A Nonlinear Long Memory Model for US Unemployment*

Dick van Dijk†
Econometric Institute
Erasmus University Rotterdam

Philip Hans Franses‡
Econometric Institute
Erasmus University Rotterdam

Richard Paap§
RIBES
Erasmus University Rotterdam

Econometric Institute Research Report EI2000-30/A

September 2000

Abstract

Two important empirical features of monthly US unemployment are that shocks to the series seem rather persistent and that unemployment seems to rise faster in recessions than that it falls during expansions. To jointly capture these features of long memory and nonlinearity, respectively, we put forward a new time series model and evaluate its empirical performance. We find that the model describes the data rather well and that it outperforms related competitive models on various measures of fit.

Keywords: fractional integration, smooth transition autoregression, time series model specification.

JEL Classification: C22, C51, C52, E24, E32.

---

*This paper was prepared for the Cardiff Conference on Long Memory and Nonlinear Time Series, July 2000. Helpful comments and suggestions from conference participants are gratefully acknowledged. The third author thanks the Netherlands Organization for Scientific Research (N.W.O.) for its financial support.

†Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: djvandijk@few.eur.nl` (corresponding author)

‡Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: franses@few.eur.nl`

§Erasmus University Rotterdam, Rotterdam Institute for Business Economic Studies, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: paap@few.eur.nl`

1 Introduction

Two key features of US unemployment, which are well documented in the literature, are that shocks to the series seem rather persistent and that it seems to rise faster in recessions than that it falls during expansions. The first feature is commonly coined as the long memory feature, see Diebold and Rudebusch (1989), Tschernig and Zimmermann (1992), Koustas and Veloce (1996) and Crato and Rothman (1998). The time series model that is used in these studies to describe this feature usually belongs to the class of fractionally integrated [FI] time series models, see Granger and Joyeux (1980) and Hosking (1981). The second feature is commonly called nonlinearity, see Montgomery *et al.* (1998), Koop and Potter (1999), and Skalin and Teräsvirta (2000), to mention just a few studies. The time series models which are considered most often for describing and forecasting the nonlinear properties of unemployment are either of the Markov-switching type, see Hamilton (1989), or of the threshold autoregressive [TAR], see Tong (1990), or smooth transition autoregressive [STAR] type, see Granger and Teräsvirta (1993), Teräsvirta (1994, 1998) and Franses, Van Dijk and Teräsvirta (2000). These three model classes assume the presence of two or more regimes, within which the data require different (linear) models for description and forecasting.

Interestingly enough, at least as far as we know, there have been no attempts to combine the features of long memory and nonlinearity into a single model, which then would be applied to US unemployment data. Such an attempt would be rather useful, as there are by now several studies, which document that one may fit a long-memory model to nonlinear data, and the other way around. For example, Granger and Teräsvirta (1999) and Davidson (2000) demonstrate that nonlinear models may generate data to which one may want to fit linear long-memory models. At the same time, data from long-memory models may seem nonlinear enough to try and fit certain nonlinear models to these observations, see Andersson, Eklund and Lyhagen (1999). As occasional structural breaks amount to a rather stylized nonlinear model (see Koop and Potter (2000) and Lundbergh, Teräsvirta and van Dijk (2000) for discussions on structural change and nonlinearity), one may expect that neglecting structural breaks also spuriously suggests the presence of the long memory property, see Bos, Franses and Ooms (1999), Diebold and Inoue (1999) and Granger and Hyung (2000). In fact, linear models which allow for occasional structural breaks have been considered for unemployment rates, see Bianchi and Zoega (1998), Papell, Murray and Ghiblawi (2000) and Coakley, Fuertes and Zoega (2000). To summarize, it seems worthwhile to try to capture both the features of long-memory and nonlinearity into a single time series model in order to be able to assess their relative importance. Within the context of such a model, one may then decide to see if nested alternative models perform better on measures of fit or forecasting. To that aim, in this paper we put forward such a model, where we decide to combine the concepts of fractional integration and smooth transition nonlinearity, as this combination seems rather straightforward and also as the resultant model appears fairly easy to analyze.

The outline of our paper is as follows. In Section 2, we discuss the representation of the model, which we will label the Fractionally Integrated Smooth Transition Autoregression [FISTAR]. In Section 3, we focus on a useful empirical strategy to fit this model to actual data. This section contains a discussion on tests for nonlinearity in FI models, estimation of FISTAR models and tests for model adequacy. In Section 4, we try to fit this model to monthly US unemployment data, as for these data it has been well-documented that long-memory and nonlinear features seems to exist, as argued above. We compare the FISTAR model with a FI model and with various STAR models, and we find that the FISTAR model has various desirable properties. We also display impulse response fucntions in order to grasp the empirical implications of the model. In Section 5, we conclude the paper with a range of possible topics for further research.

2 The Fractionally Integrated Smooth Transition Autoregressive Model

A model that allows for long memory in an observed time series y_t basically builds on

$$(1-L)^d y_t = x_t, \tag{1}$$

where x_t is a covariance-stationary process, e.g., an autoregressive moving average [ARMA] process, and the parameter d is possibly non-integer, in which case the time series y_t is called fractionally integrated [FI], see Granger and Joyeux (1980) and Hosking (1981). For non-integer d, the fractional differencing operator $(1-L)^d$ is defined by the binomial expansion,

$$(1-L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \cdots .$$

The series y_t is covariance stationary if $d < 0.5$ and invertible if $d > -0.5$. The autocorrelation function of y_t does not decline at an exponential rate, as is characteristic for covariance-stationary ARMA processes, but rather at a (much) slower hyperbolic rate. For $0 < d < 0.5$, y_t possesses long memory in the sense that the autocorrelations are not absolutely summable. Finally, for $0.5 < d < 1$, y_T is non-stationary, but the limiting value of the impulse response function is equal to 0, such that shocks do not have permanent effects.

Fractionally integrated models have been successfully implemented for exchange rates (Diebold, Husted and Rush, 1991; Cheung, 1993; and Baillie and Bollerslev, 1994), inflation rates (Baillie, Chung and Tieslau, 1996; and Hassler and Wolters, 1995), and unemployment (Diebold and Rudebusch, 1989; Tschernig and Zimmermann, 1992; Kostas and Veloce, 1996; and Crato and Rothman, 1998), see Baillie (1996) for a survey.

To capture nonlinear features in a time series y_t, one can choose from a wide variety of nonlinear models, see Franses and Van Dijk (2000) for a recent survey. A model which enjoys a fair amount of popularity, mainly due to its empirical tractability, is the smooth transition autoregressive [STAR] model, that is,

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1} + \cdots + \phi_{1,p}y_{t-p})(1 - G(s_t;\gamma,c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \cdots + \phi_{2,p}y_{t-p})G(s_t;\gamma,c) + \varepsilon_t, \quad (2)$$

where ε_t is a white noise process and the transition function $G(s_t;\gamma,c)$ usually is assumed to be the logistic function

$$G(s_t;\gamma,c) = (1 + \exp\{-\gamma(s_t - c)/\sigma_{s_t}\})^{-1} \qquad \text{with } \gamma > 0, \quad (3)$$

where s_t is the transition variable, σ_{s_t} is the standard deviation of s_t, γ is a slope parameter and c is a location parameter. The parameter restriction $\gamma > 0$ is an identifying restriction. The value of the logistic function (3), which is bounded between 0 and 1, depends on the transition variable s_t as follows. $G(s_t;\gamma,c) \to 0$ as $s_t \to -\infty$, $G(s_t;\gamma,c) = 0.5$ for $s_t = c$, and $G(s_t;\gamma,c) \to 1$ as $s_t \to +\infty$. When $\gamma \to \infty$, $G(s_t;\gamma,c)$ becomes a step function, such that the model effectively becomes a threshold autoregressive [TAR] model. For $\gamma = 0$, $G(s_t;\gamma,c) = 1/2$ for all s_t, in which case the model reduces to a linear AR model with parameters $\phi_j = (\phi_{1,j} + \phi_{j,2})/2$, $j = 0, 1, \ldots, p$. Finally, the exponent in (3) is normalized by dividing it by σ_{s_t} to make the parameter γ approximately scale-free, which is useful for finding initial estimates for the nonlinear optimization used to estimate the parameters in (2).

The transition variable s_t can be assumed to be a lagged endogenous variable ($s_t = y_{t-l}$ for certain integer $l > 0$), an exogenous variable ($s_t = z_t$), or a (possibly nonlinear) function of lagged endogenous variables, that is, $s_t = h(y_{t-1}, \ldots, y_{t-l}; \alpha)$ for some function $h(\cdot)$, which depends on the $(q \times 1)$ parameter vector α, and $e \geq 1$. In particular, if $h(y_{t-1}, \ldots, y_{t-l}; \alpha) = \sum_{j=1}^{l} \alpha_j y_{t-j}$, the model becomes an artificial neural network [ANN] with a single hidden unit, see also Medeiros and Veiga (2000). The transition variable can also be a linear time trend ($s_t = t$), which gives rise to an AR model with smoothly changing parameters.

Illustrative applications of the STAR model and the closely related TAR model to unemployment rates can be found in Montgomery *et al.* (1998), Koop and Potter (1999), Caner and Hansen (2000) and Skalin and Teräsvirta (2000), among others.

In this paper we combine the two representations in (1) and (2) into the following new time series model,

$$(1 - L)^d y_t = x_t, \quad (4)$$

$$x_t = \phi_1' w_t(1 - G(s_t;\gamma,c)) + \phi_2' w_t G(s_t;\gamma,c) + \varepsilon_t, \quad (5)$$

where $w_t = (1, \tilde{w}_t')'$, $\tilde{w}_t = (x_{t-1}, \ldots, x_{t-p})$, $\phi_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,p})'$, $i = 1, 2$, and $G(s_t; \gamma, c)$ is the logistic function given in (3). We assume that ε_t is a martingale difference sequence with respect to the history of the time series up to time $t-1$, which is denoted as $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_{1-(p-1)}, y_{1-p}\}$, that is, $\mathrm{E}[\varepsilon_t|\Omega_{t-1}] = 0$. For simplicity, we also assume that the conditional variance of ε_t is constant, $\mathrm{E}[\varepsilon_t^2|\Omega_{t-1}] = \sigma^2$, although this assumption can be relaxed if necessary.

This fractionally integrated smooth transition autoregressive [FISTAR] model restricts the long-run properties of the time series y_t to be constant, as these are determined by the fractional differencing parameter d. However, it does allow for different short-run dynamics in the two regimes corresponding with $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$, through ϕ_1 and ϕ_2. This makes the model potentially useful for capturing both nonlinear and long-memory features of the time series y_t.

The FISTAR model (4)-(5) can be expressed as an infinite order STAR model as

$$y_t = \left(\phi_{1,0} + \sum_{j=1}^{\infty} \pi_{1,j} y_{t-j}\right)(1 - G(s_t; \gamma, c)) + \left(\phi_{2,0} + \sum_{j=1}^{\infty} \pi_{2,j} y_{t-j}\right) G(s_t; \gamma, c) + \varepsilon_t, \quad (6)$$

where

$$\pi_i(L) = \phi_i(L)(1-L)^d, \qquad i = 1, 2, \quad (7)$$

with $\pi_i(L) = 1 - \pi_{i,1}L - \pi_{i,2}L^2 - \cdots$, and $\phi_i(L)$ defined similarly.

3 Specification of FISTAR models

Granger (1993) strongly recommends a 'specific-to-general' strategy for building nonlinear time series models. This implies starting with a simple or restricted model and proceeding to more complicated ones only if diagnostic tests indicate that the maintained model is inadequate. It is straightforward to extend the procedure proposed by Teräsvirta (1994) for STAR models to obtain such a 'specific-to-general' modeling cycle for FISTAR models. The resulting specification procedure can be summarized as follows.

1. Specify an appropriate linear ARFI model with autoregressive order p [ARFI(p)] for the time series under investigation;

2. Test the null hypothesis of linearity against the alternative of a FISTAR model. If linearity is rejected, select the appropriate transition variable s_t;

3. Estimate the parameters in the FISTAR model;

4. Evaluate the model using diagnostic tests.

Steps 2 to 4 in this specification procedure are discussed in turn below. This section concludes with some remarks on forecasting with FISTAR models, impulse response analysis and model selection.

