# Modeling Asymmetric Persistence over the Business Cycle

Philip Hans Franses*
*Econometric Institute*
*Erasmus University Rotterdam*

and

Richard Paap†
*RIBES*
*Erasmus University Rotterdam*

Econometric Institute Report 9852/A

**Abstract**

We address the issue of time varying persistence of shocks to macroeconomic time series variables by proposing a new and parsimonious time series model. Our model assumes that this time varying persistence depends on a linear combination of lagged explanatory variables, where this combination characterizes the business cycle regimes. The key feature of our model is that an autoregressive parameter takes larger values only when this indicator variable exceeds a stochastic threshold. The parameters and the (lags of the) variables that constitute the indicator variable have to be determined from the data. Other forms of censoring amount to straightforward extensions. Our application to US unemployment shows that the model fits very well. A linear combination of lagged (differenced) industrial production, oil price, interest spread and stock returns amounts to an adequate indicator of an upcoming recession, which corresponds with explosive behavior of unemployment. Also, the out-of-sample forecasts from our model oftentimes improve those from linear and other nonlinear models.

**Keywords**: asymmetric persistence, nonlinear time series, business cycle, censored regression.

*Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL - 3000 DR Rotterdam, The Netherlands, e-mail: franses@few.eur.nl.

†Corresponding author: Erasmus University Rotterdam, RIBES, P.O. Box 1738, NL - 3000 DR Rotterdam, The Netherlands, e-mail: paap@few.eur.nl.

# 1 Introduction

There is ample empirical evidence for asymmetries in the business cycle. Usually these asymmetries are found to correspond to different dynamic properties over the business cycle of time series of aggregated macroeconomic variables like output and unemployment. Early references and important more recent studies include Neftçi (1984), Hamilton (1989) and Teräsvirta and Anderson (1992), among many others. An interesting object of research, which corresponds to these asymmetries, relates to the question whether shocks to macroeconomic aggregates have different long-run impact across the business cycle stages, see for example Beaudry and Koop (1993) and Elwood (1998). In the present paper, we also address this issue of time varying persistence of shocks, by proposing a new and parsimonious time series model and applying it to US unemployment.

Our new (nonlinear) time series model contains a couple of features, which are in clear distinction to related models proposed in the literature. First of all, we allow the time varying persistence of shocks to depend explicitly on exogenous variables, which together characterize the business cycle. This in contrast to related time varying parameter models, such as for example the stochastic (or randomized) unit root process, see Granger and Swanson (1997) and Leybourne, McCabe and Tremayne (1996), where the generating mechanism for the parameter variation is purely random. Consequently, these models need the Kalman filter for parameter estimation, see also Grillenzoni (1993), whereas our model yields closed-form expressions for the likelihood. Secondly, by making persistence to depend on exogenous variables, where the dependence has to be estimated using the available data, we do not *a priori* assume knowledge of the business cycle peaks and troughs. In fact, a useful by-product of our model is that we can estimate the turning points, based on the data at hand. As we are uncertain about the exact dates of these turning points, we introduce an additional error term, which indicates the confidence we can have about the obtained chronology. Finally, and this is mainly governed by our application to unemployment data, we impose that the persistence of shocks is higher in recessions than in expansions. This corresponds with the findings in Blanchard and

Summers (1987) and Bianchi and Zoega (1998), where it is found that in recessions unemployment tends to display seemingly explosive behavior. It must be stressed here though that our model can easily be adapted for other applications.

The outline of our paper is as follows. In Section 2 we introduce our AutoRegressive model with Censored Latent Effects Parameters [AR-CLEP]. As we aim to allow for an increase of the persistence (AR parameter) in recessions only, we introduce censored latent effects of lagged explanatory variables. In Section 3, we outline some possible extensions of our basic model and we compare it with closely-related nonlinear models. In Section 4, we apply our model to monthly US unemployment, and compare its out-of-sample forecasting quality with some of its alternatives. Finally, in Section 5 we conclude with a few remarks.

# 2 The AR-CLEP Model

In this section we start off with a discussion of the representation of the simple AR(1)-CLEP model. Next, we consider parameter estimation in Section 2.2, and the construction of residuals for diagnostic purposes in Section 2.3. Finally, in Section 2.4 we show how out-of-sample forecasts can be generated.

## 2.1 Representation

Consider the following AR(1) model with a time varying autoregressive parameter for a time series $\{y_t\}_{t=1}^T$

$$y_t = \mu + (\rho + \rho_t) y_{t-1} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$, and $\rho_t$ is a censored latent variable defined by

$$\rho_t = \begin{cases} x_t'\beta + u_t & \text{if } x_t'\beta + u_t \geq 0 \\ 0 & \text{if } x_t'\beta + u_t < 0 \end{cases} \tag{2}$$

with $u_t \sim \text{NID}(0, \sigma_u^2)$, $x_t$ a $(k \times 1)$ vector of exogenous variables including a constant and with $\beta$ an unknown $(k \times 1)$ parameter vector. Hence, the autoregressive parameter of the AR model equals $\rho$ unless $x_t'\beta$ is larger than the stochastic threshold level $-u_t$, in which case the AR coefficient equals $\rho + \rho_t$, indicating that there is more persistence in the time series. Notice that $\rho_t$ is not fixed, but takes values that depend on $x_t$ and $\beta$. If $\sigma_u = 0$, (1) with (2) reduces to a threshold model with a nonstochastic threshold value. By allowing $u_t \neq 0$, we introduce more uncertainty concerning the timing of a higher degree of persistence.

As $\rho_t$ in (1) is a latent variable we can only make probability statements about its values. The probability that $\rho_t = 0$ equals the probability that $x_t'\beta + u_t < 0$, that is

$$\Pr[\rho_t = 0|x_t] = \int_{-\infty}^{-x_t'\beta} \frac{1}{\sigma_u}\phi(u_t/\sigma_u)du_t = \Phi\left(\frac{-x_t'\beta}{\sigma_u}\right), \tag{3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function [pdf] and the cumulative density function [CDF] of the standard normal distribution, respectively. The probability that $\rho_t \neq 0$ is thus given by

$$\Pr[\rho_t \neq 0|x_t] = \int_{-x_t'\beta}^{\infty} \frac{1}{\sigma_u}\phi(u_t/\sigma_u)du_t = 1 - \Phi\left(\frac{-x_t'\beta}{\sigma_u}\right) = \Phi\left(\frac{x_t'\beta}{\sigma_u}\right) \tag{4}$$

The expected value of the autoregressive parameter $\rho + \rho_t$ follows from

$$\begin{aligned}
\mathrm{E}[\rho_t|x_t] &= \Pr[\rho_t = 0|x_t]\ \mathrm{E}[\rho_t|\rho_t = 0, x_t] + \Pr[\rho_t \neq 0|x_t]\ \mathrm{E}[\rho_t|\rho_t \neq 0, x_t] \\
&= 0 + \Phi\left(\frac{x_t'\beta}{\sigma_u}\right)\left(x_t'\beta + \sigma_u \frac{\phi(x_t'\beta/\sigma_u)}{1 - \Phi(-x_t'\beta/\sigma_u)}\right) \\
&= x_t'\beta\ \Phi\left(\frac{x_t'\beta}{\sigma_u}\right) + \sigma_u \phi\left(\frac{x_t'\beta}{\sigma_u}\right),
\end{aligned} \tag{5}$$

where we use that if $z \sim \mathrm{N}(a, \sigma^2)$ then

$$\mathrm{E}[z|z > b] = a + \sigma \frac{\phi((b-a)/\sigma)}{1 - \Phi((b-a)/\sigma)}\ , \tag{6}$$

see Johnson and Kotz (1970, p. 81–83).

