LSTAR models. Among the AR specifications, $p = 5$ is preferred. Among the LSTAR specifications, the evidence is in favor of $(p = 2, d = 1)$, with $(p = 3, d = 1)$ being a close contender. The Bayes factor in favor of the first of the two models is $\exp(956.206 - 955.722) = 1.623$; on the Jeffreys scale, the evidence against the less parsimonious model is "not worth more than a bare mention" (Jeffreys, 1961, Appendix B). Indeed, assuming prior odds of unity, the posterior probability of $(p = 2, d = 1)$ against $(p = 3, d = 1)$ is $1.623/2.623 = 0.619$. In the MSAR case, a two-state model with $p = 3$ is very clearly preferred.

Marginal likelihoods are well-known to be sensitive to the prior specification (and are not defined for improper priors). In Table 2, we therefore present some marginal likelihood estimates obtained when all the prior variances (except the one of σ^2) are doubled. The marginal likelihoods are indeed uniformly lower, but the ranking between models remains unchanged.

The conclusions that emerge from Tables 1 and 2 are clear. First, Bayes factors present very strong evidence against linearity. Secondly, the LSTAR model with $s_t = U_{t-1} - U_{t-13}$ is uniformly, and very strongly, preferred to the two- and three-state MSAR models. Other choices of s_t, such as $U_{t-1} - U_{t-d-1}$ for various values of d, yielded lower marginal likelihoods; the same was true for the transition variable $(y_{t-1} - y_{t-6})/5$, found by Koop and Potter (1999) to maximize the posterior odds in their TAR model, albeit for a different sample. An ESTAR model was also tried, with inferior results.

Even though care was taken to ensure comparable priors across models, an element of uncertainty due to prior sensitivity remains. Also, the marginal likelihood criterion is known to favor parsimony; if one treats the discrete latent variables in the MSAR model as unknown parameters, LSTAR becomes much more parsimonious than MSAR. In the author's opinion, an ultimate comparison of both models should therefore rely on complementary evidence. Such evidence might be provided by posterior predictive p-values; this is the topic of the next section.

5. Posterior predictive p-values.

Let $s_i(x, \theta)$, for $i = 1, \ldots, n$, be a collection of statistics commonly used for misspecification testing, x being a vector of data assumed to be generated by a given model with parameter vector θ. In the applications of this section, $s_i(x, \theta)$ will be based on the generalized residuals $u_t(x, \theta)$, defined as:

$$u_t(x, \theta) = \Phi^{-1}\left[P(X_t \leq x_t \mid x^{t-1}, \theta) \right] \tag{5.1}$$

where $\Phi(.)$ is the normal integral, and x^{t-1} contains all observations on x_s up to $t - 1$. If the probabilities on the right-hand side of (5.1) are indeed those implied by the process generating x, the generalized residuals are independent standard normal. In the LSTAR model, $u_t(x, \theta)$ is simply the standardized residual implied by (2.1). In the MSAR model (3.1), we have:

$$P[X_t \leq x_t \mid x^{t-1}, \theta] = \sum_{k=0}^{1} P[X_t \leq x_t \mid x^{t-1}, \theta, G_t = k] P[G_t = k \mid x^{t-1}, \theta] \tag{5.2}$$

where the conditional probabilities of the regimes are computed as in (3.2).

The predictive distribution of $s_i(x, \theta)$ can be readily simulated if a posterior sample from $p(\theta \mid y)$ is available: for each replication θ from this sample, one simply generates x by recursively simulating the model, computes the generalized residual series $\{u_t(x, \theta)\}_{t=1}^{T}$, and computes the resulting $s_i(x, \theta)$ for $i = 1, \ldots, n$. The posterior predictive p-value for statistic i is defined as:

$$P\left[s_i(x, \theta) > E[s_i(y, \theta) \mid y] \right]$$

and is estimated as the percentage of replications of $s_i(x, \theta)$ that exceed the posterior average of the values computed from the actual data. More details on this approach can be found in Gelman and Meng (1996) or Koop (2003).

We will compute three statistics $s_i(x, \theta)$ from the generalized residual series $\{u_t(x, \theta)\}_{t=1}^{T}$, namely:

(1) The Bera-Jarque statistic, used as an indicator of error non-normality;

(2) An F statistic for the nullity of the autoregression coefficients in an AR(12) model of the generalized residuals. This is used as an indicator of error autocorrelation, and is denoted by AC(12);

(3) An F statistic for the nullity of the autoregression coefficients in an AR(12) model of the squared generalized residuals. This is used as an indicator of error conditional heteroscedasticity, and is denoted by ARCH(12).

Table 3 presents the estimated posterior predictive p-values for each statistic and for the LSTAR and MSAR models that were found, in Section 4, to have the highest marginal likelihoods. The MSAR model was identified using the first autoregression coefficient: the constraint $\phi_{21} > 0$ in (3.1) was very clearly suggested by the output of the unconstrained permutation sampler. In each case, the estimated p-values are based on 10000 replications. The estimated predictive distributions of the statistics were quite close to the relevant asymptotic Chi-square and F distributions, suggesting that asymptotic approximations would be reasonable.

The estimated p-values in Table 3 are all larger than 0.01, with the exception of that for ARCH(12) in the LSTAR model with $p = 2$. So, there is weak evidence of conditional heteroscedasticity in this model.

Since, as discussed in Section 4, the posterior probabilities of the two preferred LSTAR formulations are nearly equal, it is legitimate to use other evidence, such as the one provided in this section, for discriminating between these two models (additional evidence in favor of $p = 3$ will be given in Section 8). There is another, more subjective, reason for preferring a three-lag LSTAR model: since our ultimate aim is the comparison of LSTAR and MSAR models, it makes sense to choose the same autoregressive order for both. It might otherwise be difficult to disentangle those differences that are due to model order from those that are due to model class.

For these reasons, and for the sake of brevity, we will concentrate in the sequel on the 2-state MSAR model with $p = 3$ and on the LSTAR model with $p = 3$ and $s_t = U_{t-1} - U_{t-13}$. However, we will occasionally mention results obtained with the two-lag LSTAR model when this appears noteworthy.

6. MCMC ESTIMATION

In this section, we will present Markov chain Monte Carlo estimates of the LSTAR model (2.1)–(2.2) with $p = 3$ and $s_t = U_{t-1} - U_{t-13}$, and of the MSAR model (3.1) with $p = 3$ and the identification constraint $\phi_{21} > 0$. The prior distributions are the same as those used for estimating the marginal likelihoods in Table 1, and were described in the third paragraph of Section 4.

In Table 4, θ_α denotes the estimated posterior quantile at probability α and s_θ is the estimated posterior standard error. The regression coefficient estimates for both models are generally similar, even though some differences can be noticed (in particular, LSTAR predicts regime switching in the second autoregression coefficient, but this is not the case for MSAR since the confidence interval for ϕ_{22} contains zero in the second model but not in the first). In the MSAR model,

$G_t = 0$ can be associated with expansion and $G_t = 1$ with contraction. In the LSTAR model, the data do not appear to be very informative on the shape parameter of the transition function, since the posterior results for γ almost reproduce the $N(3, 0.1)$ prior for this parameter: the 95% prior confidence interval is $[2.38, 3.62]$ and the p-value of a Bera-Jarque statistic for testing the normality of the posterior replications of γ is 0.39. This is not surprising, since a wide range of values of γ will typically lead to similar shapes of the transition function, as noted by Teräsvirta (1994); indeed, this reason is commonly invoked to explain the failure of likelihood maximization algorithms in STAR models.