3.1 Testing linearity

Consider again the FISTAR model in (4)-(5), but now with $G(s_t; \gamma, c)$ given by the general k-th order logistic function

$$G(s_t; \gamma, c) = \left(1 + \exp\left\{-\frac{\gamma}{\sigma_{s_t}^k}\prod_{i=1}^{k}(s_t - c_i)\right\}\right)^{-1} \qquad \text{with } \gamma > 0, c_1 \leq \ldots \leq c_k, \tag{8}$$

where $k \geq 1$. The additional parameter restrictions $c_1 \leq \ldots \leq c_k$ also are identifying restrictions. For odd k, the value of the logistic function (8) is bounded between 0 and 1, with $G(s_t; \gamma, c) \to 0$ as $s_t \to -\infty$, and $G(s_t; \gamma, c) \to 1$ as $s_t \to +\infty$. On the other hand, for even k, $G(s_t; \gamma, c)$ is bounded between a and 1 with $0 \leq a \leq 1/2$, and $G(s_t; \gamma, c) \to 1$ for $s_t \to \pm\infty$. For all k, there are k smooth transitions between the minimum and maximum values of $G(s_t; \gamma, c)$ as s_t increases from $-\infty$ to $+\infty$. Thus, the use of this transition function is advantageous as it allows to capture different forms of regime-switching behavior by varying the value of k, see Teräsvirta (1998). Here it is useful to derive tests of linearity against the alternative of a FISTAR model.

The null hypothesis of linearity can be expressed as equality of the autoregressive parameters in the two regimes of the FISTAR model in (4)-(5). Thus, $\mathsf{H}_0 : \phi_{1,j} = \phi_{2,j}$ for all $j = 0, 1, \ldots, p$. As in the STAR model, the testing problem is complicated by the presence of unidentified nuisance parameters under the null hypothesis, as the parameters in the transition function (γ and c) drop out of the model when the null hypothesis holds true. Note that besides equality of the AR parameters in the two regimes, the alternative null hypothesis $\mathsf{H}_0' : \gamma = 0$ also gives rise to a linear ARFI model.

To circumvent the identification problem we follow the approach of Luukkonen, Saikkonen and Teräsvirta (1988) and approximate $G(s_t; \gamma, c)$ by a first-order Taylor expansion around the null hypothesis $\mathsf{H}_0' : \gamma = 0$,

$$\begin{aligned} T_1(s_t; \gamma, c) &= G(s_t; 0, c) + \gamma \left.\frac{\partial G(s_t; \gamma, c)}{\partial \gamma}\right|_{\gamma=0} + R_1(s_t; \gamma, c) \\ &= \frac{1}{2} + \frac{\gamma}{4\sigma_{s_t}^k}\prod_{i=1}^{k}(s_t - c_i) + R_1(s_t; \gamma, c), \end{aligned} \tag{9}$$

where $R_1(s_t; \gamma, c)$ is a remainder term. Substituting $T_1(\cdot)$ for $G(\cdot)$ in (5) and rearranging terms yields the auxiliary regression

$$x_t = \beta_0' w_t + \beta_1' w_t s_t + \cdots + \beta_k' w_t s_t^k + e_t, \tag{10}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' w_t R_1(s_t; \gamma, c)$. Notice that under the null hypothesis, $R_1(s_t; \gamma, c) = 0$ and $e_t = \varepsilon_t$. Consequently, the remainder term of the Taylor approximation does not affect the properties of the residuals under the null hypothesis and hence the distribution theory for the test statistics. The relationships between the parameters $\beta_i =$

$(\beta_{i,0}, \beta_{i,1}, \dots, \beta_{i,p})'$, $i = 0, 1, \dots, k$, in the auxiliary regression (10) and the parameters in the original FISTAR model are such that the null hypothesis can be reformulated as $\mathsf{H}_0'' : \beta_i = 0$, $i = 1, \dots, k$.

The null hypothesis of linearity in (10) can conveniently be tested by means of a Lagrange Multiplier [LM] test. Assuming the errors to be normally distributed with variance σ^2, the conditional log-likelihood for observation t is given by

$$l_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 - \frac{e_t^2}{2\sigma^2}. \tag{11}$$

Because the information matrix is block diagonal, the error variance σ^2 can be assumed to be fixed. The remaining partial derivatives evaluated under the null hypothesis are given by

$$\left.\frac{\partial l_t}{\partial \beta_i}\right|_{\mathsf{H}_0''} = \frac{1}{\sigma^2}\hat{\varepsilon}_t w_t s_t^i, \qquad i = 0, 1, \dots, k, \tag{12}$$

$$\left.\frac{\partial l_t}{\partial d}\right|_{\mathsf{H}_0''} = \frac{1}{\sigma^2} e_t \left.\frac{\partial e_t}{\partial d}\right|_{\mathsf{H}_0''} = -\frac{1}{\sigma^2}\hat{\varepsilon}_t \sum_{j=1}^{t-1} \frac{\hat{\varepsilon}_{t-j}}{j}, \tag{13}$$

where $\hat{\varepsilon}_t$ are the residuals obtained from the ARFI model under the null hypothesis. The LM-type test statistic to test H_0'' is given by $\mathrm{LM} = \hat{l}_\beta' \mathrm{cov}(\hat{l}_\beta)^{-1}\hat{l}_\beta$, where $l_\beta = (l_{\beta,1}, \dots, l_{\beta,n})'$ and $l_{\beta,t} = (\partial l_t/\partial\beta_1, \dots, \partial l_t/\partial\beta_k)'$ and hats indicate that all elements of l_β should be evaluated under the null hypothesis. The expressions for the partial derivatives of the log-likelihood given above suggest that the statistic can be computed in a few steps as follows:

1. Estimate an ARFI model, obtain the residuals $\hat{\varepsilon}_t$, and compute the sum of squared residuals under the null hypothesis, $\mathrm{SSR}_0 = \sum_{t=1}^n \hat{\varepsilon}_t^2$.

2. Regress the residuals $\hat{\varepsilon}_t$ on w_t, $-\sum_{j=1}^{t-1}\frac{\hat{\varepsilon}_{t-j}}{j}$, and the auxiliary regressors $w_t s_t^i$, $i = 1, \dots, k$, and compute the sum of squared residuals under the alternative, SSR_1.

3. Compute the χ^2-version of the test statistic as $\mathrm{LM}_k = \frac{\mathrm{SSR}_0 - \mathrm{SSR}_1}{\mathrm{SSR}_0/n}$, or the F-version as

$$\mathrm{LM}_k = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/kp}{\mathrm{SSR}_1/(n - kp - (p+1))}, \tag{14}$$

   where n denotes the sample size.

Under the null hypothesis, the χ^2- and F-version of LM_k are χ^2 distributed with kp degrees of freedom, and F distributed with kp and $n - kp - (p+1)$ degrees of freedom, respectively. As usual, the F version of the test statistic is preferable to the χ^2 variant in small samples because its size and power properties are better.

Although the LM-type test in (14) can be computed for any value of k, it seems reasonable to consider only a number of small values, say, $k = 1$, 2 and 3. Finally, note that the only difference between the statistic derived above and the test of linearity against the alternative of a STAR model is the inclusion of $-\sum_{j=1}^{t-1} \frac{\hat{\varepsilon}_{t-j}}{j}$, which originates from the gradient of the likelihood with respect to the fractional differencing parameter d, in the auxiliary regression in step 2.

3.2 Estimation

To estimate the parameters in a FISTAR model we modify Beran's (1995) approximate maximum likelihood [AML] estimator for invertible and possible non-stationary ARFIMA models to allow for different autoregressive dynamics in the two regimes. The resulting estimator amounts to minimizing the sum of squared residuals

$$Q_n(\theta) = \sum_{t=1}^{n} e_t^2(\theta), \tag{15}$$

where $\theta \equiv (\phi_1', \phi_2', d, \gamma, c)$ and the residuals $e_t(\theta)$ are computed as

$$e_t(\theta) = y_t - (\phi_{1,0} + \sum_{j=1}^{\infty} \pi_{1,j} y_{t-j})(1 - G(s_t;\gamma,c)) - (\phi_{2,0} + \sum_{j=1}^{\infty} \pi_{2,j} y_{t-j}) G(s_t;\gamma,c) + \varepsilon_t,$$

where the $\pi_{1,j}$ and $\pi_{2,j}$ are the autoregressive coefficients in the infinite order STAR representation of the FISTAR model, which can be obtained from (7).

For the linear ARFI model, the AML estimator is asymptotically efficient if the errors ε_t are normally distributed. Under less restrictive regularity conditions, it is root-T consistent and asymptotically normal. We cannot claim efficiency of the AML estimator for the FISTAR model, but consistency and asymptotic normality follow from the fact that the FISTAR model satisfies conditions M1-M7 in Wooldridge (1994, pp. 2653–2655).

Issues that deserve particular attention, and which we discuss next, are concentrating the sum of squares function, the choice of the starting values for the optimization algorithm, and the estimates of the smoothness parameter γ.

Note that when the fractional differencing parameter d and the parameters in the transition function $G(s_t;\gamma,c)$ are known and fixed, the FISTAR model is linear in the remaining autoregressive parameters. Thus, conditional upon d, γ and c, estimates of ϕ_1 and ϕ_2 can be obtained by ordinary least squares [OLS] as

$$\hat{\phi}(d,\gamma,c)' = \left(\sum_{t=1}^{n} w_t(d,\gamma,c) w_t(d,\gamma,c)' \right)^{-1} \left(\sum_{t=1}^{n} w_t(d,\gamma,c) x_t \right), \tag{16}$$

where $w_t(d,\gamma,c) = (w_t'(1 - G(s_t;\gamma,c)), w_t' G(s_t;\gamma,c))'$, and the notation $\hat{\phi}(d,\gamma,c)$ is used to indicate that the estimate of ϕ is conditional upon d, γ and c. This implies that the sum of squares function can be concentrated with respect to ϕ_1 and ϕ_2, as

$$Q_n(d,\gamma,c) = \sum_{t=1}^{n}(x_t - \hat{\phi}(d,\gamma,c)'w_t(d,\gamma,c))^2.$$

This reduces the dimensionality of the estimation problem, as $Q_n(d,\gamma,c)$ needs to be minimized with respect to the three parameters d, γ, and c only. The above immediately suggests that a grid search over d, γ, and c is a convenient method to obtain starting values for the nonlinear optimization.

The remarks in Teräsvirta (1994, 1998) concerning the parameter estimates of STAR models apply to the FISTAR model as well. In particular, it might appear to be difficult to accurately estimate the smoothness of the transition between the different regimes, characterized by γ, when this parameter is large. This is due to the fact that for large values of γ, the logistic function $G(s_t;\gamma,c)$ in (3) is close to a step function. To obtain an accurate estimate of γ one then needs many observations in the immediate neighborhood of the threshold c, because even large changes in γ only have a small effect on the shape of the transition function. The estimate of γ may therefore be rather imprecise and often appear to be insignificant when judged by its t-statistic. This should, however, not be interpreted as evidence for only weak nonlinearity or parameter instability. The reason is that the t-statistic does not have its customary asymptotic t-distribution under the hypothesis $\gamma = 0$, due to the identification problems mentioned above.

3.3 Diagnostic tests

The tests of no residual autocorrelation, no remaining nonlinearity, and parameter constancy developed by Eitrheim and Teräsvirta (1996) for the standard STAR model can be modified in a straightforward manner to obtain similar diagnostic tests for the FISTAR model.

In particular, the null hypothesis of no autocorrelation in the residuals ε_t in (4)-(5) can be tested against the alternative of serial dependence up to order q, that is, under the alternative ε_t satisfies

$$\varepsilon_t = \alpha_1\varepsilon_{t-1} + \cdots + \alpha_q\varepsilon_{t-q} + e_t, \tag{17}$$

with $e_t \sim$ i.i.d.$(0,\sigma^2)$. The null hypothesis is given by $\mathsf{H}_0 : \alpha_1 = \ldots = \alpha_q = 0$ which, following Eitrheim and Teräsvirta (1996), is tested by an LM test.