## 2.2 Estimation

The parameters in the model (1) with (2) can be estimated using maximum likelihood. The model parameters are given by $\theta = \{\mu, \rho, \sigma_\varepsilon, \beta, \sigma_u\}$. To derive the likelihood function, we first consider the pdf of $y_t$ given its past contained in $Y_{t-1} = \{y_{t-1}, \ldots, y_1\}$ and given $\rho_t$, that is,

$$f(y_t|Y_{t-1}, \rho_t; \theta) = \frac{1}{\sigma_\varepsilon}\phi\left(\frac{e_t - \rho_t y_{t-1}}{\sigma_\varepsilon}\right). \tag{7}$$

Note for (7) that we write $\varepsilon_t$ as the difference between $e_t = y_t - \mu - \rho y_{t-1}$ and $\rho_t y_{t-1}$, to indicate that the pdf is conditional on $\rho_t$. To obtain the unconditional pdf of $y_t$ given $Y_{t-1}$, we have to integrate over the unobserved variable $\rho_t$. Therefore, the unconditional pdf of $y_t$ given $Y_{t-1}$ and $x_t$ equals

$$
\begin{aligned}
f(y_t|Y_{t-1}, x_t; \theta) = \Pr[\rho_t = 0|x_t] f(y_t|Y_{t-1}, x_t, \rho_t; \theta)|_{\rho_t=0} \\
+ \int_{-x_t'\beta}^{\infty} \frac{1}{\sigma_u}\phi\left(\frac{u_t}{\sigma_u}\right) f(y_t|Y_{t-1}, x_t, \rho_t; \theta)|_{\rho_t = x_t\beta+u_t} du_t, \quad (8)
\end{aligned}
$$

where $\Pr[\rho_t = 0|x_t]$ is defined in (3). The log likelihood function simply equals the sum of the logarithms of these unconditional pdfs, that is,

$$
\mathcal{L}(Y_T|X_T; \theta) = \sum_{t=1}^{T} \ln(f(y_t|Y_{t-1}, x_t; \theta)), \quad (9)
$$

where $X_T = (x_T, \ldots, x_1)$. This likelihood function can be maximized using standard numerical optimization algorithms, like for example Gauss-Newton. For further use, in the appendix we derive the first order derivatives of the log likelihood function, which determine the first order conditions and provide estimators for the standard errors. We can decrease the computational burden of calculating the integral in (8) using

$$
\begin{aligned}
&\int_b^{\infty} \frac{1}{\sigma_1}\phi(z/\sigma_1)\frac{1}{\sigma_2}\phi((z-a)/\sigma_2)dz \\
&= \int_b^{\infty} \frac{1}{\sigma_1\sqrt{2\pi}} \exp(-\frac{1}{2}(z/\sigma_1)^2) \frac{1}{\sigma_2\sqrt{2\pi}} \exp(-\frac{1}{2}((z-a)/\sigma_2)^2) \\
&= \frac{\sigma}{\sigma_1\sigma_2\sqrt{2\pi}} \exp(\frac{a^2}{2}(\sigma^2\sigma_2^{-4} - \sigma_2^{-2})) \int_b^{\infty} \frac{1}{\sigma}\phi((z - \sigma_2^{-2}\sigma^2 a)/\sigma)dz \\
&= \frac{\sigma}{\sigma_1\sigma_2\sqrt{2\pi}} \exp(\frac{a^2}{2\sigma_2^2}(\sigma^2/\sigma_2^2 - 1))\Phi((\sigma_2^{-2}\sigma^2 a - b)/\sigma)
\end{aligned} \quad (10)
$$

where $\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. Note that in our case $z = u_t$, $\sigma_1 = \sigma_u$, $\sigma_2 = \sigma_\varepsilon/y_{t-1}$, $b = -x_t'\beta$, and $a = (y_t - (\rho + x_t'\beta)y_{t-1})/y_{t-1}$.

Once the parameters are estimated, we can have a closer look at the value of $\rho_t$. The conditional probability that $\rho_t = 0$ given $Y_t$ and $x_t$ equals

$$
\Pr[\rho_t = 0|Y_t, x_t] = \frac{\Pr[\rho_t = 0|x_t] f(y_t|Y_{t-1}, x_t, \rho_t; \theta)|_{\rho_t=0}}{f(y_t|Y_{t-1}, x_t; \theta)}. \quad (11)
$$

These probabilities can be used to analyze whether there is a higher degree of persistence at time $t$. For this purpose we use the conditional expected value of $\rho_t$ given $Y_t$, which reads as

$$\mathrm{E}[\rho_t|Y_t, x_t] = \frac{\int_{-x_t'\beta}^{\infty}(x_t'\beta + u_t)\frac{1}{\sigma_u}\,\phi\left(\frac{u_t}{\sigma_u}\right) f(y_t|Y_{t-1}, x_t, \rho_t; \theta)|_{\rho_t = x_t'\beta + u_t} du_t}{f(y_t|Y_{t-1}, x_t; \theta)}, \qquad (12)$$

which gives an estimate of the additional persistence in the time series at time $t$.

### 2.3 Residuals and Diagnostics

As the added autoregressive parameter $\rho_t$ in the AR-CLEP model is unobserved, there are several options to calculate residuals. Here, we define residuals as the expectation of $\varepsilon_t$ given $Y_t$ and $x_t$

$$\begin{aligned}\hat{\varepsilon}_t &= \mathrm{E}[\varepsilon_t|Y_t, x_t] \\ &= \mathrm{E}[y_t - \hat{\mu} - \hat{\rho}y_{t-1} - \rho_t y_{t-1}|Y_t, x_t] \\ &= y_t - \hat{\mu} - \hat{\rho}y_{t-1} - \mathrm{E}[\rho_t|Y_t, x_t]y_{t-1},\end{aligned} \qquad (13)$$

where $\mathrm{E}[\rho_t|Y_t, x_t]$ is defined in (12). These residuals can be used in diagnostic tests of the adequacy of the model.

To test for the presence of serial correlation in the residuals we consider the auxiliary regression

$$\hat{\varepsilon}_t = \omega + \alpha\hat{\varepsilon}_{t-1} + \eta_t. \qquad (14)$$

The null hypothesis of no first order serial correlation ($\alpha = 0$) can be tested using an $F$-test. Tests for higher order serial correlation can be constructed by adding higher order lags of $\hat{\varepsilon}_t$. Likewise, the hypothesis of no first order ARCH effects can be examined via the auxiliary regression

$$\hat{\varepsilon}_t^2 = \kappa + \psi\hat{\varepsilon}_{t-1}^2 + \xi_t. \qquad (15)$$

The null hypothesis corresponds with the restriction $\psi = 0$, and it can again be tested with an $F$-test. Tests for higher order ARCH effects proceed in the same way. Finally, we construct a $\chi^2(2)$ test for normality of the residuals.

## 2.4 Forecasting

The AR-CLEP model in (1) and (2) uses exogenous variables to explain the parameter variation the time series model. Forecasts from this model are therefore conditional on the values of these explanatory variables.