In Figure 2, we present kernel density estimates of the marginal posteriors of the intercepts in both states (α_1 and $\alpha_1 + \alpha_2$) and of the sums of the autoregression coefficients in both states ($\sum_{j=1}^{3} \phi_{1j}$ and $\sum_{j=1}^{3} [\phi_{1j} + \phi_{2j}]$). Both models predict a unit root during expansions and a highly persistent process during contractions (the results obtained with the two-lag LSTAR model were almost identical). The presence of unit roots, or roots very close to unity, implies that it would be misleading to speak of an equilibrium unemployment rate in a particular regime.

It is also of interest to compare the estimated transition function of the LSTAR model with the smoothed probability of the second regime in the MSAR model, which is $P[G_t = 1 \mid y_1, \ldots, y_T]$. This probability can be estimated from the MCMC output by the percentage of replications of G_t having a value of unity. The two functions are plotted in Figure 3. They are generally similar, but the "transition function" of the MSAR model appears better able to anticipate turning points. This is not too surprising, since the smoothed probability is conditional on all the sample observations whereas the LSTAR transition function uses only past observations. For comparison, the NBER business cycles (available on www.nber.org/cycles.html, accessed on January 4, 2007) are also reported as shaded areas.

Finally, Figure 4 reports a kernel density estimate of the ergodic probability of contraction in the MSAR model, given by $(1 - \mathbf{p})/(2 - \mathbf{p} - \mathbf{q})$; of the expected duration of an expansion, given by $1/(1 - \mathbf{p})$; and of the expected duration of a contraction, given by $1/(1 - \mathbf{q})$. The posteriors of the expected durations are strongly leptokurtic and skewed to the right, with a median of approximately 30 months for expansions and 17 months for contractions.

The MSAR model (3.1) assumes that the innovation variance σ^2 is constant. In a previous version of this paper, the more general model described in Deschamps (2006) was estimated. This model assumes Student-t disturbances with regime-dependent scale parameters. In this more general model, the point estimates of the Student scale parameters are 0.00126 in expansions and 0.00125 in contractions; and the point estimate of the Student degrees of freedom is equal to 94. The point and interval posterior estimates for the other parameters (regression coefficients and transition probabilities) were almost identical to those in Table 4. Together with the predictive

p-values reported in Section 5, this confirms that a normal homoscedastic model is appropriate for this sample.

7. MAXIMUM LIKELIHOOD ESTIMATION

Table 5 presents maximum likelihood estimates of the preceding three-lag LSTAR model and two-state, three-lag MSAR model. The LSTAR model was estimated using E-Views 5.1; the MSAR model was estimated by the EM algorithm described in Smith *et al.* (2006), section 2.2. In order to ensure the convergence of this algorithm, it proved necessary to concentrate the likelihood with respect to the regression coefficients and equation variance, and to maximize the concentrated loglikelihood $L(\mathbf{p}, \mathbf{q})$ by a bivariate grid search on the square $[0.9, 0.99] \times [0.9, 0.99]$, with increments of 0.01 for each coordinate; this was followed by a more refined grid search on $[0.986, 0.995] \times [0.986, 0.995]$ with increments of 0.001.

The regression coefficient estimates in Table 5 are quite close to the corresponding ones in Table 4. The estimated standard error of $\hat{\gamma}$ in Table 5 is, however, larger than s_γ in Table 4, reflecting the relatively tight prior on γ. Another noteworthy observation is the near equality of the estimated persistence probabilities $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ in Table 5; furthermore, the estimates are larger than in Table 4. They imply nearly equal ergodic probabilities, and nearly equal expected durations for both cycles. This is at variance with the behavior exhibited in Figure 4, and conflicts somewhat with historical evidence. However, the smoothed probability graph was very similar to the one in Figure 3.

Figures 5 and 6 present distribution graphs and correlograms of the generalized residuals $u_t(y, \hat{\theta})$ defined in Section 5. Table 6 presents p-values of misspecification diagnostics computed from these residuals; the values for an AR(5) model are also reported. They confirm the previous results: there is evidence against linearity, and an LSTAR model with $p = 2$ appears slightly misspecified.

8. FORECAST EVALUATION

8.1. Predictive densities.

Simulating predictive densities from the LSTAR model is a straightforward exercise. Upon defining β as in (2.3), and:

$$F(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-13}, \beta, \gamma, c)$$

as the deterministic part of the right-hand side of (2.1) with $p = 3$ and s_t replaced by:

$$s_t^* = \left(\frac{1}{1 + \exp(-y_{t-1})} - \frac{1}{1 + \exp(-y_{t-13})} \right) \times 100,$$

13

we generate, for each posterior replication $(\beta^{(i)}, \gamma^{(i)}, c^{(i)}, \sigma^{(i)})$ and for $h = 1, \ldots, H$, draws $y_{T+h|T}^{(i)}$ from normal distributions with expectations $F(y_{T+h-1}^*, y_{T+h-2}^*, y_{T+h-3}^*, y_{T+h-13}^*, \beta^{(i)}, \gamma^{(i)}, c^{(i)})$ and standard error $\sigma^{(i)}$; here y_{T+h-j}^* is the observed value y_{T+h-j} if $j \geq h$, and the previously simulated value $y_{T+h-j|T}^{(i)}$ otherwise.

Simulating predictive densities from the MSAR model can be done by the method of composition (Albert and Chib, 1993). It involves generating simulated paths:

$$G_{T+1|T}^{(i)}, \ldots, G_{T+H|T}^{(i)}$$

from the Markov evolution equation, using as initial regime probabilities $\pi_T^{(i)}$ and $1 - \pi_T^{(i)}$, where:

$$\pi_T^{(i)} = P[G_T = 0 \mid y_1, \ldots, y_T, \theta^{(i)}]$$

can be computed by the filter described in Hamilton (1994, pp. 692–693), using as input the ith MCMC replication of the MSAR parameters:

$$\theta^{(i)} = (\gamma_1^{(i)}, \gamma_2^{(i)}, \beta_{11}^{(i)}, \ldots, \beta_{2p}^{(i)}, \mathbf{p}^{(i)}, \mathbf{q}^{(i)}, \sigma^{2(i)}).$$

Once the future path $(G_{T+h|T}^{(i)}, h = 1, \ldots, H)$ has been generated, it is a simple matter to generate predictions $(y_{T+h|T}^{(i)}, h = 1, \ldots, H)$ from equation (3.1), by a method analogous to that used for the LSTAR model.

It may be of interest to present an illustration involving two information sets. The first information set (S_1) consists of all observations from 1959:10 to 1998:12, the second (S_2) adds all observations through 2001:12. Each information set ends approximately eighteen months before a turning point. The author reestimated the previous MSAR and LSTAR models, as well as a benchmark AR(5) model of the logistic transform y_t, on the samples defined by S_1 and S_2. For the MSAR and LSTAR models, the prior described in Section 4 was used; the AR model was estimated with the Gibbs sampler and a fully diffuse prior.

In the top panels of Figure 7 (corresponding to S_1) and Figure 8 (corresponding to S_2), we plot the predictive medians of U_t obtained from all three models through 2005:3, together with the realized values. There are substantial differences in the long run predictive behavior of the three models. However, in each case, convergence to an equilibrium value is not manifest, reflecting the high persistence of the process.