The null hypothesis of no remaining nonlinearity can be tested against the additive 3-regime FISTAR model, which is obtained by combining (4) with

$$x_t = \phi_1'w_t + (\phi_2 - \phi_1)'w_tG_1(s_t) + (\phi_3 - \phi_2)'w_tG_2(s_t) + e_t, \tag{18}$$

where $G_i(s_t) \equiv G_i(s_t;\gamma_i,c_i)$, $i = 1,2$ are logistic functions as in (8). The null hypothesis of no remaining nonlinearity is given by either $\mathsf{H}_0 : \phi_2 = \phi_3$ or $\mathsf{H}_0' : \gamma_2 = 0$. Obviously then, this testing problem suffers from the presence of unidentified nuisance parameters under the null hypothesis, as in the case of testing linearity against the FISTAR alternative discussed in Section 3.1. The proposed solution is the same as well, that is, avoid the identification problem by replacing the second transition function $G_2(s_t)$ by a suitable Taylor-approximation around the null hypothesis, such that the null hypothesis can be tested by a standard LM variable addition test.

In case the additional nonlinearity is thought to be determined by a variable different from s_t, a more natural model to consider under the alternative hypothesis is the 4-regime FISTAR model, given by (4) with

$$x_t = (\phi_1' w_t(1 - G_1(s_{1t})) + \phi_2' w_t G_1(s_{1t}))(1 - G_2(s_{2t})) \\ + ((1 - G_1(s_{1t}))\phi_3' w_t + G_1(s_{1t})\phi_4' w_t) G_2(s_{2t}) + e_t, \quad (19)$$

where $s_{1t} \equiv s_t$, see also Franses and van Dijk (1999). A test of no remaining nonlinearity again is obtained by approximating $G_2(s_t)$ by a suitable Taylor-expansion around the null hypothesis $\mathsf{H}_0' : \gamma_2 = 0$.

A test of parameter constancy in the FISTAR model against the alternative of smoothly changing parameters can also be based on (19) by taking $s_{2t} = t$. This can perhaps be easiest seen from the alternative representation

$$x_t = \phi_{1t}' w_t(1 - G_1(s_{1t}; \gamma_1, c_1)) + \phi_{2t}' w_t G_1(s_{1t}; \gamma_1, c_1) + e_t, \quad (20)$$

where

$$\phi_{1t} = (1 - G_2(t))\phi_1 + G_2(t)\phi_3, \quad (21)$$

$$\phi_{2t} = (1 - G_2(t))\phi_3 + G_2(t)\phi_4. \quad (22)$$

For all three diagnostic tests, the only difference with the corresponding tests in the STAR model is that one needs to include the gradient of e_t with respect to the fractional differencing parameter d, evaluated under the null hypothesis, in the auxiliary regressions which are used to compute the test statistics. Under the null hypothesis $e_t = \varepsilon_t$, such that

$$\left.\frac{\partial e_t}{\partial d}\right|_{\mathsf{H}_0} = \frac{\partial \varepsilon_t}{\partial d} = -\sum_{j=1}^{t-1} \frac{\varepsilon_{t-j}}{j}.$$

Note that this is exactly identical to the difference between the linearity tests against STAR and FISTAR alternatives, as discussed in Section 3.1.

3.4 Further issues

Forecasts from the FISTAR model can be obtained from the infinite order STAR representation given in (6). Of course, in practice, only a finite past of the time series y_t

is available, so that the STAR representation has to be truncated at some point, which introduces an additional prediction error. Closed-form expressions for multiple-step ahead forecasts from STAR models are not available, such that one has to resort to Monte Carlo or bootstrap methods to obtain these, see Granger and Teräsvirta (1993, Section 8.1).

Criteria such as the Akaike Information Criteria [AIC] and the Schwarz Information Criteria [BIC] can be used to choose between competing models, such as ARFI or STAR and FISTAR models.

A useful way of examining the dynamic behaviour of an estimated FISTAR model is to consider the effects of the shocks ε_t on the future patterns of the time series y_t by means of impulse response analysis. It is well-known that in nonlinear models the impact of a shock depends on the history of the process, on the sign and the size of the shock, and on the shocks that occur between the time the impulse is given, say, t, and the time the response is measured, say, $t+h$ for some $h>0$. The generalized impulse response function [GI], introduced by Koop, Pesaran and Potter (1996), offers a convenient way to deal with these issues. The GI for a specific shock $\varepsilon_t=\delta$ and history $\omega_{t-1}=\{y_{t-1},y_{t-2},\dots\}$ is defined as

$$\mathrm{GI}_y(h,\delta,\omega_{t-1})=\mathrm{E}[y_{t+h}|\varepsilon_t=\delta,\omega_{t-1}]-\mathrm{E}[y_{t+h}|\omega_{t-1}], \tag{23}$$

for $h=0,1,2,\dots$. In the GI, the expectation of y_{t+h} given that a shock δ occurs at time t is conditioned only on the history and this shock. Put differently, the problem of dealing with shocks occurring in intermediate time periods is handled by averaging them out. Given this choice, the natural benchmark profile for the impulse response is the expectation of y_{t+h} conditional only on the history of the process ω_{t-1}. Thus, in the benchmark profile the current shock is averaged out as well.

The GI is a function of δ and ω_{t-1} that are realizations of the random variables ε_t and Ω_{t-1}. Hence, $\mathrm{GI}_y(h,\delta,\omega_{t-1})$ itself is a realization of a random variable, defined as

$$\mathrm{GI}_y(h,\varepsilon_t,\Omega_{t-1})=\mathrm{E}[y_{t+h}|\varepsilon_t,\Omega_{t-1}]-\mathrm{E}[y_{t+h}|\Omega_{t-1}]. \tag{24}$$

In addition, one might consider the GI to be random conditional on particular subsets $\mathcal{S}$ and $\mathcal{H}$ of shocks and histories respectively, that is, $\mathrm{GI}_y(h,\mathcal{S},\mathcal{H})$. For example, one might condition on all histories in a particular regime and consider only negative shocks.

If the time series y_t is not persistent, the impulse responses eventually converge to zero for all possible shocks and all possible histories. Hence, the distribution of $\mathrm{GI}_y(h,\varepsilon_t,\Omega_{t-1})$ collapses to a spike at 0 as $h\to\infty$. By contrast, for persistent time series the dispersion of the distribution of $\mathrm{GI}_y(h,\varepsilon_t,\Omega_{t-1})$ is positive for all h. Koop *et al.* (1996) suggest that the dispersion of the distribution of $\mathrm{GI}_y(h,\varepsilon_t,\Omega_{t-1})$ at finite horizons can be interpreted as a measure of the persistence of shocks.

The absorption rate developed by van Dijk, Franses and Boswijk (2000) is an alternative measure of the speed at which the effect of a shock dies out. Assume that the eventual response to a shock $\mathrm{GI}_y(\infty, \delta, \omega_{t-1})$ is finite and define the indicator function $I_y(\pi, h, \delta, \omega_{t-1})$ as

$$I_y(\pi, h, \delta, \omega_{t-1}) \equiv I[|\mathrm{GI}_y(h, \delta, \omega_{t-1}) - \mathrm{GI}_y(\infty, \delta, \omega_{t-1})| \leq \pi|\delta - \mathrm{GI}_y(\infty, \delta, \omega_{t-1})|]$$

for certain π such that $0 \leq \pi \leq 1$, where $I[A] = 1$ if the event A occurs and 0 otherwise. In words, the indicator function $I_y(\pi, h, \delta, \omega_{t-1})$ is equal to 1 if the absolute difference between the GI at horizon h and the eventual response to the shock δ is less than or equal to a fraction π of the absolute difference between the shock δ, which is equal to the initial impact of the shock or the GI at horizon 0, and the eventual response. Put differently, $I_y(\pi, h, \delta, \omega_{t-1}) = 1$ if at least a fraction $1 - \pi$ of the initial effect of δ has been absorbed after h periods.

The 'π-life' or 'π-absorption time' of the shock δ at time t, denoted as $\mathrm{N}_y(\pi, \delta, \omega_{t-1})$, can now be defined as the minimum horizon beyond which the difference between the impulse responses at all larger horizons and the eventual response is less than or equal to a fraction π of the difference between the initial impact and the eventual response. That is, $\mathrm{N}_y(\pi, \delta, \omega_{t-1}) = m$ if $I_y(\pi, h, \delta, \omega_{t-1}) = 1$ for all $h \geq m$ and $I_y(\pi, m-1, \delta, \omega_{t-1}) = 0$.

Just like the shock- and history-specific GI in (23) can be regarded as a realization of the random variable $\mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1})$ in (24), the π-absorption time $\mathrm{N}_y(\pi, \delta, \omega_{t-1})$ can be regarded as a realization of the random variable $\mathrm{N}_y(\pi, \varepsilon_t, \Omega_{t-1})$. Conditional versions $\mathrm{N}_y(\pi, \mathcal{S}, \mathcal{H})$ for particular subsets $\mathcal{S}$ and $\mathcal{H}$ of shocks and histories, respectively, can be defined in a straightforward manner, see van Dijk, Franses and Boswijk (2000) for details.

The GI can also be used to assess the significance of nonlinear or asymmetric effects over time. Potter (1994) defines a measure of asymmetric response to a particular shock $\varepsilon_t = \delta$ given a particular history ω_{t-1} as the sum of the GI for this particular shock and the GI for the shock of the same magnitude but with opposite sign, that is,

$$\mathrm{ASY}_y(h, \delta, \omega_{t-1}) = \mathrm{GI}_y(h, \delta, \omega_{t-1}) + \mathrm{GI}_y(h, -\delta, \omega_{t-1}). \tag{25}$$

Alternatively, one could consider the distribution of the random asymmetry measure

$$\mathrm{ASY}_y(h, \varepsilon_t, \Omega_{t-1}) = \mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1}) + \mathrm{GI}_y(h, -\varepsilon_t, \Omega_{t-1}). \tag{26}$$

$\mathrm{ASY}_y(h, \varepsilon_t, \Omega_{t-1})$ should be equal to zero almost surely if positive and negative shocks have exactly the same effect (with opposite sign). More generally, we say that shocks have a symmetric effect (on average) when $\mathrm{ASY}_y(h, \mathcal{P}, \Omega_{t-1})$, where $\mathcal{P} = \{\varepsilon_t | \varepsilon_t > 0\}$ is the set of all positive shocks, has a symmetric distribution with mean equal to zero. The dispersion of this distribution might be interpreted as a measure of the asymmetry in the effects of positive and negative shocks.

Possible asymmetry in the absorption of positive and negative shocks can be examined in a similar way. For a specific shock $\varepsilon_t = \delta$ and history ω_{t-1}, a measure of asymmetric absorption can be defined as the difference in π-absorption times of δ and $-\delta$, that is,

$$\mathrm{ASYN}_y(\pi, \delta, \omega_{t-1}) = \mathrm{N}_y(\pi, \delta, \omega_{t-1}) - \mathrm{N}_y(\pi, -\delta, \omega_{t-1}). \tag{27}$$

If δ has symmetric absorption speed at ω_{t-1}, $\mathrm{ASYN}_y(\pi, \delta, \omega_{t-1}) = 0$ for all values of π. As before, the asymmetry measure in (27) can be regarded as a realization of the random variable

$$\mathrm{ASYN}_y(\pi, \varepsilon_t, \Omega_{t-1}) = \mathrm{N}_y(\pi, \varepsilon_t, \Omega_{t-1}) - \mathrm{N}_y(\pi, -\varepsilon_t, \Omega_{t-1}). \tag{28}$$

If positive and negative shocks have symmetric effects, in the sense that they are absorbed at the same speed on average, $\mathrm{ASYN}_Y(\pi, \mathcal{P}, \Omega_{t-1})$ should have a distribution with mean equal to zero.