First, we consider the out-of-sample forecast of the value of the AR coefficient. A forecast for $\rho_{T+1}$ given $X_{T+1}$ is given by the unconditional expectation of $\rho_{T+1}$ given $x_{T+1}$, see (5). Multi-step ahead forecasts of $\rho_t$ can be computed in the same way. If the vector $x_t$ contains lagged explanatory variables, some of their future values may be known at time $T$. However, for multi-step ahead forecasts it is likely that some future values of the explanatory variables are unknown at time $T$ and that they themselves have to be replaced by forecasts.

The one-step ahead forecast at time $T$ conditional on $Y_T$ and $x_{T+1}$ follows from

$$\mathrm{E}[y_{T+1}|Y_T, x_{T+1}] = \int_{-\infty}^{\infty} y_{T+1} f(y_{T+1}|Y_T, x_{T+1}; \theta) dy_{T+1}, \tag{16}$$

where $f(y_{T+1}|Y_T, x_{T+1}; \theta)$ is defined as in (8). In general, the $h$-step ahead forecast, conditional on $(x_{T+h}, \ldots, x_{T+1})$ and $Y_T$, equals

$$\begin{aligned}&\mathrm{E}[y_{T+h}|Y_T, x_{T+h}, \ldots, x_{T+1}] = \\ &\quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_{T+h} f(y_{T+h}|Y_{T+h-1}, x_{T+h}; \theta) \cdots f(y_{T+1}|Y_T, x_{T+1}; \theta)\, dy_{T+h} \cdots dy_{T+1}.\end{aligned} \tag{17}$$

If it is not tractable to evaluate the multiple integrals using numerical integration, we can resort to straightforward simulation techniques.

# 3 Extensions and Relation to Other Models

In this section we consider some extensions of our basic model in (1) and (2). Next, we relate our model to some popular alternative models for modeling time varying autoregressive parameters.

## 3.1 Extensions

Our model can be extended in several ways. The most obvious extension is to allow for higher order dynamics in the autoregressive part of the model, that is in equation (1). If one wants to use the sum of the autoregressive parameters as a measure of persistence one may opt for the following AR($p$) model

$$y_t = (\rho + \rho_t) y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \varepsilon_t, \tag{18}$$

as the $\rho$ $(+\rho_t)$ parameter equals the sum of the autoregressive parameters of an AR($p$) model in levels.

Another possibility is to consider an AR($p-1$) model for $(y_t - (\rho + \rho_t) y_{t-1})$ that is

$$(y_t - (\rho + \rho_t) y_{t-1}) = \sum_{i=1}^{p-1} \alpha_i (y_{t-i} - (\rho + \rho_t) y_{t-1-i}) + \varepsilon_t, \tag{19}$$

where the persistence measure now concerns the largest root in the autoregression. Extensions to moving average structures proceed in a similar way.

With our application below in mind, we currently assume that changes in the autoregressive parameter are into one direction. Of course, one may also want to allow for changes in the other or in both directions. This can be done by allowing $\rho_t$ to enter the regression (1) if $x_t'\beta$ is smaller than $c + u_t$ where $c$ is some unknown constant, which has to be estimated, or by introducing a second censored regression model to model explicitly negative values of $\rho_t$. We relegate these issues to our further research.

## 3.2 Related Time Varying Parameter Models

Clearly, the AR-CLEP model is a time series model with time varying parameters. In the last two decades several other model specifications have been proposed to model parameter variation over time, and some of these models bear similarities with our model, although there are also some important differences.

The threshold model advocated in for example Tong (1983), has been quite successful in describing parameter variation. A simple threshold model is given by (1) in combination with

$$\rho_t = \begin{cases} \rho^* & \text{if } x_t'\beta \geq 0 \\ 0 & \text{if } x_t'\beta < 0 \end{cases} \tag{20}$$

However, identification and estimation (especially of $\beta$) of this threshold model is very complicated, as the value of $\rho_t$ depends on an unknown linear combination of variables exceeding a threshold value, see also Chen (1995). To overcome this, the indicator function in (20) may be replaced by a smooth function, see Granger and Teräsvirta (1993). This smoothed threshold model allows for a continuum of possible values for the AR parameter between $\rho$ and $\rho + \rho^*$ by considering

$$\rho_t = \rho^* F(x_t'\beta) \tag{21}$$

where $F(\cdot)$ is a continuous function which takes values in the region $[0, 1]$. For example, $F$ can be the logistic function

$$F(z) = \frac{1}{1 + \exp(z)}, \tag{22}$$

see also Teräsvirta and Anderson (1992).

These threshold models impose an upperbound on the value of the autoregressive parameter, which contrasts with our model. For example, if $\rho^* > 0$, the maximum value of the autoregressive parameter is $\rho + \rho^*$. An alternative model may now be the model proposed in Haggan and Ozaki (1981), who consider the exponential [EAR] model given by (1) and

$$\rho_t = \exp(x_t'\beta). \tag{23}$$

As $\rho_t$ is now always positive, this model assumes that the autoregressive parameter is always in excess of $\rho$. Only for a very small value of $x_t'\beta$, $\rho_t$ gets close to 0, and the autoregressive parameter approximates $\rho$.

For the above models, the parameter variation is explained by a linear combination of exogenous variables $x_t'\beta$. Recently, there have been several applications of time series models with time varying autoregressive parameters, where this variation is modeled by an extra random variable instead of by exogenous variables. As an example, one may define that $\rho_t$ itself follows an AR(1) model, that is,

$$\rho_t = \gamma \rho_{t-1} + u_t, \tag{24}$$

with $u_t \sim \text{NID}(0, \sigma_u^2)$, see Harvey (1981) and Grillenzoni (1993) among others. The autoregressive parameter in (1) then follows a stationary autoregressive process around the mean $\rho$ if $|\gamma| < 1$. A special case of this model is analyzed in Leybourne, McCabe and Mills (1996) and Leybourne, McCabe and Tremayne (1996). These authors impose $\rho$ to be 1 and discuss several special cases of (24), including $\gamma = 0$ and $\gamma = 1$. These so-called randomized unit root processes allow for time periods with explosive behavior in $y_t$, where $\rho + \rho_t$ is larger than 1, but also for periods with $\rho + \rho_t < 1$ corresponding to non-explosive or error correcting behavior. A related model is considered in Granger and Swanson (1997) who assume that $\exp(\rho + \rho_t)$ follows a first order stationary autoregressive process. Under the restriction that $\text{E}[\exp(\rho + \rho_t)] = 1$, they obtain a so-called stochastic unit root [STUR] process, which also allows for sometimes explosive and sometimes approximately stationary behavior. A common property of these models is that the time variation in the autoregressive parameters is only explained by an unobserved stochastic process, which in contrast to our model, where we include observed explanatory variables. Hence the forecasting ability of these models only depends on the possible correlation in the unobserved stochastic processes $u_t$ and/or $\rho_t$.