The bottom panels of Figures 7 and 8 show the bounds of 95% highest density prediction intervals obtained from the LSTAR model, together with the medians and the observed values. The densities are reasonably symmetric at an horizon of one month, but exhibit considerable skewness at intermediate horizons. The lengths of the intervals turned out to be similar for all three models, and appear to be very conservative.

8.2. Classical forecast evaluation.

A formal investigation of the predictive performance of the MSAR, LSTAR, and AR models will now be presented. Even though the approach of this subsection is classical, the forecasts will be based on Bayesian predictive densities. We consider, in all, 295 information sets, each including a training sample consisting of all observations between 1959:10 and 1979:11. The t-th information set is obtained by adding t subsequent monthly observations to this training sample. We estimate the models on each information set, using the same priors as before, and simulate in each case six out of sample predictive densities at horizons of $h = 1, \ldots, 6$. The simulated densities are then analyzed using three complementary approaches, which are now discussed.

The first approach applies only to the 295 one-step forecasts ($h = 1$). We compute the percentage p_t of simulated predictions that are less than the observed value y_t, and compute the transform $\phi_t = \Phi^{-1}(p_t)$, where $\Phi(.)$ is the normal integral. The percentage p_t is of course a Monte Carlo estimate of the probability integral transform discussed in Diebold $et\ al.$ (1998); and the transforms ϕ_t should be independent standard normal under the correct predictive distribution, as emphasized by Berkowitz (2001). This can be investigated by Kolmogorov-Smirnov tests and by regressing powers of ϕ_t on their past values.

The second approach, proposed by Diebold and Mariano (1995), is based on a comparison of the mean absolute prediction errors, defined as $\sum_{t=1}^{295} \mid y_t - \hat{y}_{t,h,i} \mid /295$, where $\hat{y}_{t,h,i}$ is the median h-step forecast of y_t using model i, based on the information set available at $t - h$. We test the nullity of the expected loss differentials, using heteroscedasticity and autocorrelation consistent (HAC) variance estimates. Specifically, for $(i, j) = (1, 2), (1, 3), (2, 3)$ and for $h = 1, \ldots, 6$, we regress the series:

$$D_{t,ijh} = \mid y_t - \hat{y}_{t,h,i} \mid - \mid y_t - \hat{y}_{t,h,j} \mid$$

on a constant term c_{ijh}, and estimate the variance of $\hat{c}_{ijh}$ by the method of Andrews (1991), using a Bartlett kernel and VAR(1) prewhitening. We then test the null hypothesis that $c_{ijh} = 0$ against a two-sided alternative, using the t statistic. When $h = 1$, it is well known that the forecast errors based on the true model and the true parameter values are independent. Since this is not necessarily so when the forecasts are based on the true model with unknown parameters, however, a semiparametric variance estimate is appropriate even for one-step forecasts.

The third approach is based on the six regressions of y_t on a constant term and on the three covariates $\hat{y}_{t,h,1}$, $\hat{y}_{t,h,2}$, $\hat{y}_{t,h,3}$, for each horizon $h = 1, \ldots, 6$. For the reasons given in the previous paragraph, HAC variance estimation is appropriate, even when $h = 1$. Following West and McCracken (1998), predictor i will be called efficient relative to the predictors $j \neq i$ at horizon h if a robust F test of $y_t = \hat{y}_{t,h,i} + \epsilon_t$ against the unrestricted model does not reject the null hypothesis.

The results are in Tables 7, 8, 9, and 10. The diagnostics in Table 7 strongly indicate that the AR(5) model is misspecified; however, none of the diagnostics for the MSAR and LSTAR models are significant at the 1% level. In Table 8, the mean absolute prediction errors are systematically lowest for the LSTAR model and highest for the AR model; however, none of the Diebold-Mariano statistics are significant.

The clearest evidence is provided by the efficiency tests in Tables 9 and 10. The LSTAR model is the only one which passes the test at all horizons. It is also noteworthy that according to this efficiency benchmark, the AR model actually performs better than MSAR at horizons 4, 5, and 6. A comparison between the last columns of Tables 9 and 10 leads to the choice of $p = 3$ rather than $p = 2$ for the LSTAR model.

The preceding results suggest that all three models can provide a fairly good description of typical unemployment rate data, so that the nonlinear behavior implied by LSTAR and MSAR can be difficult to detect in practice. They also suggest that the first two approaches lack the power needed to discriminate between the LSTAR and MSAR models, at least for the modest sample sizes typically available in macroeconomic data. However, they suggest that the efficiency test is likely to be an effective tool for model discrimination in our case. Finally, the results suggest that studying predictions at horizons greater than one month is of somewhat marginal additional value; in the opinion of the author, this is due to the high persistence of the process that is investigated, which leads to very high correlations between successive forecasts at horizons larger than one, resulting in loss of power.

8.3. Bayesian forecast evaluation.

A Bayesian implementation of the tests presented in Section 8.2 involves generating, in the spirit of Section 5, simulated series of unemployment data using posterior parameter samples generated by MCMC. One would then compute the statistics of Section 8.2 for each simulated data sample, and estimate predictive p-values.

In order to compute the efficiency statistics of Section 8.2, it is necessary to estimate all models on each information set. If 1000 data samples are generated from LSTAR processes and 1000 from MSAR processes, this would necessitate $3 \times 2000 \times 300 = 1800000$ MCMC simulations when 300 forecasts must be generated using 3 different models. This is clearly infeasible.

However, a "poor man's version" of the statistics of Section 8.2 is available. If, as the information set expands, one does not update the posteriors on which the predictives are based, the number of required MCMC simulations is equal to 6000 rather than the previous one, and is well within the capabilities of a recent workstation (optimized FORTRAN programs were used). Since some of the available information is excluded in computing the forecasts, there is no particular

reason for classical p-values to be larger than the size of an efficiency test, even if the predictive model is correct; but this becomes irrelevant in the Bayesian context, since the actual distribution of the statistic is simulated. For the same reason, it becomes unnecessary to compute HAC versions of the efficiency or Diebold-Mariano statistics.

The author selected two samples of size 1000 from the 10000 posterior LSTAR and MSAR replications summarized in Table 4, and generated 2000 corresponding unemployment series ranging from 1959:10 to 2004:12. The MSAR(3), LSTAR(3), and AR(5) models were then estimated by MCMC using the first 243 observations of each artificial sample. One-step ahead median forecasts of the 300 remaining observations, ranging from 1980:1 to 2004:12, were then generated from each model, using expanding information sets but without updating the posteriors.

The estimated predictives of the efficiency and Diebold-Mariano statistics turn out to exhibit high asymmetry and extreme outliers. Unfortunately, 95%, and even 90%, highest density intervals are too wide to be useful. However, conventional boxplots are informative when outliers are omitted. The three panels of Figure 9 present the diagrams corresponding to the MSAR and LSTAR efficiency statistics and to the squared LSTAR-MSAR Diebold-Mariano statistic (at horizon 1); within each panel, the left-hand box corresponds to the MSAR data generating process (DGP) and the right-hand box to the LSTAR one. The vertical separation of the boxes within each of the top two panels can be given an interpretation similar to the classical notion of power, which is the probability under the alternative that the statistic exceeds a quantile of its distribution under the null.