Note that symmetry in $\mathrm{GI}_y(h, \delta, \omega_{t-1})$, that is, $\mathrm{ASY}_y(h, \delta, \omega_{t-1}) = 0$ for all $h \geq 0$ in (25), implies symmetry in the absorption speed, that is, $\mathrm{ASYN}_y(\pi, \delta, \omega_{t-1}) = 0$ for all $\pi > 0$. Interestingly, the reverse does not hold, that is, a shock can have symmetric absorption speed but an asymmetric impulse response. Also, $\mathrm{ASYN}_y(\pi, \delta, \omega_{t-1}) \neq 0$ for certain $\pi > 0$ implies that $\mathrm{ASY}_y(h, \delta, \omega_{t-1}) \neq 0$ for certain $h \geq 0$, whereas the reverse does not hold.

4 Monthly US unemployment rate

In this section, we apply the FISTAR model to characterize the persistence and nonlinearity properties in the monthly US unemployment rate.

4.1 Data

The series we consider represents the seasonally adjusted unemployment rate, covering the period July 1968 until December 1999 (378 observations). The series is constructed by taking the ratio of the unemployment level and civilian labor force, which are obtained from the *Bureau of Labor Statistics.*

- insert Figure 1 about here -

From Figure 1 which plots the series, it is clearly seen that the two dominant features of the unemployment rate are asymmetric behaviour over the business cycle and high persistence. The behaviour of the unemployment rate over the business cycle can be characterized as steep increases during recessions, followed by slow(er) declines during expansions. Several theories, such as asymmetric labor adjustment costs of enterprises, insider-outsider relationships, and asymmetries in job destruction and reconstruction have been developed to explain this asymmetry in the dynamic behaviour of the unemployment rate. Both TAR and STAR models have been applied to unemployment rates to describe this type of nonlinearity, see Hansen (1997), Montgomery *et al.* (1998), Rothman (1998), Koop and Potter (1999), Caner and Hansen (2000), and Skalin and Teräsvirta (2000), among others. In general, it is found that these models improve upon linear models both in describing the in-sample properties of the unemployment rate and in out-of-sample forecasting.

From Figure 1 it is also clear that the unemployment rate is persistent. In fact, the persistence of the unemployment rate also has received much attention. The two competing viewpoints are the 'natural rate' hypothesis and the hysteresis hypothesis of Blanchard and Summers (1987). Under the natural rate hypothesis, the unemployment rate is mean-reverting, whereas it is non-stationary under the hysteresis hypothesis. Thus, the two hypotheses imply that different transformations (levels and first differences, respectively) of the unemployment rate can be appropriate. Using a fractionally integrated model we avoid having to take position on the persistence properties of the unemployment rate *a priori*, but instead we let the data decide which characterization is most appropriate. It should be noted that (both theoretical and empirical) models which allow the natural rate to change over time have been put forward as another alternative for the hysteresis hypothesis, see Coakley *et al.* (2000) for a recent review, but this is not a direction we pursue here.

4.2 An empirical FISTAR model

Following the modelling cycle as outlined in Section 3, we start by specifying a linear ARFI model. We allow for a maximum autoregressive order of $p_{\max} = 18$, such that the effective estimation sample runs from January 1970 until December 1999 (360 observations). Both AIC and BIC indicate that an ARFI model with $p = 4$ is adequate. The third column of Table 1 contains summary statistics and diagnostic tests for this model. In particular, the estimate of d is equal to 0.84, suggesting that the unemployment rate is non-stationary but mean-reverting, confirming the findings of Diebold and Rudebusch (1989), Tschernig and Zimmermann (1992), Koustas and Veloce (1996), and Crato and Rothman (1998). This linear long-memory model appears adequate in that the errors seem serially uncorrelated, whereas the excess kurtosis and heteroskedasticity appear to be caused entirely by large positive residuals in January 1975, April 1980 and February 1986.

- insert Table 1 about here -

The next stage is to test linearity against FISTAR nonlinearity using the LM-type statistics developed in Section 3.1. As we are concerned with the behaviour of the unemployment rate over the business cycle, the transition variable in the FISTAR model should reflect the property that recession and expansion regimes are sustained periods of increase and decline in the unemployment rate, respectively. This makes the monthly change in the unemployment rate unsuitable as an indicator of the business cycle regime as it is too noisy. Following Skalin and Teräsvirta (2000), we therefore consider the twelve-month difference as transition variable, that is, $s_t = \Delta_{12} y_{t-l} \equiv y_{t-l} - y_{t-l-12}$, $l = 1, \ldots, l_{\max}$. We set the maximum value of the delay parameter $l_{\max}$ equal to 6.

The upper panel of Table 2 contains p-values of the F-version of the LM$_k$ statistic, $k = 1, 2, 3$, as given in (14), with $\Delta_{12} y_{t-l}$, $l = 1, \ldots, 6$, as transition variable. LM-type tests against the alternative of smoothly changing parameters, where $s_t = t$, are given as well. Linearity is rejected quite convincingly against the alternative of a FISTAR model with $s_t = \Delta_{12} y_{t-l}$ for all values of l considered, where the minimum p-values are attained in case $l = 1$. Constancy of the autoregressive parameters (based on the tests with t as transition variable) is also rejected by the LM$_1$ test, but the evidence for structural change is much less compelling that the evidence for nonlinearity.

- insert Table 2 about here -

Based on these test results, we proceed by estimating a FISTAR model with $s_t = \Delta_{12} y_{t-1}$ and an autoregressive order equal to 4 in both regimes. The AML estimates of the parameters in this model are:

$$(1-L)^{\hat{d}} y_t = x_t, \qquad \hat{d} = \underset{(0.15)}{0.43}, \tag{29}$$

$$\begin{aligned} x_t = {} & (\underset{(0.016)}{-0.050} + \underset{(0.17)}{0.30}\, x_{t-1} + \underset{(0.079)}{0.22}\, x_{t-2} + \underset{(0.078)}{0.15}\, x_{t-3} + \underset{(0.081)}{0.14}\, x_{t-4}) \times (1 - G_1(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})) \\ & + (\underset{(0.15)}{0.074} + \underset{(0.17)}{0.80}\, x_{t-1} + \underset{(0.11)}{0.18}\, x_{t-2} - \underset{(0.11)}{0.019}\, x_{t-3} - \underset{(0.077)}{0.11}\, x_{t-4}) \times G_1(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c}), \end{aligned} \tag{30}$$

$$G_1(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c}) = (1 + \exp\{-\underset{(14.4)}{14.5}\, (\Delta_{12} y_{t-1} - \underset{(0.077)}{0.15})/\sigma_{\Delta_{12} y_{t-1}}\})^{-1}, \tag{31}$$

where standard errors are given in parentheses below the parameter estimates.

The second column of Table 1 and the lower panel of Table 2 contain diagnostic tests for the estimated model. It is seen that the FISTAR does not improve the distributional properties of the residuals very much, in the sense that it cannot account for the positive skewness, excess kurtosis and ARCH effects that were found in the linear ARFI model, but again these deviations from normality and homoskedasticity appear to be due to only a few observations. Results of the diagnostic tests in the lower block of Table 2 suggest that the FISTAR model is adequate as there is no evidence for time-variation in the parameters or remaining nonlinearity. Note that AIC prefers the FISTAR model whereas BIC prefers the parsimonious ARFI model.

Figure 2 shows plots of the transition function in the estimated FISTAR model, both over time and against the transition variable $\Delta_{12} y_{t-1}$. The estimates of the parameters γ and c are such that the change of the logistic function $G(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})$ from 0 to 1 takes place for values of $\Delta_{12}y_{t-1}$ between -0.2 and 0.5. The bottom panel of Figure 2 also contains the (rescaled) unemployment rate, where circles indicate individual peaks and troughs as dated by the NBER. These peaks and troughs differ from the reference business cycle turning points, as the unemployment rate is, on average, leading at peaks and lagging at troughs. The two regimes in the FISTAR model correspond reasonably close with the contractions and expansions as identified by these turning points simply because $G(\Delta_{12}y_{t-1};\hat{\gamma},\hat{c})$ is a monotonic transformation of the transition variable $\Delta_{12}y_{t-1}$. As the transition variable is the change in the unemployment rate over the previous year, the switches between the regimes do not coincide exactly with the peaks and troughs of the unemployment rate but usually take place a few months later.

- insert Figure 2 about here -

4.3 Comparison with other models

The estimate of the fractional differencing parameter d (= 0.43) in (29) is significantly different from both 0 and 1. To further examine whether the FISTAR model improves upon STAR models for levels and first differences of the unemployment rate, we estimate FISTAR models with $d = 0$ and $d = 1$ imposed. The fourth and fifth columns of Table 1 contain estimates of the parameters in the transition functions and summary statistics for these models. The FISTAR model is preferred over the STAR model in first differences according to both AIC and BIC. AIC is indifferent between the FISTAR model and the STAR model for levels, whereas BIC prefers the latter model. The diagnostic tests for these models reveal that the STAR model for first differences is not adequate, as the null hypothesis of no remaining nonlinearity is rejected. Parameter constancy is rejected in the STAR model for levels. Detailed results of these misspecification tests are available upon request. Finally, we also estimate a STAR model for first differences, augmented with a lagged level term, that is,

$$\Delta y_t = \left(\phi_{1,0} + \rho_1 y_{t-1} + \sum_{j=1}^{p} \phi_{1,j} y_{t-j}\right)(1 - G(s_t;\gamma,c)) + \left(\phi_{2,0} + \rho_2 y_{t-1} + \sum_{j=1}^{p} \phi_{2,j} y_{t-j}\right) G(s_t;\gamma,c) + \varepsilon_t. \quad (32)$$

This model has been put forward by Skalin and Teräsvirta (2000) as a stationary nonlinear model which can suggest nonstationarity when analyzed with linear tools and which can generate the typical asymmetry of unemployment rates. The final column of Table 1 contains relevant summary statistics for this model. Both AIC and BIC are smaller than for the FISTAR model. Unreported diagnostic tests of no remaining nonlinearity and parameter constancy are satisfactory as well, as neither of these hypotheses can be rejected at conventional significance levels.

To understand the properties of the estimated FISTAR model and to compare it with the other models, it is useful to consider the skeleton of the model, that is, the deterministic part of (29)-(31). The skeleton is such that the FISTAR model contains a limit cycle of 292 months in which the unemployment rate fluctuates between 4% and 8%. This limit cycle is shown in Figure 3. The range of the cycle corresponds quite closely with the unemployment rates observed during the 1990s, suggesting that the high unemployment rate during the 1980s only was a temporary phenomenon. The periodicity of the cycle is much longer than that observed in the empirical time series, which is about 10 years. The limit cycle also contains marked asymmetry, as the parts of the cycle during which the series increases and decreases are much different in length (86 and 206 months, respectively). Figure 3 also shows deterministic extrapolation with the estimated STAR model for first differences which includes a lagged level term. It turns out that this model also contains a limit cycle, with a comparable range (5-8%), but a much higher periodicity (136 months) and somewhat more pronounced asymmetry (the increase and decrease in the unemployment rate last for 35 and 91 months, respectively).

- insert Figure 3 about here -

To gain further insight in the dynamic properties of the estimated FISTAR model, we assess the propagation of shocks by computing generalized impulse response functions. We compute history- and shock-specific GIs as defined in (23) for all observations in the period from July 1974 until December 1999 and values of the normalized initial shock equal to $\delta/\hat{\sigma}_\varepsilon = \pm 3, \pm 2.8, \ldots, \pm 0.4, \pm 0.2$, where $\hat{\sigma}_\varepsilon$ denotes the estimated standard deviation of the residuals from the FISTAR model. For each combination of history and initial shock, we compute $\text{GI}_y(h, \delta, \omega_{t-1})$ for horizons $h = 0, 1, \ldots, N$ with $N = 120$. To generate future sample paths of y_t from the FISTAR model, we use the infinite order STAR representation truncated at 120 lags. The conditional expectations in (23) are estimated as the means over 1000 realizations of y_{t+h}, with and without using the selected initial shock to obtain y_t and using randomly sampled residuals of the estimated FISTAR model elsewhere. We follow the same procedure to compute GIs for the STAR model in first differences with a lagged level term. All GIs are normalized such that they equal $\delta/\hat{\sigma}_\varepsilon$ at $h = 0$.