Finally, we mention the Markov switching model, popularized by Hamilton (1989). In these models, a binary random variable is used to describe the regimes. A simple Markov switching model is given by (1) and

$$\rho_t = \rho^* s_t, \tag{25}$$

where $s_t$ is either 0 or 1, and where $s_t$ follows an unobserved first order Markov process with transition probabilities

$$\begin{aligned} &\Pr[s_t = 0 | s_{t-1} = 0] = p_t, \quad \Pr[s_t = 1 | s_{t-1} = 0] = 1 - p_t, \\ &\Pr[s_t = 1 | s_{t-1} = 1] = q_t, \quad \Pr[s_t = 0 | s_{t-1} = 1] = 1 - q_t. \end{aligned} \tag{26}$$

Hamilton (1989) assumes that the transition probabilities are constant over time, that is $p_t = p$ and $q_t = q$, $\forall t$. Filardo (1994) and Diebold *et al.* (1994) relax this assumption and make the transition probabilities a function of explanatory variables

$$p_t = \frac{1}{1 + \exp(x_t' \beta_p)} \text{ and } q_t = \frac{1}{1 + \exp(x_t' \beta_q)}. \tag{27}$$

In contrast to this model, our AR-CLEP model does not only forecast the occurrence of positive values of $\rho_t$, which correspond with a recession, but also yields an estimate of its size, which provides an explicit expression of the magnitude of persistence of explosiveness in a recession. Hence, our model does not restrict the degree of extra persistence to a constant value, as in (25).

# 4 Application

To illustrate our model we apply it to seasonally adjusted monthly unemployment rate of the United States. Figure 1 shows a graph of the logarithm of this series for the sample period 1969.01–1997.12. It is clear from this graph that short (seemingly explosive) periods of rapidly increasing unemployment are followed by longer periods with a slow decline in unemployment, possibly to some natural level of unemployment.

To explain possible variation in persistence over the business cycle, we use monthly seasonally adjusted US industrial production ($i_t$), the log of the oil price in dollars deflated by seasonally adjusted US CPI ($o_t$), the log of the Dow Jones index ($d_t$) and the difference between the 10 year treasury with constant maturity and a 3-month treasury bill rate of the United States ($r_t$)[1]. The inclusion of the oil price is based on the results in Hamilton (1983), Tatum (1988) and Mork (1989), while the results in for instance Harvey (1988), Estrella and Hardouvelis (1991) and Estrella and Mishkin (1997) suggest that the term structure of interest is a good predictor for turning points.

[1]The data are available at the internet site of the Federal Reserve Bank of St. Louis except for the Dow Jones index, which is taken from Datastream.

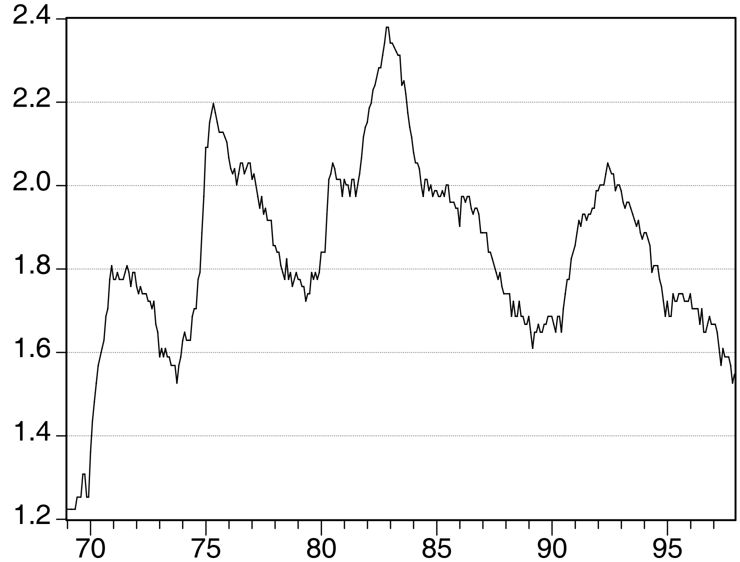


Figure 1: The logarithm of US unemployment rate, 1969.01–1997.12

## 4.1 AR-CLEP Model

We define $y_t$ as the log of the unemployment rate. The explanatory variables in the censored regression (2) are lagged values of $\Delta i_t$, $r_t$, $\Delta o_t$ and $\Delta d_t$, where first differences are taken to remove possible trends from the explanatory series. The estimation period is 1969.01–1997.12. Pre-sample observations are used as starting values.

First, we need to determine the order of the autoregression and the lag structure of the explanatory variables in the censored regression. A first order autoregression turns out to be appropriate to capture the dynamics in the $y_t$ series. The diagnostic tests below will confirm this. The specification of the lag structure of the explanatory variables is based on the following approach. First we add $r_t$, $\Delta d_t$, $\Delta o_t$ and $\Delta i_t$ one by one to the censored regression. In each step the optimal lag structure of the added variable is based on the maximum value of the likelihood over a range of possible lag structures. At the end of this procedure, we check whether a change in the lag structure of one of the explanatory variables gives a higher value of the likelihood. This results in the following model:

$$y_t = \underset{(0.012)}{0.021} + (\underset{(0.006)}{0.985} + \rho_t \mathbb{I}[\rho_t > 0])y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}\ (0, \underset{(0.001)}{0.023^2}) \tag{28}$$

with

$$\rho_t = \underset{(0.005)}{0.005} - \underset{(0.004)}{0.004}\ \Delta i_{t-2} - \underset{(0.003)}{0.009}\ r_{t-9} + \underset{(0.021)}{0.045}\ \Delta o_{t-10} - \underset{(0.051)}{0.144}\ \Delta d_{t-7} + u_t, \qquad u_t \sim \text{NID}\ (0, \underset{(0.004)}{0.016^2}), \tag{29}$$

where standard errors appear between parentheses, and where $\mathbb{I}[\cdot]$ is an indicator function which is 1 if the argument holds and zero otherwise.

To check for possible misspecification, we calculate the residuals defined in (13). Figure 2 shows a graph of these residuals. To check for serial correlation in these residuals we run regression (14) and test for the significance of the $\alpha$ parameter. The resulting $F$-statistic is 2.79, which is not significant at the 5% level. The $F-$statistic for ARCH effects based on auxiliary regression (15) equals 0.02 with $p$-value 0.87. The $\chi^2(2)$ normality test statistic equals 3.58 with $p$-value 0.17 such that normality cannot be rejected. Finally, to test for neglected dynamics in the censored regression (29), we add the first lag of the explanatory variables to the censored regression model, that is $\Delta i_{t-3}$, $r_{t-10}$, $\Delta o_{t-11}$ and $\Delta d_{t-8}$. The likelihood ratio statistic for zero restrictions on these four variables equals 8.54, which is not significant at the 5% level. Hence, there do not seem to be serious neglected dynamics in the censored regression and in the AR regression. In sum, our model in (28) and (29), which only contains 9 parameters, does not seem to be seriously misspecified.