In the top two panels of Figure 9, conventional credible intervals overlap to a very large extent, so that none of the "poor man's efficiency tests" can be called powerful. While this may in part be due to the lack of posterior updating, it may also reflect the fact that the ability of forecast evaluation to discriminate between competing models is only asymptotic (as pointed out by a referee). Nevertheless, concentrating on the interquartile ranges, bounded by the edges of the boxes in Figure 9, allows clear conclusions to emerge. First, it appears easier for the MSAR model to emulate the LSTAR DGP than the converse: the vertical separation of the boxes is larger in the middle panel than in the top panel, and the squared Diebold-Mariano statistic (in the bottom panel) tends to be lower under the STAR DGP. Secondly, it is only under the STAR DGP that the observed values of the three statistics fall within the interquartile predictive ranges. This observation implicitly sets the size of a classical test equal to 25%; however, many classical statisticians would agree that choosing a large value for size is legitimate when the test is known to lack power.

Bayesian forecast evaluation thus confirms the results obtained with classical tests: since the observed value of the squared Diebold-Mariano statistic does not exceed the third quartile of either

17

predictive distribution in the bottom panel of Figure 9, this test is not useful for discriminating between the two models; however, both the MSAR and LSTAR efficiency tests point to the predictive superiority of the LSTAR model.

9. DISCUSSION AND CONCLUSIONS

This paper has attempted to compare the performance of LSTAR, MSAR, and AR models for forecasting the monthly US unemployment rate. Unlike many past contributions, it has used fully Bayesian methods for estimating predictive densities.

The LSTAR model incorporates strong prior knowledge on the factors determining the onset of transitions between regimes, in the form of a particular transition function and transition variable; by contrast, in the MSAR model, such prior knowledge only consists in a flexible evolution equation. It is therefore not surprising that *an appropriate* LSTAR model should make better use of available information and perform better on predictive efficiency tests. The fact that both models yield insignificant misspecification diagnostics and have insignificant mean loss differentials, however, is an additional argument in favor of the particular LSTAR model used in this paper: if two formulations with different prior assumptions yield results that are essentially similar, our confidence in the stronger assumptions increases, and one will tend to favor the more structured alternative. In this sense, the MSAR and LSTAR approaches can be said to be cross-validating and complementary.

It has been pointed out by a referee that the basic MSAR model can be extended to specify transition probabilities that depend on past observables. A simple formulation would model the transition probabilities as logistic normal processes. By a suitable choice of parameters, a logistic normal can approximate a Beta distribution; this leads to a natural candidate-generating density in a Metropolis-Hastings step. This was tried by the author, but did not result in a well-behaved MCMC sampler. The reason appears to be due to the shape of the logistic function, which is nearly constant over a large range of argument values. A more sophisticated approach, followed by Filardo and Gordon (1998), uses a latent probit process to model the transition probabilities and appears to work well in the MCMC context. Unfortunately, this generalization obviously complicates an MCMC sampler, the bridge sampling estimation of marginal likelihoods, and the estimation of predictive densities; it is not certain that the primary aims of this paper could have been achieved with the more general model. In the opinion of the author, the more limited objective that consists in comparing two relatively simple formulations that clearly differ in their prior assumptions on structural change is also interesting, if only to validate the choice of a particular LSTAR transition variable: indeed, in Table 1, the ranking between the MSAR and LSTAR models would have been reversed if $d = 4$ rather than $d = 1$ had been chosen.

The data used in this paper are available from January 1948. When the MSAR model is estimated using the full sample from 1948 to 2004, a third regime can be identified. In this regime, the equation variance is approximately six times higher than in the other two; and the autoregression coefficients take values that are intermediate between those for expansions and contractions. The probability that this third regime occurs after 1960, however, is estimated to be less than 20% in any period, most estimates being very close to zero. Adding the pre-1960 data to each of the 295 information sets defined in Section 8.2 and using the three-regime MSAR model for conducting the simulated prediction exercises improves neither efficiency nor mean absolute prediction errors. For this reason, the data prior to 1960 were considered to be mostly of historical interest, and the results obtained with the extended MSAR model are not reported.

The evidence in favor of nonlinear dynamics appears to be stronger in this paper than in previous contributions. This may be partly due to our use of a monthly data frequency, rather than the quarterly frequency in, e.g., Rothman (1998) and Montgomery *et al.* (1998). Indeed, in another context, Klaassen (2005) finds that using lower frequency data can mask evidence in favor of regime switching.

This paper has also illustrated the importance of allowing *all* the parameters of a hidden Markov model of the unemployment rate to switch between regimes. This can be seen by comparing our results with those obtained by Bianchi and Zoega (1998). By allowing only the intercept term and the equation variance to vary across regimes, these authors all but eliminate the possibility of detecting Markov switching behavior in the US unemployment rate, and conclude in favor of a one-state model for this series; this conclusion is at considerable variance with the results of the present study.

Finally, this paper has illustrated the potential of regressing observations on out of sample point predictions for discriminating between non-nested models. It is perhaps surprising that this simple technique, which appears to have been originally suggested by Granger and Ramanathan (1984), has not been used more often in the literature.

References

Albert J, Chib S. 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* **11**: 1–15.

Andrews DWK. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.

Berkowitz J. 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* **19**: 465–474. DOI: 10.1198/07350010152596718

Bianchi M, Zoega G. 1998. Unemployment persistence: does the size of the shock matter? *Journal of Applied Econometrics* **13**: 283–304. DOI: 10.1002/(SICI)1099-1255(199805/06)13:3<283::AID-JAE469>3.0.CO;2-S

Chan F, McAleer M. 2002. Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *Journal of Applied Econometrics* **17**: 509–534. DOI: 10.1002/jae.686

Chan F, McAleer M. 2003. Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics* **13**: 581–592. DOI: 10.1080/0960310022000029295

Chib S. 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75**: 79–97.

Chib S, Jeliazkov I. 2001. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96**: 270–281. DOI: 10.1198/016214501750332848

Deschamps PJ. 2006. A flexible prior distribution for Markov switching autoregressions with Student-t errors. *Journal of Econometrics* **133**: 153–190. DOI: 10.1016/j.jeconom.2005.03.012

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Diebold FX, Gunther TA, Tay AS. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39**: 863–883.

Filardo AJ, Gordon SF. 1998. Business cycle durations. *Journal of Econometrics* **85**: 99–123.

Frühwirth-Schnatter S. 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96**: 194–209. DOI: 10.1198/016214501750333063

Frühwirth-Schnatter S. 2004. Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal* **7**: 143–167. DOI: 10.1111/j.1368-423X.2004.00125.x

Frühwirth-Schnatter S. 2006. *Finite Mixture and Markov Switching Models*. Springer: New York.

Gelfand AE, Dey DK. 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, B,* **56**: 501–514.

Gelman A, Meng XL. 1996. Model checking and model improvement. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman and Hall: London.

Granger CWJ, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3**: 197–204.

Hamilton JD. 1994. *Time Series Analysis*. Princeton University Press: Princeton.

Jeffreys H. 1961. *Theory of Probability (3d. edition)*. Oxford University Press: Oxford.

Klaassen F. 2005. Long swings in exchange rates: are they really in the data? *Journal of Business and Economic Statistics* **23**: 87–95. DOI: 10.1198/073500104000000505

Koop G, Potter SM. 1999. Dynamic asymmetries in U.S. unemployment. *Journal of Business and Economic Statistics* **17**: 298–312.

Koop G. 2003. *Bayesian Econometrics*. Wiley: Chichester.