The GI's for specific histories and shocks are used to estimate the density of $\text{GI}_y(h, \mathcal{S}, \mathcal{H})$, where $\mathcal{S}$ and $\mathcal{H}$ denote sets of selected shocks and histories, respectively. The set of shocks $\mathcal{S}$ is the set of all negative or positive shocks, whereas the set $\mathcal{H}$ consists of the histories for which the value of the transition function $G(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})$ in (31) is greater ('recession') and less ('expansion') than 0.5. The densities are obtained with a standard Nadaraya-Watson kernel estimator, using $\phi(\delta/\hat{\sigma}_\varepsilon)$ as weight for $\text{GI}_y(h, \delta, \omega_{t-1})$, where $\phi(z)$ denotes the standard normal probability distribution. The reason for using this weighting scheme is that the standardized shocks $\delta/\hat{\sigma}_\varepsilon$ then effectively are sampled from a discretized normal distribution and the resulting distribution of $\text{GI}_y(h, \varepsilon_t, \Omega_{t-1})$ should resemble a normal distribution if the effect of shocks is symmetric and proportional to their magnitude (as is the case in linear models). Finally, highest density regions are then estimated using the density quantile method outlined in Hyndman (1995, 1996).

- insert Figure 4 about here -

Figure 4 shows HDRs for distributions of $\text{GI}_y(h, \mathcal{S}, \mathcal{B})$ for $h = 0, 6, 12, \ldots, 120$ for the FISTAR model. It appears that several interesting asymmetries in the impulse responses exist. First, shocks occurring during recessions tend to be magnified during the first 12 months, after which their effect declines gradually towards zero. Shocks occurring during expansions reach their maximum effect only after 18 months. Second, the effect of positive shocks during expansions is much larger than the effect of negative shocks during the first 3 years after impact. On the other hand, there does not appear so much asymmetry between the impulse responses for positive and negative shocks occurring during recessions. The latter observations are confirmed by the measure of asymmetric impulse response $\text{ASY}_y(h, \delta, \omega_{t-1})$ defined in (25). The upper panel of Table 3 contains means of the random asymmetry measures $\text{ASY}_y(h, \mathcal{S}, \mathcal{H})$ for $h = 12, 24, 36, 48$ and 60, for different sets of shocks $\mathcal{S}$ defined as A(ll) $= \{\varepsilon_t | \varepsilon_t > 0\}$, S(mall) $= \{\varepsilon_t | 1 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 0\}$, M(edium) $= \{\varepsilon_t | 2 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 1\}$ and L(arge) $= \{\varepsilon_t | 3 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 2\}$. The set $\mathcal{H}$ consists of all histories ('unconditional') or only of those histories for which the transition function $G(\Delta_{12} y_{t-1}; \gamma, c)$ in (31) is larger ('recession') and smaller ('expansion') than 0.5. To judge whether the mean of $\text{ASY}_y(h, \mathcal{S}, \mathcal{H})$ is significantly different from zero, we use $\sigma_{\text{ASY}_y(h,\mathcal{S},\mathcal{H})}/\sqrt{n_\mathcal{S}}$, where $\sigma_{\text{ASY}_y(h,\mathcal{S},\mathcal{H})}$ is the standard deviation of $\text{ASY}_y(h, \mathcal{S}, \mathcal{H})$ and $n_\mathcal{S}$ is the number of shocks δ in the set $\mathcal{S}$ for which $\text{ASY}_y(h, \delta, \omega_{t-1})$ is computed, as standard error for the mean. The reason for dividing by $n_\mathcal{S}$ is that different realizations $\text{ASY}_y(h, \delta, \omega_{t-1})$ are not independent across histories ω_{t-1} but are independent across shocks δ. It is seen that the absolute value of the mean asymmetry measure is invariably larger (in absolute value) for shocks occurring during expansions. Also noteworthy is that the mean asymmetry is actually negative for small shocks during expansions at horizons up to 4 years and beyond 6 years, suggesting that negative shocks have a larger impact than positive shocks. For medium-sized and large shocks, mean asymmetry is positive at these horizons. Symmetry of the impulse responses to shocks occurring during recessions is rejected only for large shocks at horizons between 2 and 3 years and between 5 and 6 years. Asymmetry for shocks occurring during expansions is detected for all sizes of shocks at horizons up to 3 years, while asymmetry is found for medium-sized and large shocks at longer horizons as well.

- insert Table 3 about here -

Figure 5 contains HDRs for distributions of $\text{GI}_y(h, \mathcal{S}, \mathcal{H})$ for $h = 0, 6, 12, \ldots, 120$ for the STAR model for first differences with lagged level term. Comparing these with the HDRs for the FISTAR model in Figure 4, it appears that roughly the same type of asymmetries are captured by the two models. In the STAR model it also is the case that shocks occurring during recessions reach their maximum impact sooner than shocks occurring during expansions, while the asymmetry between positive and negative shocks appears to be somewhat more pronounced for shocks occurring during expansions.

- insert Figures 5 and 6 about here -

A notable difference between the GIs of the two models, which is not immediately apparent from the HDRs is that the impulse responses for the STAR model decay in an strongly oscillatory fashion, such that, for example, the largest positive shocks actually generate the most negative responses at horizons between 48 and 96 months, and *vice versa*. The impulse responses in the FISTAR model also oscillate, but to a much lesser extent. The response to positive shocks is negative at horizons between 54 and 75 months, while this effect also is much less pronounced than in the STAR model. This is seen clearly from Figure 6, which shows the mean of $\text{GI}_y(h, \mathcal{S}, \mathcal{H})$ for $\mathcal{S} = \{\varepsilon_t | \varepsilon_t = k\sigma\}$, $k = -3, -2, \ldots, 3$. Also note that the mean asymmetry measures for the STAR model, reported in the lower panel of Table 3, show a similar cyclical pattern, which again is much more pronounced than the pattern in the mean asymmetry measures for the FISTAR model. Another difference between the impulse responses in the two models, which can be seen from Figures 4 and 5 is that the GIs in the STAR model decay much faster than the GIs in the FISTAR model.

- insert Tables 4 and 5 about here -

Finally, we report absorption times of shocks in the FISTAR and STAR models in Table 4. It appears that in both models large shocks are absorbed somewhat faster than small shocks, and shocks in recessions are absorbed slower than shocks in expansions. The half-lives of shocks ($\pi = 0.50$) are similar in the two models, but the differences in absorption times can be seen to become larger for smaller values of π.

Asymmetry measures of the absorption times are given in Table 5. In both models, medium-sized and large positive shocks in recessions are absorbed faster than negative shocks of equal size, while small positive shocks are absorbed slower. The reverse holds for expansions. Note that the asymmetry is more pronounced in FISTAR model than in the STAR model. In sum, it appears that the FISTAR model does highlight interesting features of the data, which are not captured by the STAR model.

5 Concluding remarks

In this paper we proposed a new time series model which can capture long memory and nonlinearity at the same time. Upon fitting it to three decades of monthly US unemployment, we found that a rather parsimonious version of the model fits the data well. When we compared the model with various possible competitive models, we found that a linear fractionally integrated model could certainly be improved by including nonlinear features. Indeed, once we added these, there remained no evidence of nonlinearity and time-varying parameters. However, the introduction of long-memory into a STAR model did not give a much better fit. The key distinction between these two models lies in the fact that our combined model has other long-run properties. We highlighted these using various impulse response functions.

There are various directions for further research. The first one originates from the fact that we considered seasonally adjusted data, while perhaps it would have been better to consider the original unemployment rate series. This would require to augment the FISTAR model with an explicit description of the seasonal patterns of the series. Second, our model assumes that the long-memory and nonlinear characteristics of the time series are constant over time, whereas recent research indicates that structural change might be an important feature of unemployment rates, see Bianchi and Zoega (1998), Papell *et al.* (2000) and Coakley *et al.* (2000). Extending the FISTAR model to allow for structural changes, possibly along the lines of the time-varying STAR model put forward by Lundbergh *et al.* (2000), might be worthwhile. Both these extensions would lead to a time series model that captures three features jointly. A third further research topic amounts to comparing ARFI, STAR and FISTAR models in terms of out-of-sample forecasting properties. In particular, one would hope that long-memory models would forecast better for long horizons and that nonlinear models would outperform linear models if nonlinearity is an important feature of the data. However, Ray (1993) and Crato and Ray (1996) demonstrate that using AR(MA) models to predict long-memory time series does not result in a large loss of forecasting accuracy. Similarly, Clements and Krolzig (1998) show that AR models have a competitive forecasting performance for nonlinear (Markov Switching and threshold autoregressive) time series. Finally, extensions to allow for more than two regimes and to multivariate series also seem interesting areas for further research.

References

Andersson, M.K., (2000), Do long-memory models have long memory?, *International Journal of Forecasting* **16**, 121–124.

Andersson M.K., B. Eklund and J. Lyhagen (1999), A simple linear time series model with misleading nonlinear properties, *Economics Letters* **65**, 281–284.

Baillie, R.T. (1996), Long memory processes and fractional integration in econometrics, *Journal of Econometrics* **73**, 5–59.

Baillie, R.T. and T. Bollerslev (1994), The long memory of the forward premium, *Journal of International Money and Finance* **13**, 565–571.

Baillie R.T., C.F. Chung and M.A. Tieslau (1996), Analysing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* **11**, 23–40.

Baum C.F., J.T. Barkoulas and M. Caglayan (1999), Persistence in international inflation rates, *Southern Economic Journal* **65**, 900–913.

Beran, J. (1995), Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society* **B 57**, 659–672.

Bianchi, M. and G. Zoega (1998), Unemployment persistence: does the size of the shock matter?, *Journal of Applied Econometrics* **13**, 283–304.

Blanchard, O.J. and L.H. Summers (1987), Hysteresis in unemployment, *European Economic Review* **31**, 288–295.

Bos, C., P.H. Franses and M. Ooms (1999), Re-analyzing inflation rates: evidence of long memory and level shifts, *Empirical Economics* **24**, 427–449.

Caner, M. and B.E. Hansen (2000), Threshold autoregression with a unit root, unpublished manuscript, University of Wisconsin, Madison.

Cheung, Y.-W. (1993), Long memory in foreign exchange rates, *Journal of Business & Economic Statistics* **11**, 93–101.

Clements, M.P. and H.-M. Krolzig (1998), A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *Econometrics Journal* **1**, C47–C75.

Coakley, J., A.-M. Fuertes and G. Zoega (2000), Evaluating the persistence and structuralist theories of unemployment, CEPR Discussion Paper No. 2438, Center for Economic Policy Research.

Crato N. and B.K. Ray (1996), Model selection and forecasting for long-range dependent processes, *Journal of Forecasting* **15**, 107–125.

Crato, N. and P. Rothman (1996), Measuring hysteresis in unemployment rates with long memory models, unpublished manuscript, East Carolina University.

Davidson, J. (2000), When is a time series I(0)? Evaluating the memory properties of nonlinear dynamic models, unpublished manuscript, Cardiff University.

Diebold, F.X., S. Husted and M. Rush (1991), Real exchange rates under the gold standard, *Journal of Political Economy* **99**, 1252–1271.

Diebold, F.X. and A. Inoue (1999), Long memory and structural change, unpublished manuscript, University of Pennsylvania.

Diebold, F.X. and G.D. Rudebusch (1989), Long memory and persistence in aggregate output, *Journal of Monetary Economics* **24**, 189–209.

Eitrheim, Ø and T. Teräsvirta (1996), Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

Franses, P.H. and D. van Dijk (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

Granger, C.W.J. (1993), Strategies for modelling nonlinear time-series relationships, *The Economic Record* **69**, 233–238.

Granger, C.W.J. and R. Joyeux (1980), An introduction to long-range time series models and fractional differencing, *Journal of Time Series Analysis* **1**, 15–30.