Given its empirical adequacy, we may now interpret the parameters in (28) and (29). If $\rho_t = 0$, the autoregressive coefficient equals 0.985. The large AR(1) coefficient indicates a slow decay towards an equilibrium. This equilibrium may be interpreted as the natural unemployment rate. In our case this natural unemployment rate equals $\exp(0.021/(1-0.985) + 0.5 \times 0.023^2) = 3.969$. The persistence in the time series increases if $\rho_t$ exceeds zero. The time variation in persistence is explained by the explanatory variables in (29). Their coefficients have the expected sign. Negative growth

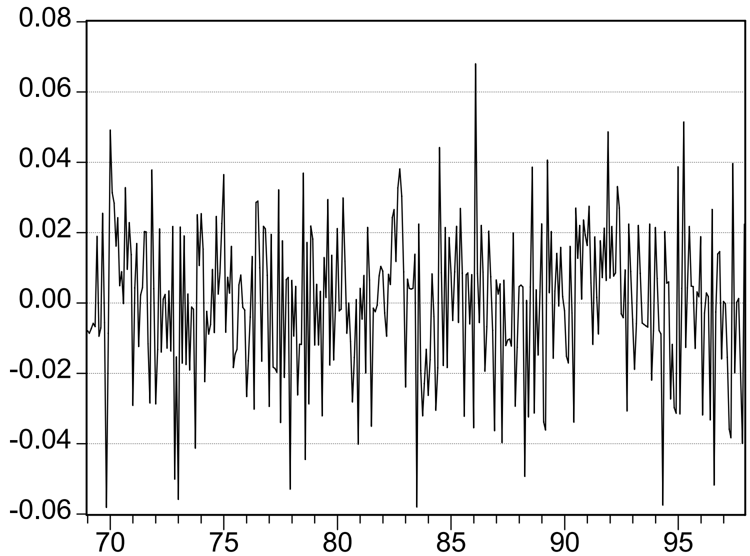


Figure 2: Residuals from the AR-CLEP model of the log of US unemployment rate

in industrial production, a negative difference between the long and short term interest rate and a negative Dow Jones return increase the probability of a positive value of $\rho_t$. The same applies for an increase in the oil price. The lag structure suggests that industrial production affects the persistence after 2 months, the term structure of interest after 9 months, the oil price after 10 months and the Dow Jones returns after 7 months. Furthermore, the variance of error term of the censored regression (29) appears to differ significantly from zero, and hence our model cannot be reduced to a threshold model with a nonstochastic threshold.

Figure 3 shows the estimated conditional probabilities $\Pr[\rho_t \neq 0|Y_t, x_t]$ defined in (11). Remember that large values of these probabilities indicate periods with higher persistence. The periods with $\Pr[\rho_t \neq 0|Y_t, x_t]$ are found at the beginning and middle of the 1970s and in the beginning of the 1980s and 1990s, and they appear to correspond with periods of increasing unemployment.

The conditional probabilities can be used to determine business cycle turning points, see Hamilton (1990) for a similar approach in Markov switching models. We may define

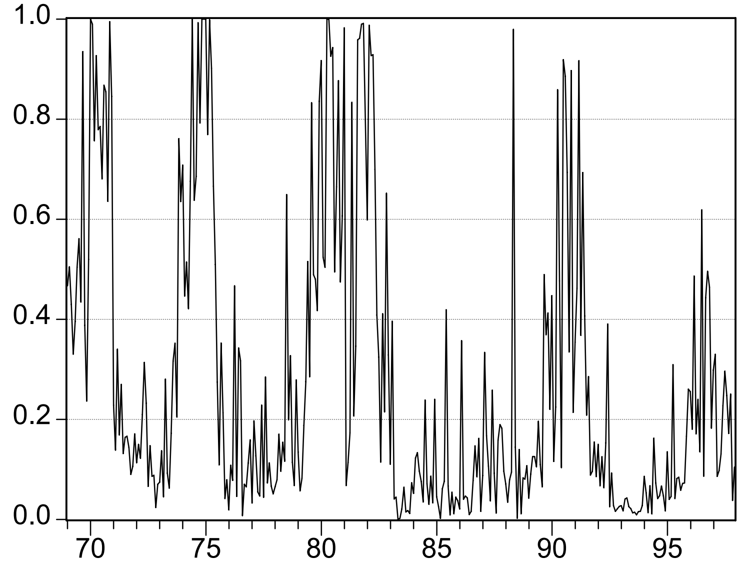


Figure 3: Conditional probabilities $\Pr[\rho_t \neq 0 | Y_t, x_t]$

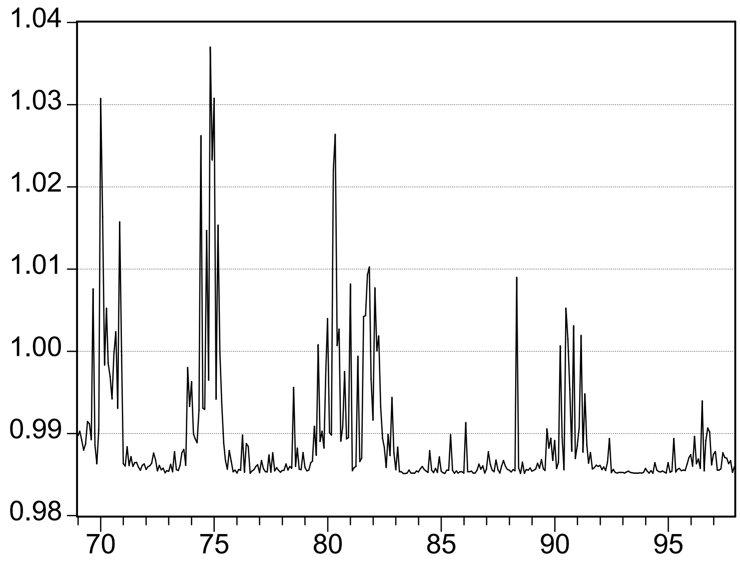


Figure 4: Conditional expectation $\rho + \mathrm{E}[\rho_t | Y_t, x_t]$

Table 1: Peaks and troughs for US unemployment based on non-smoothed and smoothed conditional probabilities.

| Unemployment[1] | | | | NBER | |
|---|---|---|---|---|---|
| non-smoothed | | smoothed | | | |
| peak | trough | peak | trough | peak | trough |
| 69.11 | 70.12 | 69.11 | 70.12 | 69.12 | 70.11 |
| 74.04 | 75.06 | 74.03 | 75.05 | 73.11 | 75.03 |
| 79.11 | 80.07 | 79.10 | 81.01 | 80.01 | 80.07 |
| 81.07 | 82.05 | 81.06 | 82.05 | 81.07 | 82.11 |
| – | – | 90.06 | 91.04 | 90.07 | 91.03 |

[1] A recession is defined by 6 consecutive months for which $\Pr[\rho_t \neq 0 \mid Y_t, x_t] > 0.5$. A peak corresponds with the last expansion observation before a recession and a trough with the last observation in a recession.

a recession as 6 consecutive months for which $\Pr[\rho_t \neq 0|Y_t, x_t] > 0.5$. A trough corresponds with the last observation in a recession and a peak with the last observation in an expansion. The first two columns of Table 1 display the peaks and troughs resulting from the conditional probabilities. The peaks and troughs correspond reasonably well with the official NBER peaks and troughs displayed in the last two columns of Table 1 except for the recession in the 1990s. For this last period, we cannot find 6 consecutive months for which $\Pr[\rho_t \neq 0|Y_t, x_t] > 0.5$, although according to Figure 3 a recession seems plausible. Hence, the rule we have defined to determine a recession may be too restrictive. Therefore, we also consider turning point analysis based on smoothed probabilities. The smoothed conditional probability that $\rho_t \neq 0$ at time $t$ is the average of the conditional probabilities at time $t-1$, $t$ and $t+1$. The third and fourth columns of Table 1 show the peaks and troughs based on these smoothed conditional probabilities. Now we also detect the recession in the beginning of the 1990s.