Lopes HF, Salazar E. 2006. Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis* **27**: 99–117. DOI: 10.1111/j.1467-9892.2005.00455.x

Meng XL, Wong WH. 1996. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* **6**: 831–860.

Mira A, Nicholls G. 2004. Bridge estimation of the probability density at a point. *Statistica Sinica* **14**: 603–612.

Montgomery AL, Zarnowitz V, Tsay RS, Tiao GC. 1998. Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association* **93**: 478–493.

Rothman P. 1998. Forecasting asymmetric unemployment rates. *Review of Economics and Statistics* **80**: 164–168.

Smith A, Naik PA, Tsai CL. 2006. Markov-switching model selection using Kullback-Leibler divergence. *Journal of Econometrics* **134**: 553–577. DOI: 10.1016/j.jeconom.2005.07.005

Teräsvirta T. 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* **89**: 208–218.

Tierney, L. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**: 1701–1762.

Van Dijk D, Teräsvirta T, Franses PH. 2002. Smooth transition autoregressive models — a survey of recent developments. *Econometric Reviews* **21**: 1–47. DOI: 10.1081/ETC-120008723

West KD, McCracken MW. 1998. Regression-based tests of predictive ability. *International Economic Review* **39**: 817–840.

		p					
		1	2	3	4	5	6
AR		919.566	916.410	927.651	937.031	**937.795**	935.637
		0.000	0.001	0.001	0.001	0.001	0.001
2-state MSAR		932.093	946.035	**949.597**	945.624	941.595	937.158
		0.008	0.008	0.010	0.016	0.017	0.034
3-state MSAR		934.640	941.001	943.708	939.379	934.560	929.930
		0.040	0.158	0.085	0.229	0.063	0.078
LSTAR	$d = 1$	935.575	**956.206**	955.722	952.162	946.854	941.455
		0.004	0.005	0.006	0.006	0.006	0.007
	$d = 2$	936.308	954.752	953.807	951.360	946.043	940.856
		0.005	0.005	0.006	0.006	0.006	0.007
	$d = 3$	928.488	950.139	951.096	948.494	943.463	938.140
		0.005	0.005	0.005	0.006	0.006	0.006
	$d = 4$	926.544	945.231	947.103	946.198	941.172	935.886
		0.005	0.005	0.005	0.006	0.006	0.006
	$d = 5$	922.249	938.212	942.523	941.437	937.365	932.142
		0.006	0.005	0.005	0.006	0.006	0.006
	$d = 6$	919.496	933.870	938.681	938.802	934.502	929.884
		0.007	0.005	0.007	0.009	0.012	0.017

TABLE 2. *Logarithmic marginal likelihoods and*
numerical standard errors with looser priors

		p					
		1	2	3	4	5	6
AR		919.136	915.639	926.566	935.586	**935.997**	933.493
		0.000	0.001	0.001	0.001	0.001	0.001
2-state MSAR		931.312	944.596	**947.571**	942.833	938.350	932.123
		0.008	0.011	0.009	0.014	0.020	0.038
3-state MSAR		933.653	938.641	940.267	935.681	929.097	928.811
		0.037	0.065	0.246	0.137	0.163	0.145
	$d=1$	934.715	**954.603**	953.354	949.074	943.098	937.012
		0.005	0.005	0.006	0.007	0.007	0.007
LSTAR	$d=2$	935.410	953.186	951.411	948.264	942.274	936.393
		0.006	0.005	0.006	0.006	0.007	0.007
	$d=3$	927.619	948.612	948.802	945.452	939.741	933.724
		0.006	0.005	0.005	0.006	0.006	0.006

TABLE 3. *Posterior predictive p-values*

	Bera-Jarque	AC(12)	ARCH(12)
LSTAR $(p=2, d=1)$	0.6718	0.0935	0.0066
LSTAR $(p=3, d=1)$	0.5872	0.0681	0.0254
2-state MSAR $(p=3)$	0.7330	0.0149	0.0468

TABLE 4. *Posterior replication summaries for the LSTAR and MSAR models*

θ	LSTAR				MSAR			
	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	s_θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	s_θ
α_1	-0.082	-0.045	-0.006	0.019	-0.078	-0.034	0.013	0.023
ϕ_{11}	0.525	0.645	0.761	0.060	0.458	0.588	0.712	0.065
ϕ_{12}	0.165	0.291	0.417	0.064	0.126	0.255	0.380	0.065
ϕ_{13}	-0.055	0.054	0.164	0.056	0.027	0.151	0.280	0.065
α_2	-0.115	-0.045	0.023	0.035	-0.117	-0.052	0.019	0.034
ϕ_{21}	0.384	0.583	0.779	0.100	0.340	0.539	0.743	0.102
ϕ_{22}	-0.591	-0.312	-0.027	0.143	-0.362	-0.098	0.142	0.127
ϕ_{23}	-0.487	-0.294	-0.101	0.098	-0.656	-0.467	-0.261	0.100
$\sigma^2 \times 100$	0.128	0.145	0.163	0.009	0.114	0.130	0.148	0.009
γ	2.463	3.006	3.567	0.282				
c	0.043	0.188	0.326	0.072				
p					0.921	0.966	0.987	0.017
q					0.868	0.941	0.978	0.029

TABLE 5. *Maximum likelihood estimates of the LSTAR and MSAR models*

	LSTAR		MSAR	
θ	$\hat{\theta}$	$s_{\hat{\theta}}$	$\hat{\theta}$	$s_{\hat{\theta}}$
α_1	-0.047	0.020	-0.031	0.020
ϕ_{11}	0.651	0.061	0.582	0.056
ϕ_{12}	0.291	0.063	0.251	0.064
ϕ_{13}	0.046	0.056	0.162	0.056
α_2	-0.050	0.038	-0.054	0.029
ϕ_{21}	0.577	0.105	0.536	0.079
ϕ_{22}	-0.323	0.146	-0.101	0.105
ϕ_{23}	-0.279	0.100	-0.462	0.078
$\sigma^2 \times 100$	0.145	0.009	0.132	0.008
γ	3.153	0.850		
c	0.195	0.067		
p			0.990	0.006
q			0.991	0.006

TABLE 6. *P-values of misspecification diagnostics (maximum likelihood)*

	Bera-Jarque	AC(12)	ARCH(12)
AR ($p = 5$)	0.0200	0.1013	0.0008
LSTAR ($p = 2$)	0.6112	0.1068	0.0096
LSTAR ($p = 3$)	0.6252	0.1279	0.0484
MSAR ($p = 3$)	0.7602	0.0171	0.0476

TABLE 7. *Analysis of probability integral transforms (p-values)*

	Kolmogorov-Smirnov	AC(12)	ARCH(12)
AR ($p = 5$)	0.2147	0.0020	0.0005
LSTAR ($p = 2$)	0.0697	0.0861	0.0329
LSTAR ($p = 3$)	0.1182	0.0449	0.0151
MSAR ($p = 3$)	0.3066	0.0462	0.0307

TABLE 8. *Diebold-Mariano tests*

	Mean absolute prediction errors			P-values		
Horizon	LSTAR(3)	MSAR(3)	AR(5)	AR-MSAR	AR-LSTAR	LSTAR-MSAR
1	0.0283	0.0287	0.0291	0.449	0.169	0.294
2	0.0360	0.0364	0.0371	0.455	0.343	0.596
3	0.0452	0.0466	0.0475	0.621	0.262	0.276
4	0.0549	0.0572	0.0577	0.793	0.343	0.246
5	0.0642	0.0670	0.0680	0.715	0.317	0.264
6	0.0724	0.0762	0.0773	0.761	0.328	0.303