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

Granger, C.W.J. and T. Teräsvirta (1999), A simple nonlinear time series model with misleading linear properties, *Economics Letters* **62**, 161–165.

Granger, C.W.J. and N. Hyung (2000), Occasional breaks and long memory, unpublished manuscript, University of California, San Diego.

Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series subject to changes in regime, *Econometrica* **57**, 357–384.

Hansen, B.E. (1997), Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.

Hassler, U. and J. Wolters (1995), Long memory in inflation rates - international evidence, *Journal of Business & Economicd Statistics* **13**, 37–45.

Hosking, J.R.M. (1981), Fractional differencing, *Biometrika* **68**, 165–176.

Hyndman, R.J. (1995), Highest-density forecast regions for nonlinear and nonnormal time series, *Journal of Forecasting* **14**, 431–441.

Hyndman, R.J. (1996), Computing and graphing highest-density regions, *American Statistician* **50**, 120–126.

Koop, G., M.H. Pesaran and S.M. Potter (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119–147.

Koop, G. and S.M. Potter (1999), Dynamic asymmetries in US unemployment, *Journal of Business & Economic Statistics* **17**, 298–313.

Koop, G. and S.M. Potter (2000), Nonlinearity, structural breaks or outliers in economic time series?, in W.A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (eds.), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge: Cambridge University Press.

Koustas, Z. and W. Veloce (1996), Unemployment hysteresis in Canada: an approach based on long-memory time series models, *Applied Economics* **28**, 823–831.

Lundbergh, S., T. Teräsvirta and D. van Dijk (2000), Time-varying smooth transition autoregressive models, SSE/EFI Working Paper Series in Economics and Finance No. 376, Stockholm School of Economics.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Medeiros, M. and A. Veiga (2000), A flexible coefficient smooth transition time series model, SSE/EFI Working Paper Series in Economics and Finance No. 360, Stockholm School of Economics.

Montgomery, A.L., V. Zarnowitz, R.S. Tsay and G.C. Tiao (1998), Forecasting the US unemployment rate, *Journal of the American Statistical Association* **93**, 478–493.

Papell, D.H., C.J. Murray and H. Ghiblawi (2000), The structure of unemployment, *Review of Economics and Statistics* **82**, 309–315.

Potter, S.M. (1994), Asymmetric economic propagation mechanisms, in W. Semmler (ed.), *Business Cycles: Theory and Empirical Methods*, Boston: Kluwer, pp. 313–330.

Ray, B.K. (1993), Modelling long-memory processes for optimal long-range prediction, *Journal of Time Series Analysis* **14**, 511–525.

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

Skalin, J. and T. Teräsvirta (2000), Moving equilibria and asymmetries in unemployment rates, *Macroeconomic Dynamics*, to appear.

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

Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Tschernig, R. and K.F. Zimmermann (1992), Illusive persistence in German unemployment, *Recherches Économiques de Louvain* **58**, 441–453.

van Dijk, D. and P.H. Franses (1999), Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**, 311–340.

van Dijk, D., T. Teräsvirta and P.H. Franses (2000), Smooth transition autoregressive models - a survey of recent developments, SSE/EFI Working Paper Series in Economics and Finance No. 380, Stockholm School of Economics.

van Dijk, D., P.H. Franses and H.P. Boswijk (2000), Asymmetric and common absorption of shocks in nonlinear autoregressive models, Econometric Institute Research Report 2000-01/A.

Wooldridge, J.M. (1994), Estimation and inference for dependent processes, in R.F. Engle and D.L. McFadden (eds.), *Handbook of econometrics, vol. IV*, Amsterdam: Elsevier Science, pp. 2639–2738.

Table 1: Summary of estimated models for US unemployment rate

	FISTAR	ARFI	STAR $d=0$	STAR $d=1$	STAR $d=1$, Δy_{t-1}
$\hat{d}$	0.425	0.835	–	–	–
$\hat{\gamma}$	14.55	–	13.72	15.02	14.12
$\hat{c}$	0.153	–	0.106	0.077	0.109
$T\hat{\sigma}_{\varepsilon}^2$	8.44	9.50	8.53	8.85	8.37
AIC	−3.68	−3.61	−3.68	−3.65	−3.69
BIC	−3.54	−3.55	−3.56	−3.53	−3.55
SK	0.34	0.37	0.34	0.15	0.33
	(4.6×10^{-3})	(2.0×10^{-3})	(4.7×10^{-3})	(0.12)	(5.3×10^{-3})
EK	1.35	1.73	1.10	1.57	1.11
	(8.9×10^{-8})	(1.1×10^{-11})	(9.9×10^{6})	(6.7×10^{-10})	(8.2×10^{-6})
JB	34.0	53.1	25.0	38.1	25.1
	(4.1×10^{-8})	(3.0×10^{-12})	(3.8×10^{-6})	(5.4×10^{-9})	(3.6×10^{-6})
ARCH(1)	10.9	15.4	12.3	10.5	10.4
	(9.8×10^{-4})	(8.7×10^{-5})	(4.6×10^{-4})	(1.2×10^{-3})	(1.3×10^{-3})
ARCH(4)	18.1	20.5	18.2	15.9	14.7
	(1.2×10^{-3})	(4.0×10^{-4})	(1.1×10^{-3})	(3.2×10^{-3})	(5.3×10^{-3})
ARCH(12)	28.0	28.4	27.8	25.7	25.8
	(5.6×10^{-3})	(4.8×10^{-3})	(5.9×10^{-3})	0.012	0.012
$\text{LM}_{\text{SC}}(1)$	0.17	2.62	0.41	0.34	0.18
	(0.68)	(0.11)	(0.52)	(0.56)	(0.67)
$\text{LM}_{\text{SC}}(4)$	1.51	1.26	2.42	1.66	1.35
	(0.20)	(0.29)	(0.048)	(0.16)	(0.25)
$\text{LM}_{\text{SC}}(12)$	1.05	1.11	1.43	1.54	1.04
	(0.40)	(0.35)	(0.15)	(0.11)	(0.41)

The table presents diagnostic tests for the estimated FISTAR and ARFI models, and STAR models for levels ($d = 0$), first differences ($d = 1$) and first differences with lagged level term ($d = 1$, Δy_{t-1}) for the US unemployment rate over the sample period January 1970-December 1999 ($T = 360$). $\hat{\sigma}_{\varepsilon}^2$ denotes the residual variance, SK is skewness, EK excess kurtosis, JB the Jarque-Bera test of normality of the residuals, ARCH(r) is the LM test of no Autoregressive Conditional Heteroscedasticity [ARCH] up to order r, and $\text{LM}_{\text{SC}}(q)$ denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below the test statistics are p-values.

Table 2: LM-type tests of (no remaining non-)linearity and parameter constancy

Transition variable	LM_1	LM_2	LM_3
		ARFI model	
$\Delta_{12}y_{t-1}$	4.0×10^{-4}	2.1×10^{-4}	3.5×10^{-4}
$\Delta_{12}y_{t-2}$	2.1×10^{-3}	6.7×10^{-4}	1.4×10^{-3}
$\Delta_{12}y_{t-3}$	0.012	1.1×10^{-3}	8.6×10^{-3}
$\Delta_{12}y_{t-4}$	0.020	1.1×10^{-3}	5.5×10^{-4}
$\Delta_{12}y_{t-5}$	0.060	8.8×10^{-4}	3.7×10^{-3}
$\Delta_{12}y_{t-6}$	0.090	2.3×10^{-3}	5.1×10^{-3}
t	0.020	0.093	0.16
		FISTAR model	
$\Delta_{12}y_{t-1}$	0.62	0.31	0.40
$\Delta_{12}y_{t-2}$	0.41	0.35	0.39
$\Delta_{12}y_{t-3}$	0.51	0.33	0.42
$\Delta_{12}y_{t-4}$	0.50	0.33	0.52
$\Delta_{12}y_{t-5}$	0.32	0.38	0.59
$\Delta_{12}y_{t-6}$	0.47	0.44	0.55
t	0.39	0.41	0.47

p-values of LM-type test statistics of (no remaining non-)linearity and parameter constancy in the ARFI(p,d) model with $p = 4$ and the FISTAR model for the US unemployment rate.

Table 3: Asymmetry measures for impulse responses in FISTAR and STAR models

h	Unconditional				Recession				Expansion			
	A	S	M	L	A	S	M	L	A	S	M	L
						FISTAR model						
6	−0.00	−0.12*	0.21*	0.88*	−0.01	−0.02	0.03	0.05	−0.00	−0.17*	0.28*	1.22*
12	0.01	−0.14	0.27	1.04*	−0.02	0.05	−0.14	−0.53	0.02	−0.22*	0.44*	1.67*
24	0.01	−0.14	0.28	1.01*	−0.03	0.10	−0.26	−0.90*	0.03	−0.23*	0.50*	1.77*
36	0.01	−0.09	0.18	0.64*	−0.02	0.06	−0.17	−0.55*	0.02	−0.15*	0.32*	1.11*
48	0.00	−0.03	0.07	0.23	−0.01	0.01	−0.04	−0.09	0.01	−0.05	0.11*	0.35*
60	0.00	0.00	−0.00	−0.01	−0.01	−0.02	0.01	0.10*	0.00	0.01	−0.01	−0.06
72	−0.00	0.01	−0.02	−0.05	−0.01	−0.02	0.00	0.07	0.00	0.02	−0.03	−0.10*
84	−0.00	−0.01	0.01	0.04	−0.01	−0.00	−0.02	−0.03	0.00	−0.01	0.02	0.06
96	0.00	−0.02	0.04	0.14*	0.00	0.02	−0.03	−0.11	0.00	−0.03	0.06*	0.24*
108	0.01	−0.02	0.05	0.19*	0.01	0.03	−0.03	−0.13	0.01	−0.04	0.09*	0.32*
120	0.01	−0.02	0.05	0.19*	0.01	0.03	−0.03	−0.11	0.01	−0.03	0.09*	0.31*
						STAR model						
6	−0.00	−0.12*	0.19*	0.82*	−0.02	−0.05	0.05	0.18	0.00	−0.15*	0.26*	1.10*
12	0.01	−0.12*	0.24	0.94*	−0.02	0.02	−0.08	−0.28	0.02	−0.19*	0.38*	1.49*
24	0.01	−0.10	0.21	0.76*	−0.01	0.07	−0.16	−0.56	0.02	−0.17*	0.37*	1.35*
36	0.01	−0.06	0.12	0.45*	0.00	0.02	−0.03	−0.10	0.01	−0.10*	0.19*	0.69*
48	0.00	−0.01	0.03	0.10	0.01	−0.03	0.09*	0.30*	−0.00	−0.00	0.00	0.02
60	−0.00	0.02	−0.04	−0.16	0.02	−0.03	0.11*	0.36*	−0.01	0.05	−0.11*	−0.39*
72	−0.00	0.03	−0.06	−0.24*	0.01	−0.02	0.06	0.19*	−0.01	0.05*	−0.12*	−0.44*
84	−0.00	0.02	−0.04	−0.17*	−0.00	−0.00	0.00	0.00	−0.00	0.03	−0.06*	−0.25*
96	0.00	0.01	−0.01	−0.04	−0.01	0.01	−0.03	−0.10*	0.01	0.01	0.00	−0.02
108	0.01	0.00	0.02	0.06	−0.00	0.02	−0.03	−0.10*	0.01	−0.01	0.04*	0.13*
120	0.01	−0.00	0.02	0.08*	0.00	0.01	−0.02	−0.05	0.01	−0.01	0.04*	0.14*

Means of the asymmetry measure $\text{ASY}_y(h, \mathcal{S}, \mathcal{H})$ in the estimated FISTAR model and STAR model for first differences with lagged level term. Means larger than two times $\sigma_{\text{ASY}_y(h,\mathcal{S},\mathcal{H})}/\sqrt{n_\mathcal{H}}$ are marked with an asterisk, where $\sigma_{\text{ASY}_y(h,\mathcal{S},\mathcal{H})}$ is the standard deviation of $\text{ASY}_y(h, \mathcal{S}, \mathcal{H})$ and n_A is the number of shocks δ for which $\text{ASY}_y(h, \delta, \omega_{t-1})$ is computed. The different sets of shocks are defined as A(ll) $= \{\varepsilon_t | \varepsilon_t > 0\}$, S(mall) $= \{\varepsilon_t | 1 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 0\}$, M(edium) $= \{\varepsilon_t | 2 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 1\}$ and L(arge) $= \{\varepsilon_t | 3 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 2\}$. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})$ is smaller and larger than 0.5, respectively.