The persistence in the time series in each month follows from the estimated values of the conditional expectation of $\rho + \rho_t$ defined in (12). Figure 4 shows this expectation

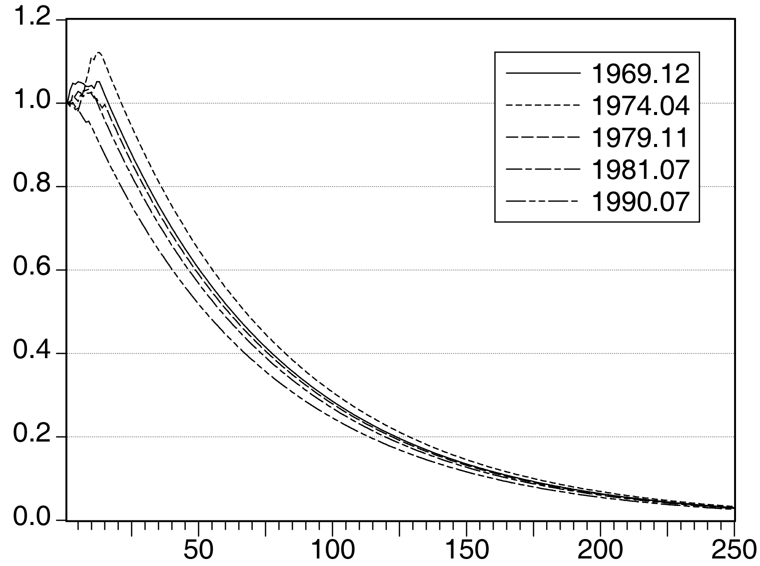


Figure 5: Effects of a unit shock on future values of the unemployment series

for each $t$. The value of the autoregressive parameter is close to 0.985 during expansions but during recessions its value increases and is sometimes larger than 1, indicating very strong persistence. Note that this explosive behavior is not so severe (with a maximum value is about 1.037) and is only temporary such that shocks eventually will die out. The magnitudes of the autoregressive parameter $(\rho + \rho_t)$ show that the recession in the beginning of the 1990s was less severe than the other recessions.

To show the influence of shocks during recessions on future values of the unemployment rate, we display the impact of a unit shock in the first month of each recession (based on the turning point results in Table 1) in Figure 5. Hence we impose a unit shock in 1969.12, 1974.04 1979.11, 1981.07 and 1990.07 and compute the impact of this shock over time based on the estimated conditional expectations $\rho + \mathrm{E}[\rho_t|Y_t, x_t]$ during the recessions. To make a direct comparison between the long-run impact of shocks during the recession, we set the AR parameter after the recession back to 0.985. The graph shows that a unit shock in 1974.04 has the largest impact on future values, while a unit shock in 1990.07 has the smallest impact. For example, the last unit shock becomes half its size after about 52 months, while that of 1974.04 takes 68 months for the same decay.

## 4.2 Forecasting Competition

To compare our AR-CLEP model with alternative time series models with time varying autoregressive parameters, we perform an out-of-sample forecast comparison. As alternative models we consider two linear and three nonlinear models. The linear alternatives are a simple AR(1) model and an AR(1) model with additional explanatory variables $y_{t-1}x_t$. This latter model we will denote as an ARXY(1). This ARXY(1) model assumes that the $x_t$ variables always exercise an effect, and that this effect can be positive and negative. Both models are estimated using ordinary least squares. For the $x_t$ variables we take the same variables and lag structure as for the AR-CLEP model in the previous subsection. To facilitate comparison, we follow the same procedure for the next three nonlinear models.

As the first nonlinear alternative model, we consider the exponential AR model of order 1 [EAR(1)] given in (1) with (23). The parameters in this model are estimated using nonlinear least squares [NLS]. For the full sample, this results in

$$y_t = \underset{(0.012)}{0.023} + (\underset{(0.008)}{0.976} + \rho_t)y_{t-1} + \varepsilon_t, \tag{30}$$

with

$$\rho_t = \exp(-\underset{(0.462)}{4.102} - \underset{(0.106)}{0.093}\Delta i_{t-2} - \underset{(0.124)}{0.247}\, r_{t-9} + \underset{(0.482)}{1.414}\Delta o_{t-10} - \underset{(1.916)}{3.928}\Delta d_{t-7}) \tag{31}$$

with estimated standard errors in between parentheses.

The second alternative model is the smoothed threshold model given by (1) with (21). The NLS estimates of this model are

$$y_t = \underset{(0.013)}{0.025} + (\underset{(0.007)}{0.983} + \rho_t)y_{t-1} + \varepsilon_t, \tag{32}$$

with

$$\rho_t = \underset{(0.003)}{0.020}\left[1+\exp\left(-\underset{(0.849)}{0.835}+\underset{(1.051)}{1.637}\,\Delta i_{t-2}+\underset{(1.372)}{2.800}\,r_{t-9}-\right.\right.$$
$$\left.\left.\underset{(9.403)}{12.076}\,\Delta o_{t-10}+\underset{(13.908)}{23.483}\,\Delta d_{t-7}\right)\right]^{-1} \quad (33)$$

with again the estimated standard errors between parentheses. For further reference, this model is denoted as LSTR(1).

Finally, we consider the Markov switching model with time varying parameters given in (1) with (25)–(27). We denote this model as MSTVP(1). The model is estimated using the EM algorithm of Dempster *et al.* (1977), see Hamilton (1990) and Diebold *et al.* (1994). Its parameter estimates are

$$y_t = \underset{(0.011)}{0.046} + (\underset{(0.006)}{0.972} + \rho_t)y_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim \text{NID}(0, \underset{(0.049\times 10^{-3})}{0.572\times 10^{-3}}) \quad (34)$$

with

$$\rho_t = \underset{(0.002)}{0.022}\ s_t \quad (35)$$

where

$$p_t = \left[1+\exp\left(-\underset{(1.300)}{3.866}-\underset{(0.936)}{2.471}\,\Delta i_{t-2}-\underset{(0.442)}{1.159}\,r_{t-9}+\right.\right.$$
$$\left.\left.\underset{(11.076)}{12.244}\,\Delta o_{t-10}-\underset{(24.719)}{49.469}\,\Delta d_{t-7}\right)\right]^{-1} \quad (36)$$

and

$$q_t = \left[1+\exp\left(-\underset{(0.747)}{1.901}-\underset{(1.259)}{1.679}\,\Delta i_{t-2}+\underset{(0.554)}{0.123}\,r_{t-9}-\right.\right.$$
$$\left.\left.\underset{(21.406)}{26.318}\,\Delta o_{t-10}-\underset{(15.454)}{5.555}\,\Delta d_{t-7}\right)\right]^{-1}. \quad (37)$$

The estimated standard errors appear in between parentheses. Some parameters in the above shown models may be set equal to zero, but for comparability reasons this is not pursued here.

To evaluate the out-of-sample forecast performance of the models we hold out the last 12, 48 and 96 months and re-estimate the parameters of the AR-CLEP and the above five models. We generate 12, 48 and 96 one-step ahead forecasts and compare the forecasted values with the true values. The one-step ahead forecasts for the AR-CLEP model are generated using (16) and hence we condition on the exogenous variables. For the other models, 1-step ahead forecasts can be generated in a straightforward way, except for the Markov switching model. The 1-step ahead forecast at time $T$ using this model is given by

$$\mathrm{E}[y_{T+1}|Y_T, x_t] = \mu + \lambda_{T+1}\rho y_T + (1 - \lambda_{T+1})(\rho + \rho^*)y_T \tag{38}$$

where $\lambda_{T+1} = (1 - s_T)p_{T+1} + s_T(1 - q_{T+1})$, that is the probability that $s_{T+1} = 0$. Since the variable $s_T$ is not observed it has to be replaced by an estimate that $s_T = 0$. We use the filter provided in Hamilton (1989) to compute smoothed conditional probabilities $\Pr[s_T = 0|Y_T, x_T]$.