TABLE 9. *Efficiency tests with 3-lag LSTAR model*

	Statistics			P-values		
Horizon	AR(5)	MSAR(3)	LSTAR(3)	AR(5)	MSAR(3)	LSTAR(3)
1	11.353	4.754	1.462	0.000	0.001	0.214
2	5.066	2.459	1.431	0.001	0.046	0.224
3	3.305	2.832	1.519	0.011	0.025	0.197
4	2.591	3.083	1.685	0.037	0.017	0.154
5	2.250	3.281	1.847	0.064	0.012	0.120
6	1.794	3.078	1.871	0.130	0.017	0.116

TABLE 10. *Efficiency tests with 2-lag LSTAR model*

	Statistics			P-values		
Horizon	AR(5)	MSAR(3)	LSTAR(2)	AR(5)	MSAR(3)	LSTAR(2)
1	11.333	5.226	3.077	0.000	0.000	0.017
2	5.182	2.564	2.198	0.001	0.039	0.069
3	3.389	2.560	2.090	0.010	0.039	0.082
4	2.742	3.036	1.916	0.029	0.018	0.108
5	2.363	3.079	2.152	0.053	0.017	0.075
6	1.963	2.988	2.016	0.100	0.019	0.092

Figure 1. Unemployment series

Unemployment rate U(t)

y(t)=ln[0.01 U(t)/(1-0.01 U(t))]

s(t)=U(t-1)-U(t-13)

28
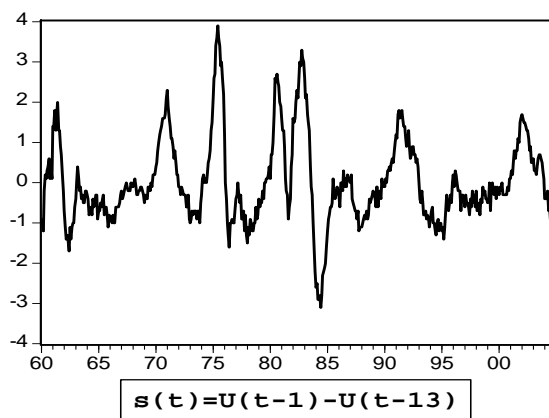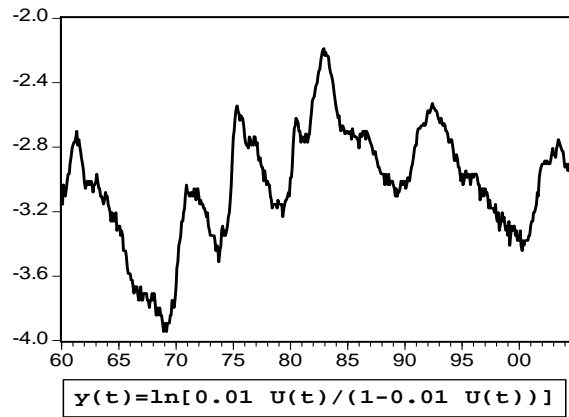
Figure 2. Kernel density estimates

EXPANSION   CONTRACTION

MSAR, constant term

MSAR, sum of autoregression coefficients

LSTAR,   constant term

LSTAR, sum of autoregression coefficients
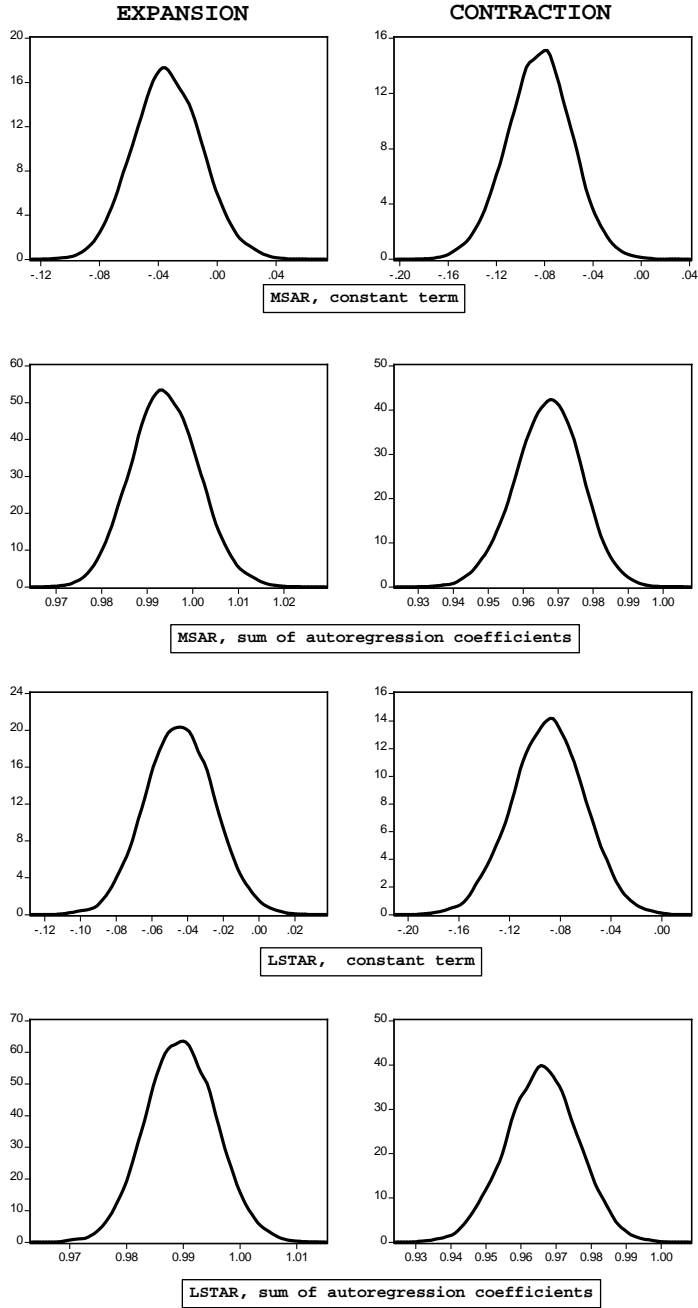
Figure 3. Transition functions