Table 4: Absorption times in FISTAR and STAR models

h	Unconditional				Recession				Expansion			
	A	S	M	L	A	S	M	L	A	S	M	L
					FISTAR model							
1.00	27.8	28.1	26.9	26.8	30.4	30.9	29.0	28.8	26.7	27.0	26.0	26.0
0.75	37.0	38.2	34.2	33.2	37.9	39.2	34.8	33.8	36.7	37.9	34.0	33.0
0.50	52.0	54.0	47.2	48.2	52.7	54.5	48.5	47.4	51.7	53.7	46.7	48.5
0.25	88.9	91.2	83.7	81.1	91.3	93.3	86.9	83.3	88.0	90.4	82.4	80.3
					STAR model							
1.00	25.4	25.7	24.7	24.5	29.7	30.2	28.7	27.6	23.5	23.8	22.9	23.2
0.75	35.6	36.1	34.4	34.5	37.7	38.4	36.1	35.2	34.6	35.0	33.6	34.2
0.50	50.3	51.3	47.7	47.6	51.8	52.8	49.2	49.1	49.6	50.7	47.0	47.0
0.25	81.4	82.6	78.4	78.3	83.1	84.1	80.7	79.8	80.6	82.0	77.4	77.7

Means of $\mathrm{N}_Y(\pi, \mathcal{S}, \mathcal{H})$ in the estimated FISTAR model and STAR model for first differences with lagged level term. The different sets of shocks are defined as A(ll) = $\{\varepsilon_t | \varepsilon_t > 0\}$, S(mall) = $\{\varepsilon_t | 1 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 0\}$, M(edium) = $\{\varepsilon_t | 2 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 1\}$ and L(arge) = $\{\varepsilon_t | 3 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 2\}$. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})$ is smaller and larger than 0.5, respectively.

Table 5: Asymmetry measures for absorption times in FISTAR and STAR models

h	Unconditional				Recession				Expansion			
	A	S	M	L	A	S	M	L	A	S	M	L
					FISTAR model							
1.00	−4.1*	−7.0*	2.4	5.9	2.5*	4.4*	−1.9	−4.4	−6.7*	−11.5*	4.1*	10.0*
0.75	−3.7*	−6.3	2.6	4.0*	2.8	4.6	−1.0	−4.5*	−6.3*	−10.7	4.1	7.5*
0.50	−4.3	−8.3	4.2	13.7	3.8	7.5	−4.7	−9.0	−7.5*	−14.6*	7.8	22.8*
0.25	−0.2	−3.4	6.6	14.0	3.9	8.1	−4.4	−19.8*	−1.8	−8.0	11.0	27.5*
					STAR model							
1.00	−3.8*	−6.7	2.7	6.3*	1.1	1.9	−0.7	−2.1	−6.1*	−10.6*	4.2*	10.1*
0.75	−4.1*	−7.2	2.6	7.9*	1.6	3.4	−2.7	−3.0	−6.7*	−11.9*	5.0*	12.7*
0.50	−3.4*	−6.6	3.6	9.4*	3.5*	6.2	−2.4	−7.0	−6.5*	−12.3*	6.3*	16.7*
0.25	−0.1	−1.9	3.6	7.2*	3.5	5.5	−1.0	−3.7	−1.8	−5.2	5.7*	12.1*

Means of the asymmetry measure $\mathrm{ASYN}_y(h, \mathcal{S}, \mathcal{H})$ in the estimated FISTAR model and STAR model for first differences with lagged level term. Means larger than two times $\sigma_{\mathrm{ASYN}_y(h,\mathcal{S},\mathcal{H})}/\sqrt{n_\mathcal{H}}$ are marked with an asterisk, where $\sigma_{\mathrm{ASYN}_y(h,\mathcal{S},\mathcal{H})}$ is the standard deviation of $\mathrm{ASYN}_y(h, \mathcal{S}, \mathcal{H})$ and n_A is the number of shocks δ for which $\mathrm{ASYN}_y(h, \delta, \omega_{t-1})$ is computed. The different sets of shocks are defined as A(ll) = $\{\varepsilon_t | \varepsilon_t > 0\}$, S(mall) = $\{\varepsilon_t | 1 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 0\}$, M(edium) = $\{\varepsilon_t | 2 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 1\}$ and L(arge) = $\{\varepsilon_t | 3 \geq \varepsilon_t/\hat{\sigma}_\varepsilon > 2\}$. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12} y_{t-1}; \hat{\gamma}, \hat{c})$ is smaller and larger than 0.5, respectively.

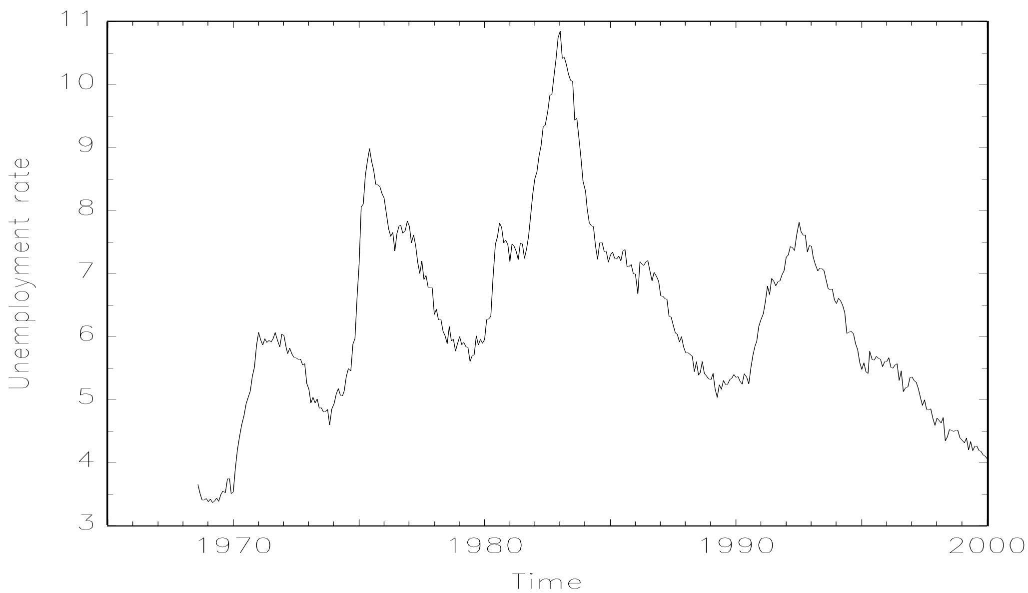
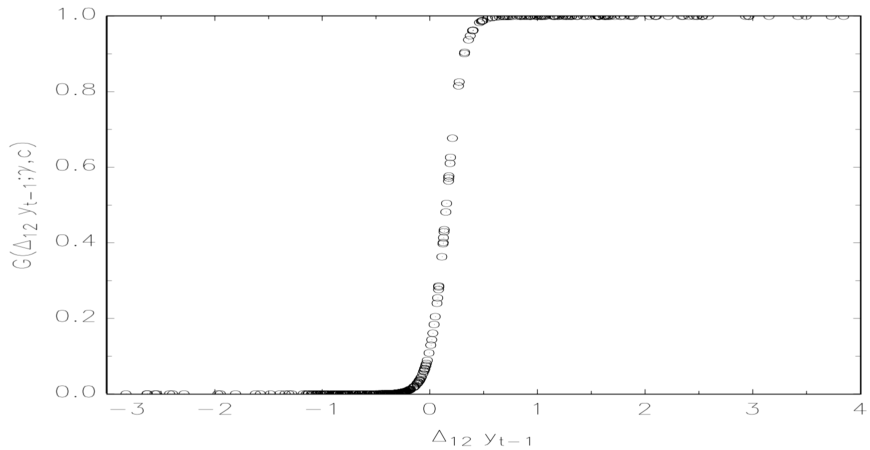
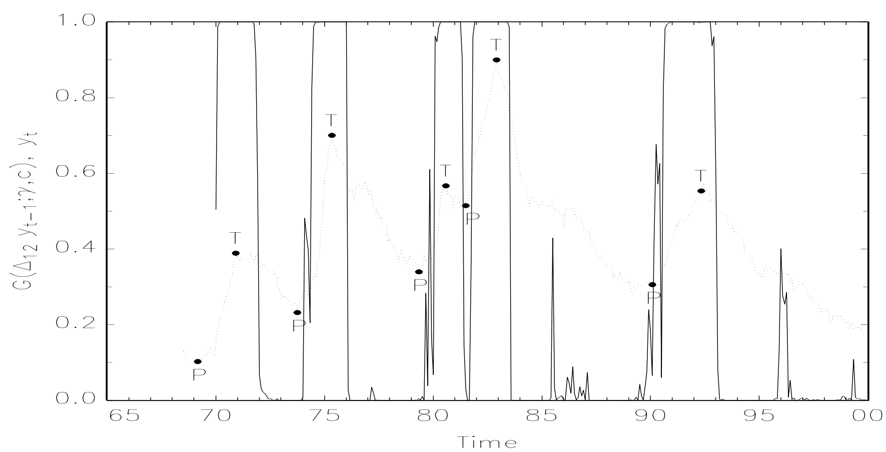


Figure 1: Monthly seasonally adjusted US unemployment rate, July 1968-December 1999.



(a) Transition function versus $\Delta_{12}y_{t-1}$



(b) Transition function versus time

Figure 2: Transition function in FISTAR model for monthly US unemployment rate against the transition variable $\Delta_{12}y_{t-1}$ and over time. The dotted line represents the rescaled monthly unemployment rate. Solid circles indicate NBER-dated unemployment peaks (P) and troughs (T).

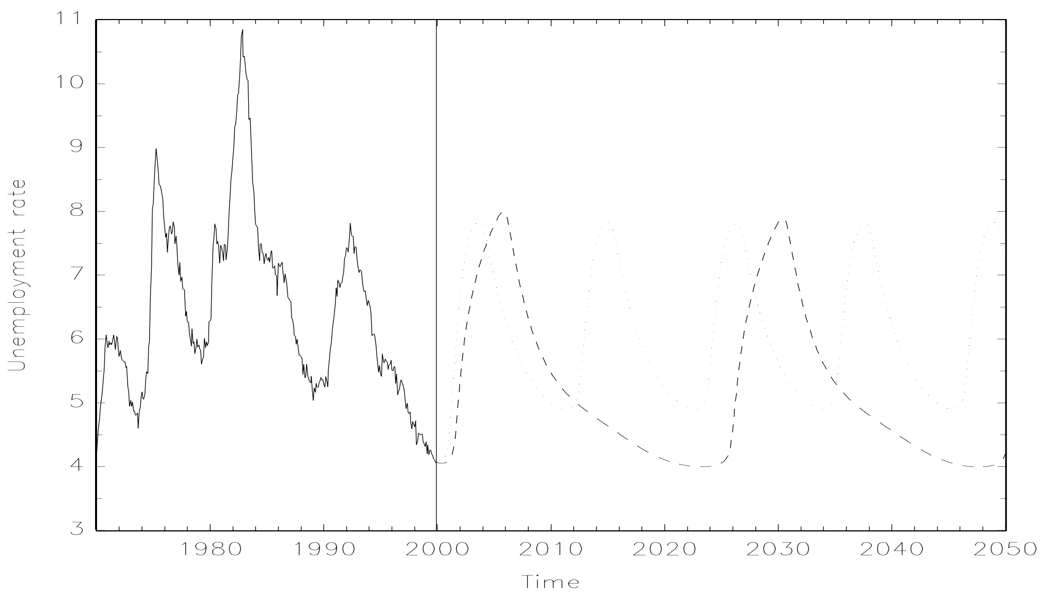
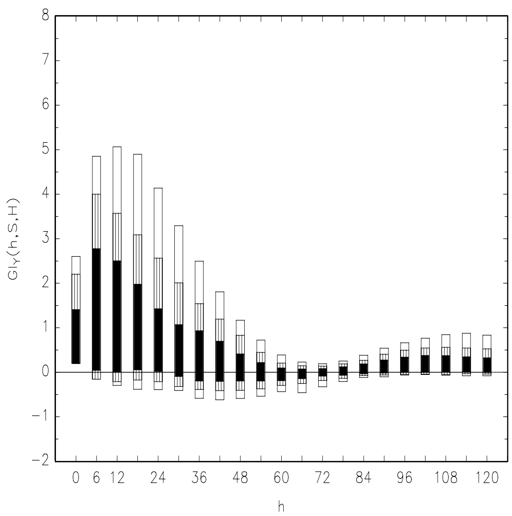
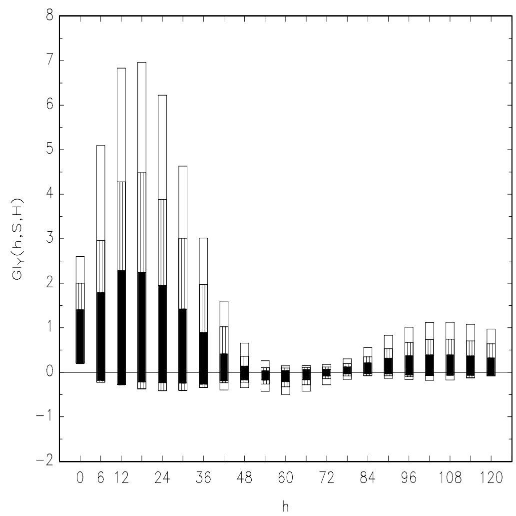