Table 2 shows the results of the forecast comparison. The first two columns show the root mean squared forecast error [RMSE] and the mean absolute percentage forecast error [MAPE]. We observe that the AR(1)-CLEP model performs best or second best, where close competitors are the LSTR(1) and the MSTVP(1) models. The next two columns give the results of the nonparametric binomial sign test. The first column displays the fraction that the forecast errors of the AR-CLEP are smaller in absolute value than the forecast errors of the alternative models. The second column shows the $p$-value of the one-sided test of equal forecast accuracy against the alternative hypothesis that the AR-CLEP model is better. It can be seen that oftentimes the differences in the forecast errors are not significant at the 10% level, although there are some exceptions, especially concerning the linear models. Finally, in the last two columns we give the results for forecast encompassing tests. These encompassing tests check whether the forecasts of the competing models have explanatory power for the forecast errors of the AR-CLEP(1) model, see Clements and Hendry (1993) for details. Clearly, for the largest forecasting sample, which includes the most recent recession, the AR(1)-CLEP model encompasses rival models but often not the other way around.

Table 2: Forecast performance comparison.

| model | criteria RMSE ×100 | MAPE | sign test[1] fraction | $p$-value | encompassing tests[2] I | II |
|---|---|---|---|---|---|---|
| | | | forecasting sample 1997.01–1997.12 | | | |
| AR(1)-CLEP | 2.68 | 1.35 | | | | |
| AR(1) | 2.90 | 1.51 | 10/12 | (0.00) | 0.89 (0.37) | 0.28 (0.61) |
| ARXY(1) | 2.90 | 1.56 | 11/12 | (0.00) | 0.38 (0.55) | 0.78 (0.40) |
| EAR(1) | 2.83 | 1.49 | 10/12 | (0.00) | 0.36 (0.56) | 0.48 (0.50) |
| LSTR(1) | 2.70 | 1.36 | 6/12 | (0.39) | 0.12 (0.73) | 0.01 (0.91) |
| MSTVP(1) | 2.71 | 1.36 | 6/12 | (0.39) | 0.91 (0.36) | 0.36 (0.56) |
| | | | forecasting sample 1994.01–1997.12 | | | |
| AR(1)-CLEP | 2.56 | 1.18 | | | | |
| AR(1) | 2.77 | 1.29 | 30/48 | (0.03) | 0.86 (0.36) | 0.00 (0.99) |
| ARXY(1) | 2.66 | 1.24 | 28/48 | (0.10) | 0.10 (0.75) | 1.23 (0.27) |
| EAR(1) | 2.61 | 1.22 | 26/48 | (0.24) | 0.16 (0.68) | 0.01 (0.92) |
| LSTR(1) | 2.61 | 1.21 | 24/48 | (0.44) | 0.26 (0.61) | 1.08 (0.30) |
| MSTVP(1) | 2.56 | 1.15 | 20/48 | (0.85) | 0.87 (0.35) | 0.03 (0.86) |
| | | | forecasting sample 1990.01–1997.12 | | | |
| AR(1)-CLEP | 2.37 | 1.06 | | | | |
| AR(1) | 2.49 | 1.11 | 52/96 | (0.18) | 0.81 (0.37) | 8.64 (0.00) |
| ARXY(1) | 2.44 | 1.11 | 51/96 | (0.24) | 0.23 (0.63) | 6.75 (0.01) |
| EAR(1) | 2.37 | 1.07 | 46/96 | (0.62) | 0.27 (0.61) | 1.76 (0.18) |
| LSTR(1) | 2.36 | 1.05 | 51/96 | (0.24) | 0.24 (0.62) | 0.21 (0.65) |
| MSTVP(1) | 2.47 | 1.07 | 53/96 | (0.13) | 2.07 (0.15) | 9.99 (0.00) |

[1] A nonparametric sign test. The first column displays the fraction $k$ that the forecast errors of the AR-CLEP model are smaller in absolute value than the forecast errors of the alternative models. The second column displays the $p$-value for the corresponding one-sided nonparametric test for equal forecast accuracy against the alternative that the AR-CLEP model is better. The $p$-value equals $\sum_{i=nk+1}^{n} p^i(1-p)^{n-i}$, where $n$ is the number of forecasts, $p=\frac{1}{2}$ and $k$ is the fraction.

[2] In column I we report encompassing tests (with $p$-values between parentheses) to test whether forecasts generated by the AR-CLEP model encompass forecasts generated by the alternative model, while in column II we investigate whether the alternative model forecasts encompass the AR-CLEP model forecasts.

## 5 Conclusion

In this paper we proposed a parsimonious time series model, which allows for a higher degree of persistence during recessions, where these recessionary periods are explained by a linear combination of lagged explanatory variables. The application to US unemployment showed that the model yields plausible inference and outperforms alternative models in terms of forecasting.

An interesting topic for further research is to see if our model can be extended to a multivariate setting. If so, it is worthwhile to see if the same or other linear combinations of lagged variables predict recession periods for all variables.

# Appendix

To derive the first derivative of the log likelihood function (9) we consider first the partial derivatives of the unconditional pdf of $y_t$ with respect to the model parameters $\theta = \{\mu, \rho, \sigma_\varepsilon, \beta, \sigma_u\}$

$$\frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \mu} = \frac{1}{f(y_t|Y_{t-1}, x_t; \theta)} \left[ \Phi\left(\frac{-x_t'\beta}{\sigma_u}\right) \frac{e_t}{\sigma_\varepsilon^2} f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \int_{-x_t'\beta}^{\infty} \frac{1}{\sigma_u}\phi\left(\frac{u_t}{\sigma_u}\right) \frac{e_t - (x_t'\beta + u_t)y_{t-1}}{\sigma_\varepsilon^2} f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t = x_t\beta + u_t} du_t \right],$$

$$\frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \rho} = \frac{1}{f(y_t|Y_{t-1}, x_t; \theta)} \left[ \Phi\left(\frac{-x_t'\beta}{\sigma_u}\right) \frac{e_t}{\sigma_\varepsilon^2} f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \int_{-x_t'\beta}^{\infty} \frac{1}{\sigma_u}\phi\left(\frac{u_t}{\sigma_u}\right) \frac{e_t - (x_t'\beta + u_t)y_{t-1}}{\sigma_\varepsilon^2} f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t = x_t\beta + u_t} du_t \right] y_{t-1},$$

$$\frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \sigma_\varepsilon} = \frac{1}{f(y_t|Y_{t-1}, x_t; \theta)} \left[ \Phi\left(\frac{-x_t'\beta}{\sigma_u}\right) \left(\frac{e_t^2}{\sigma_\varepsilon^3} - \frac{1}{\sigma_\varepsilon}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \int_{-x_t'\beta}^{\infty} \frac{1}{\sigma_u}\phi\left(\frac{u_t}{\sigma_u}\right) \left(\frac{(e_t - (x_t'\beta + u_t)y_{t-1})^2}{\sigma_\varepsilon^3} - \frac{1}{\sigma_\varepsilon}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t = x_t\beta + u_t} du_t \right],$$