MSAR smoothed contraction probability

LSTAR transition function

30
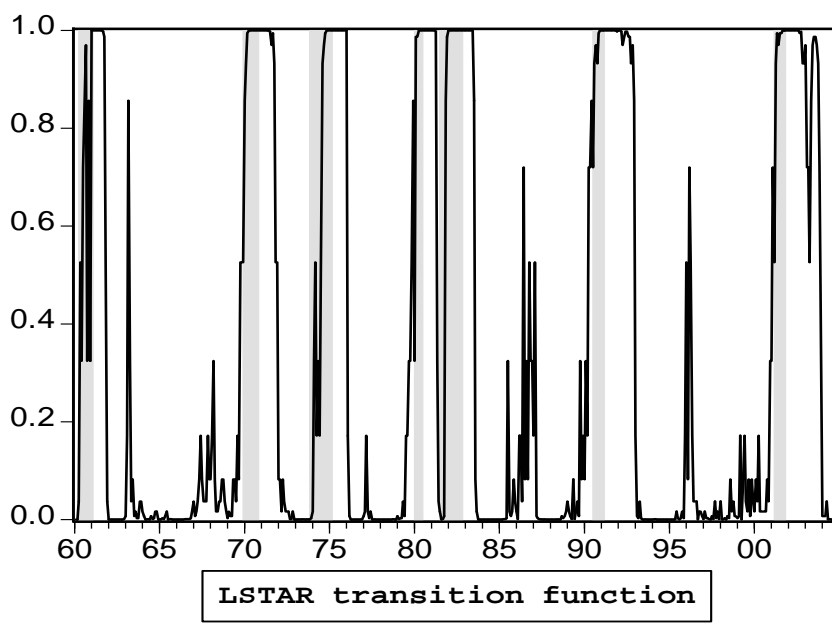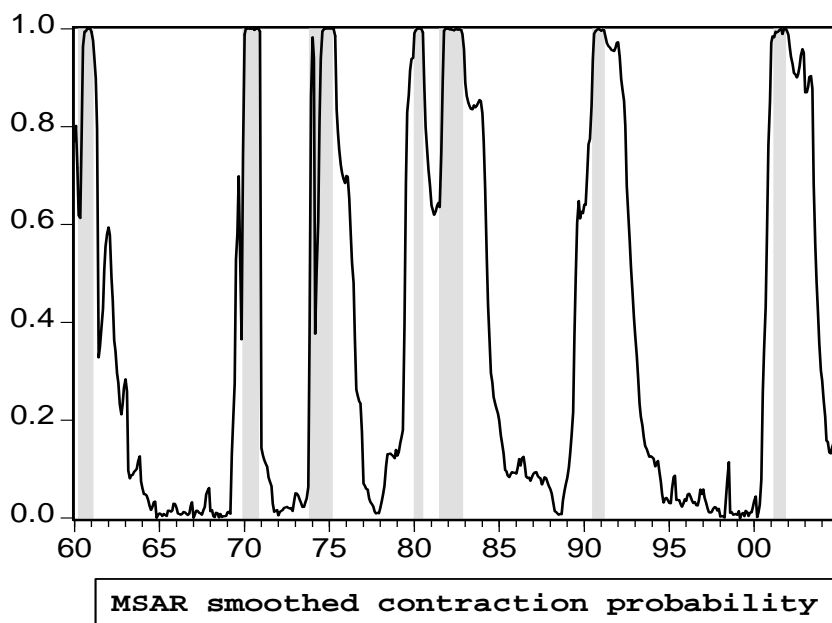
Figure 4. Long run implications of MSAR