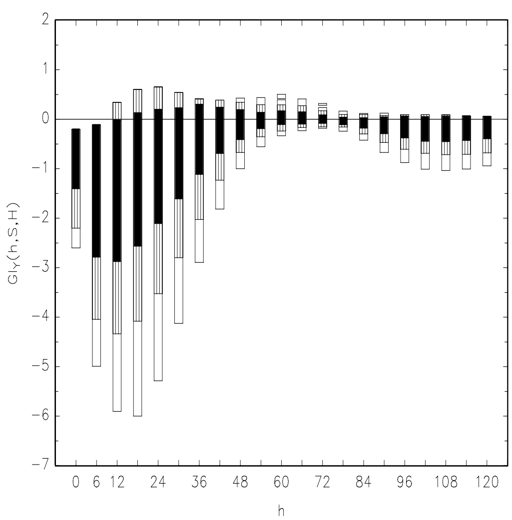
Figure 3: Deterministic extrapolation of the FISTAR model (dashed line) and the STAR model for first differences with a lagged level term (dotted line).



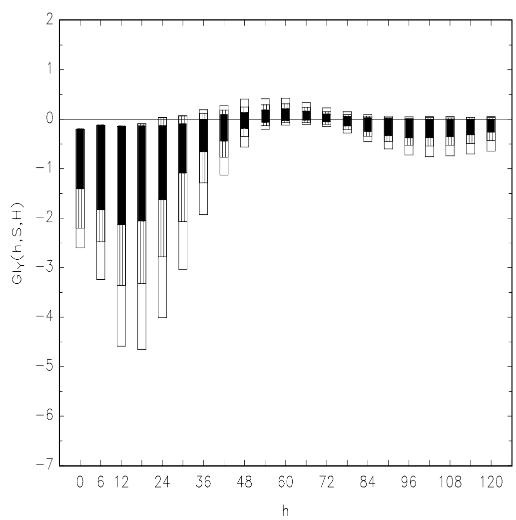
(a) positive shocks, recession



(b) positive shocks, expansion

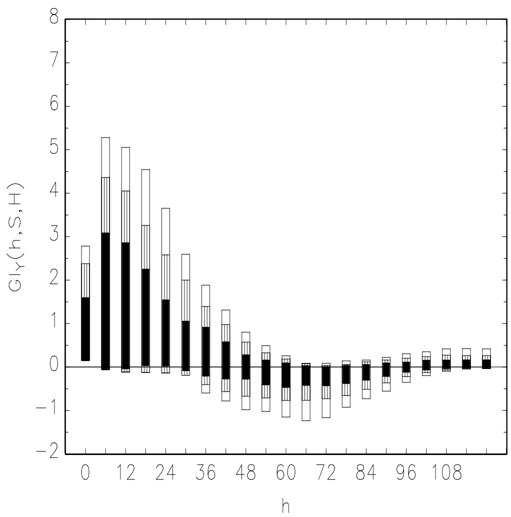


(c) negative shocks, recession

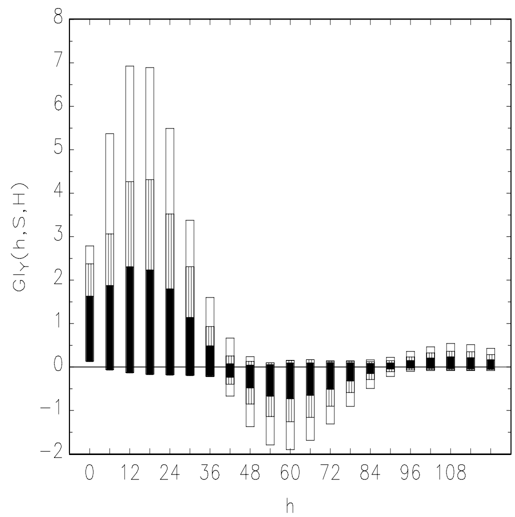


(d) negative shocks, expansion

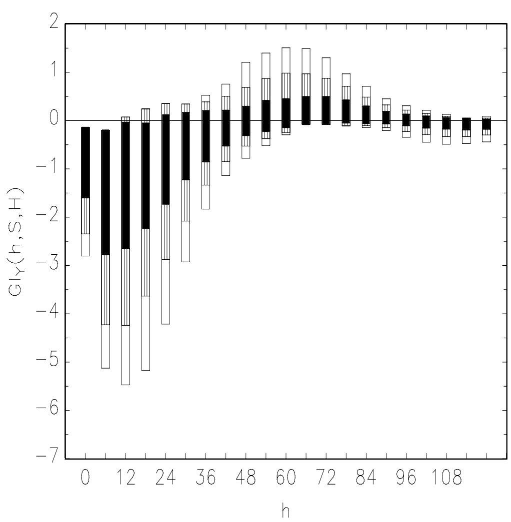
Figure 4: 50% (black), 75% (hatched) and 90% (white) highest density regions for generalized impulse responses in the FISTAR model estimated for the US unemployment rate. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12}y_{t-1};\widehat{\gamma},\widehat{c})$ is larger and smaller than 0.5, respectively.



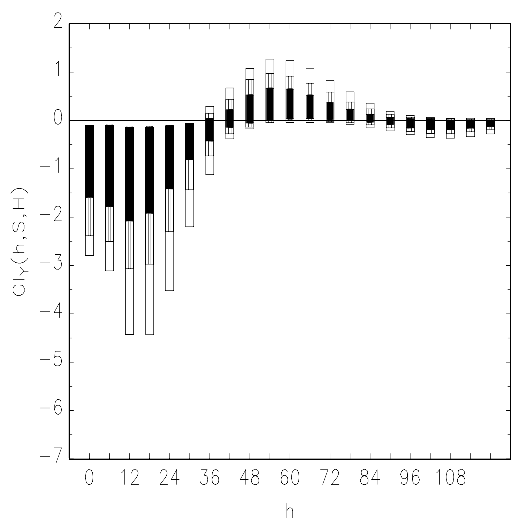
(a) positive shocks, recession



(b) positive shocks, expansion

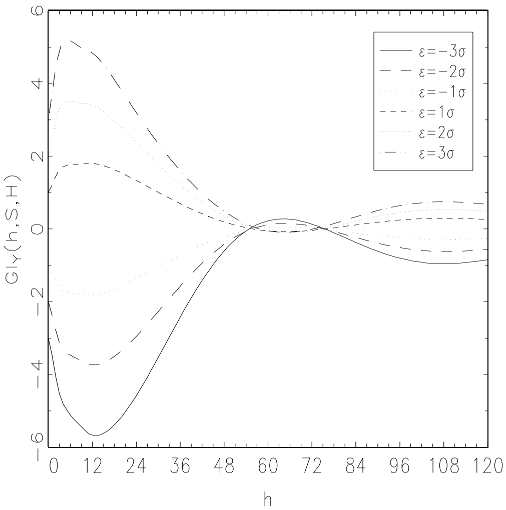


(c) negative shocks, recession

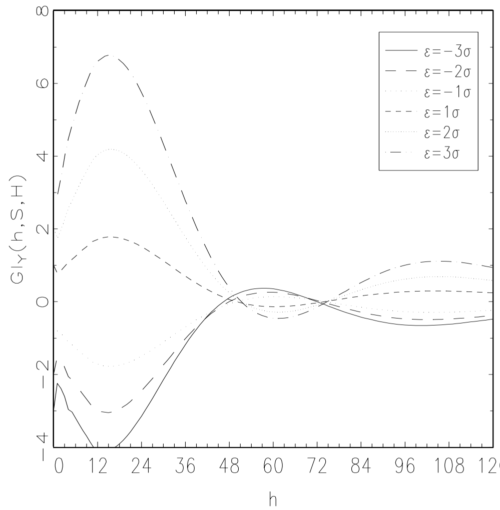


(d) negative shocks, expansion

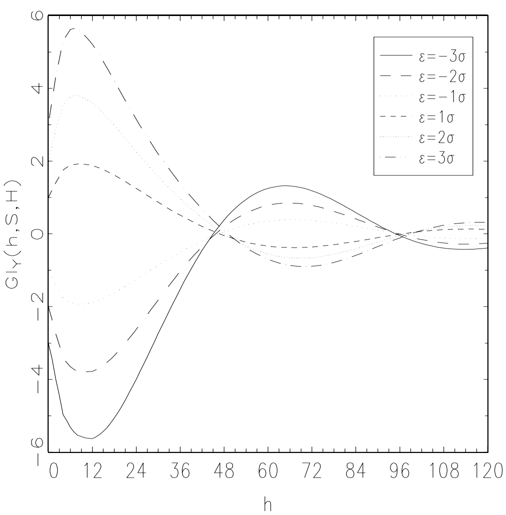
Figure 5: 50% (black), 75% (hatched) and 90% (white) highest density regions for generalized impulse responses in the STAR model for first differences with lagged level term estimated for the monthly US unemployment rate. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12}y_{t-1};\widehat{\gamma},\widehat{c})$ is larger and smaller than 0.5, respectively.



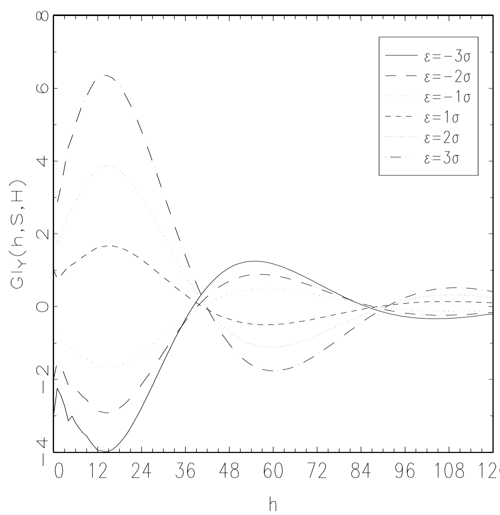
(a) FISTAR, recession



(b) FISTAR, expansion



(c) STAR, recession



(d) STAR, expansion

Figure 6: Mean of generalized impulse responses in the FISTAR model and STAR model for first differences with lagged level term estimated for the monthly US unemployment rate. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12} y_{t-1}; \widehat{\gamma}, \widehat{c})$ is larger and smaller than 0.5, respectively.