$$\frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \beta} = \frac{1}{f(y_t|Y_{t-1}, x_t; \theta)} \left[ \frac{-1}{\sigma_u}\phi\left(\frac{-x_t'\beta}{\sigma_u}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \frac{1}{\sigma_u}\phi\left(\frac{-x_t'\beta}{\sigma_u}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \int_{-x_t'\beta}^{\infty} \frac{y_{t-1}}{\sigma_u}\phi\left(\frac{u_t}{\sigma_u}\right) \frac{(e_t - (x_t'\beta + u_t)y_{t-1})}{\sigma_\varepsilon^2} f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t = x_t'\beta + u_t} du_t \right] x_t$$

and

$$\frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \sigma_u} = \frac{1}{f(y_t|Y_{t-1}, x_t; \theta)} \left[ \frac{-1}{\sigma_u^2}\phi\left(\frac{-x_t'\beta}{\sigma_u}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t=0} + \int_{-x_t'\beta}^{\infty} \left(\frac{u_t^2}{\sigma_u^4} - \frac{1}{\sigma_u^2}\right) \phi\left(\frac{u_t}{\sigma_u}\right) f(y_t|Y_{t-1}, \rho_t; \theta)|_{\rho_t = x_t\beta + u_t} du_t \right],$$

where $f(y_t|Y_{t-1}, x_t; \theta)$ and $f(y_t|Y_{t-1}, \rho_t; \theta)$ are defined in (8) and (7) respectively and where $e_t = y_t - \mu - \rho y_{t-1}$. The total derivative $g_t(y_t|Y_{t-1}, x_t; \theta) = \partial \ln(f(y_t|Y_{t-1}, x_t; \theta))/\partial\theta$ is a vector of stacked partial derivatives

$$g_t(y_t|Y_{t-1}, x_t; \theta) = \begin{pmatrix} \frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \mu} \\ \frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \rho} \\ \frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \sigma_\varepsilon} \\ \frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \beta} \\ \frac{\partial \ln(f(y_t|Y_{t-1}, x_t; \theta))}{\partial \sigma_u} \end{pmatrix}.$$

Hence the first order derivative of the log likelihood function (9) equals

$$\frac{\partial \mathcal{L}(Y_T|x_T; \theta)}{\partial \theta} = \sum_{t=1}^{T} g_t(y_t|Y_{t-1}, x_t; \theta).$$

The maximum likelihood estimator of $\theta$ denoted by $\hat{\theta}$ is the solution of the first order condition

$$\frac{\partial \mathcal{L}(Y_T|x_T; \theta)}{\partial \theta} = 0.$$

This estimator $\hat{\theta}$ is asymptotically normally distributed with mean $\theta$ and the covariance matrix is given by the inverse of the information matrix $I$. This matrix can be estimated by evaluating minus the second order derivative of the log likelihood function in $\theta = \hat{\theta}$ or by the sum of the outer products of the gradients $g_t$ evaluated in $\hat{\theta}$

$$\hat{I} = \sum_{t=1}^{T} g_t(y_t|Y_{t-1}, x_t; \hat{\theta})' g_t(y_t|Y_{t-1}, x_t; \hat{\theta}),$$

see Hamilton (1996, p.132) for a similar approach in Markov switching time series models.

# References

Beaudry, P. and G. Koop, 1993, Do Recessions Permanently Change Output?, *Journal of Monetary Economics* **31**, 149–163.

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. Summers, 1987, Hysteresis in Unemployment, *European Economic Review* **31**, 288–295.

Chen, R., 1995, Threshold Variable Selection in Open-Loop Threshold Autoregressive Models, *Journal of Time Series Analysis* **16**, 461–482.

Clements, M.P. and D. Hendry, 1993, On the Limitations of Comparing Mean Squared Forecast Errors, *Journal of Forecasting* **12**, 617–637.

Dempster, A.P., N.M. Laird and D.B. Rubin, 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society B* **39**, 1–38.

Diebold, F.X., J. Lee and G.C. Weinbach, 1994, Regime Switching and Endogenous Transition Probablities, in C.P. Hargreaves (editor), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, Oxford, chapter 10, pp. 283–302.

Elwood, S.K., 1998, Is the Persistence of Shocks to Output Asymmetric?, *Journal of Monetary Economics* **41**, 411–426.

Estrella, A. and F.S. Mishkin, 1997, The Predictive Power of the Term Structure of Interest Rates in Europe and the United Staes: Implications for the European Central Bank, *European Economic Review* **41**, 1375–1401.

Estrella, A. and G.A. Hardouvelis, 1991, The Term Structure as a Predictor of Real Economic Activity, *Journal of Finance* **46**, 555–576.

Filardo, A.J., 1994, Business-Cycle Phases and Their Transitional Dynamics, *Journal of Business & Economic Statistics* **12**, 299–308.

Granger, C.W.J. and N.R. Swanson, 1997, An Introduction to Stochastic Unit-Root Processes, *Journal of Econometrics* **80**, 35–62.

Granger, C.W.J. and T. Teräsvirta, 1993, *Modelling Nonlinear Economic Relations*, Oxford University Press, Oxford.

Grillenzoni, C., 1993, ARIMA Processes With ARIMA Parameters, *Journal of Business & Economic Statistics* **11**, 235–250.

Haggan, V. and T. Ozaki, 1981, Modelling Non-linear Random Vibrations Using an Amplitude-dependent Autoregressive Time Series Model, *Biometrika* **68**, 189–196.

Hamilton, J.D., 1983, Oil and the Macroeconomy Since World War II, *Journal of Political Economy* **91**, 228–248.

Hamilton, J.D., 1989, A New Approach to the Econometric Analysis of Nonstationary Time Series and Business Cycles, *Econometrica* **57**, 357–384.

Hamilton, J.D., 1990, Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics* **45**, 39–70.

Hamilton, J.D., 1996, Specification Testing in Markov-Switching Time-Series Models, *Journal of Econometrics* **70**, 127–158.

Harvey, A.C., 1981, *The Econometric Analysis of Time Series*, Philip Alan, Oxford.

Harvey, C.R., 1988, The Real Term Structure and Consumption Growth, *Journal of Financial Economics* **22**, 305–333.

Johnson, N.L. and S. Kotz, 1970, *Distributions in Statistics: Continuous Univariate Distributions*, Houghton Mifflin, Boston.

Leybourne, S.J., B.P.M. McCabe and A.R. Tremayne, 1996, Can Economic Time Series Be Differenced to Stationarity?, *Journal of Business & Economic Statistics* **14**, 435–446.

Leybourne, S.J., B.P.M. McCabe and T.C. Mills, 1996, Randomized Unit Processes for Modelling and Forecasting Financial Time Series: Theory and Applications, *Journal of Forecasting* **15**, 253–270.

Mork, K.A., 1989, Oil and the Macroeconomy When Prices Go Up and Down: An Extension of Hamilton's Results, *Journal of Political Economy* **97**, 740–744.

Neftçi, S.N., 1984, Are Economic Time Series Asymmetric over the Business Cycle, *Journal of Political Economy* **92**, 307–328.

Tatum, J.A., 1988, Are the Macroeconomic Effects of Oil Pice Changes Symmetric?, *Carnegie-Rochester Conference Series on Public Policy* **28**, 325–368.

Teräsvirta, T. and H.M. Anderson, 1992, Characterizing Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models, *Journal of Applied Econometrics* **7**, S119–S136.

Tong, H., 1983, *Threshold Models in Non-Linear Time Series Analysis*, Springer-Verlag, Berlin.