Ergodic probability of contraction

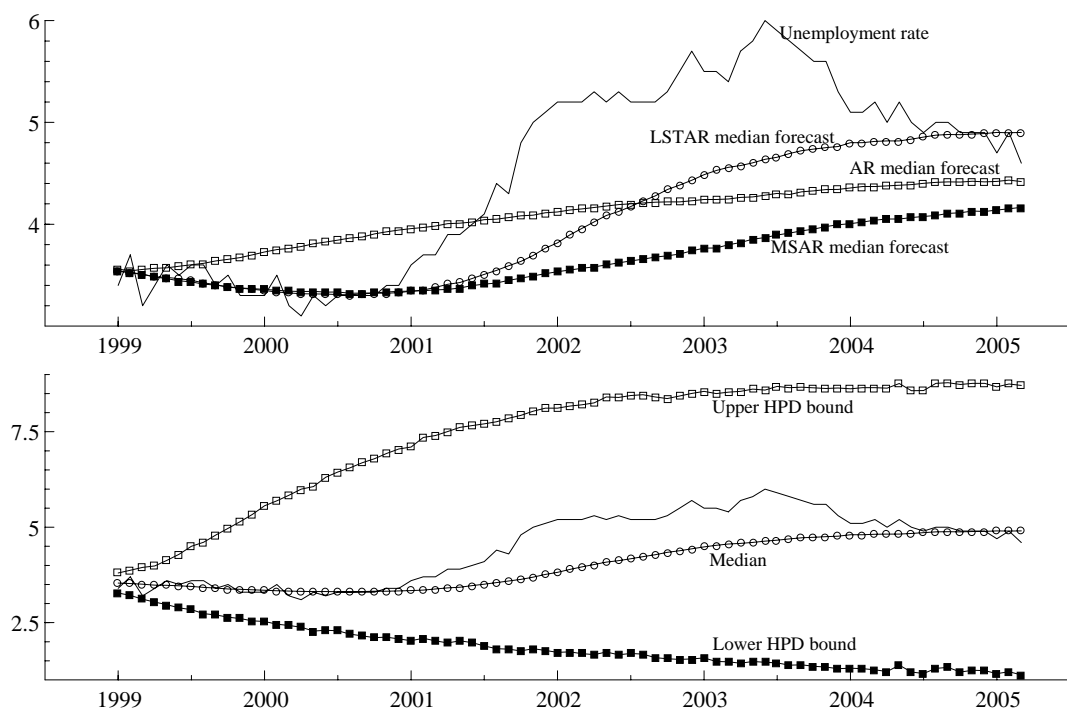Expected duration of expansion

Expected duration of contraction

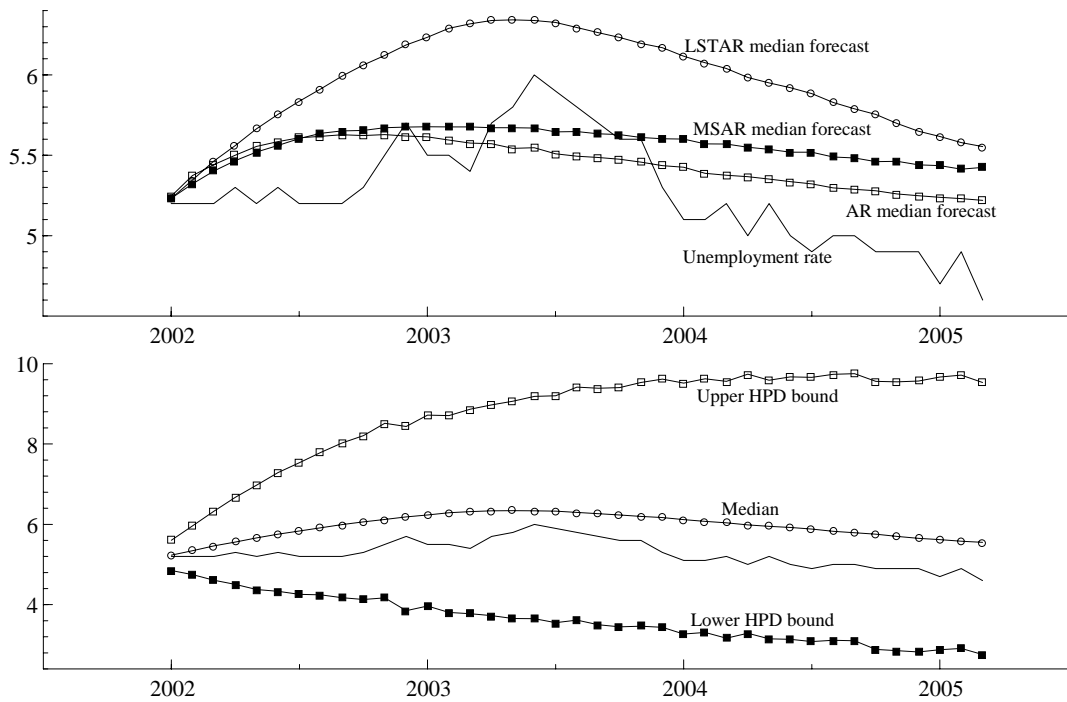31
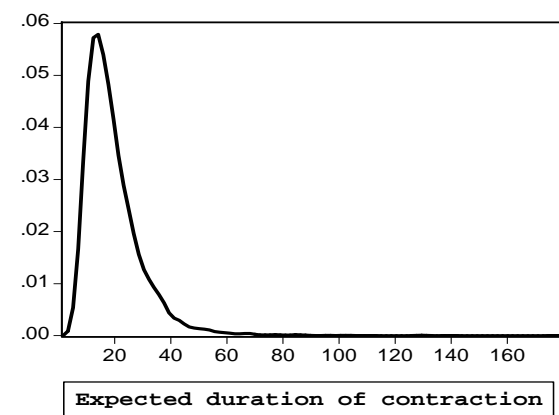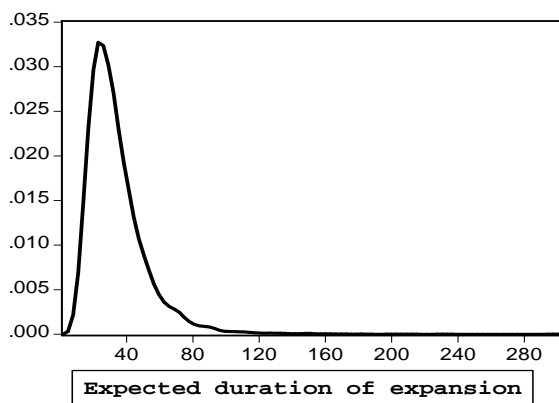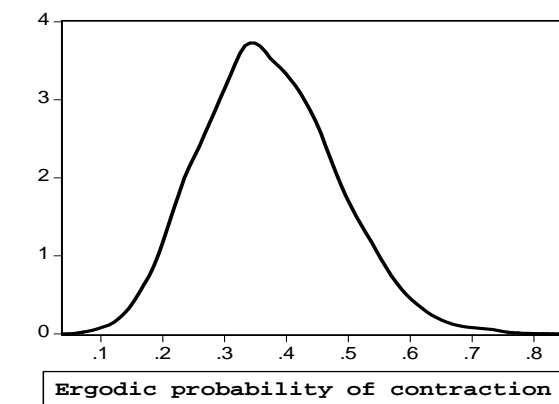
Figure 5. LSTAR (p=3) maximum likelihood residuals
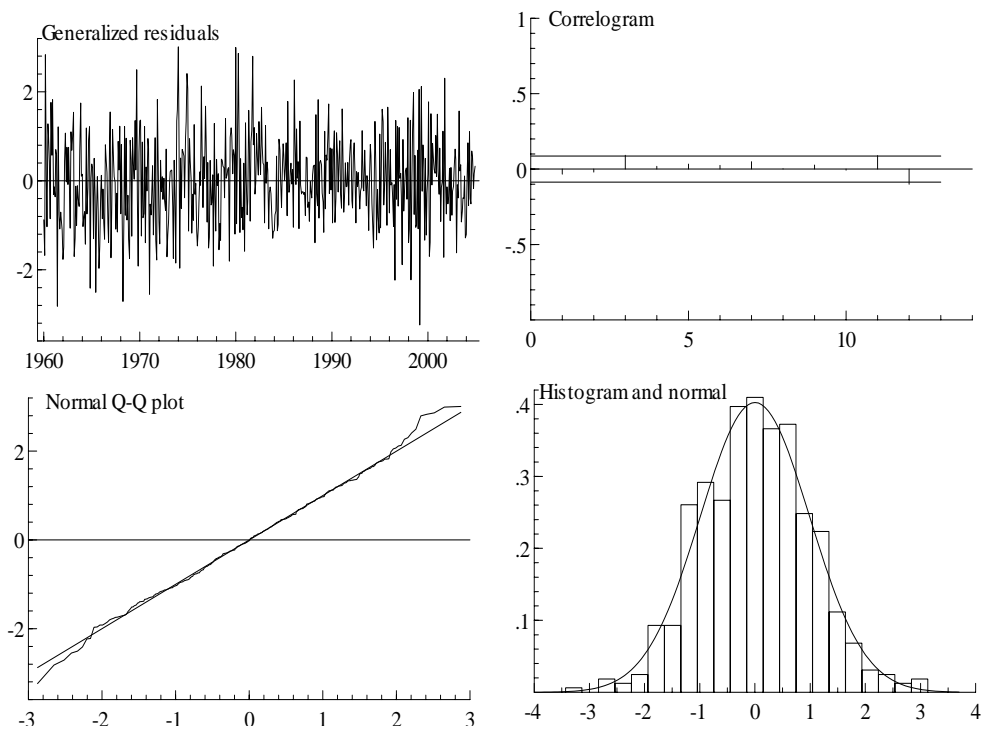
Figure 6. MSAR maximum likelihood residuals
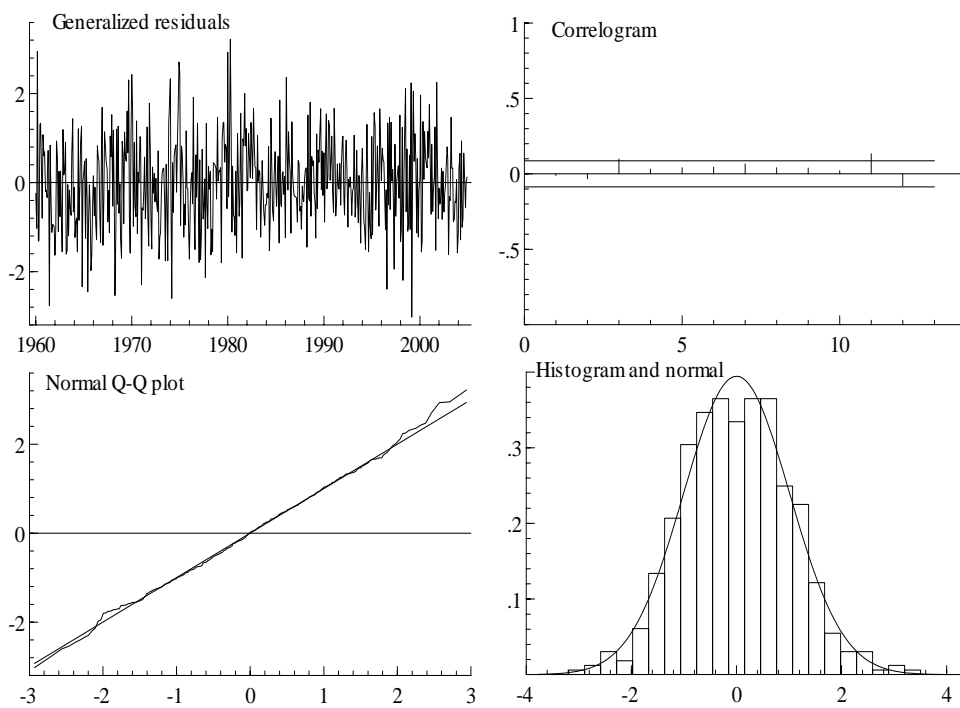
Figure 7. Median forecasts and LSTAR prediction intervals (information set S$_1$)

Figure 8. Median forecasts and LSTAR prediction intervals (information set S_2)

Figure 9. Bayesian forecast evaluation