Real-time detection of the business cycle using SETAR models
A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2004

ISBN 92-894-7386-X
ISSN 1725-4825

© European Communities, 2004
4TH EUROSTAT AND DG ECFIN
COLLOQUIUM ON MODERN TOOLS FOR BUSINESS CYCLE ANALYSIS

"GROWTH AND CYCLE IN THE EURO-ZONE"

20 TO 22 OCTOBER 2003
Luxembourg, European Parliament
Hémicycle, Schuman building

Real-time detection of the business cycle using SETAR models

Laurent Ferrara, Centre d’Observation Economique, Paris, France
Dominique Guéган, ENS-Cachan, IDHE-MORA, Cachan, France
Real-Time Detection of the Business Cycle using SETAR Models

L. Ferrara * and D. Guégan †

October 7, 2003

1 Introduction

Recently, we witnessed the development of new modern tools in business cycle analysis, mainly based on non-linear parametric modeling. Non-linear models have the great advantage to be flexible enough to take into account certain stylized facts of the economic business cycle, such as asymmetries in the phases. In this respect, much of attention has concentrated on the class of non-linear dynamic models that accommodate the possibility of regime changes.

Especially, Markov-Switching models popularized by Hamilton (1989) have been extensively used in business cycle analysis in order to describe the economic fluctuations. Among the huge amount of empirical studies, we can quote the papers of Sichel (1994), Lahiri and Wang (1994), Potter (1995), Anas and Ferrara (2002a), Chauvet and Piger (2003), Clements and Krolzig (2003) or Ferrara (2003) as regards the US economy and the papers of Krolzig (2001, 2003) or Krolzig and Toro (2001) as regards the Euro-zone economy. Generally, the output of these applications is twofold. The authors aim either at dating the turning points of the cycle or at detecting in real-time the current regime of the economy.

However, a clear distinction must be done between dating and detecting the turning points of the cycle. Dating is an ex post exercise for which several parametric and non-parametric methodologies are available. It turns out that simple non-parametric procedures, such as the famous Bry and Boschan (1971) procedure still used by the Dating Committee of the NBER,

*Centre d’Observation Economique, 27 Avenue de Friedland, 75382 Paris cedex 08, e-mail: lferrara@ccip.fr
†E.N.S. Cachan, Département d’Economie et Gestion, MORAC-NRS, CNRS UMR 8833, Senior Academic Fellow de l’Institut Europlace de finance (IEF), 61 Avenue du President Wilson, 94235, Cachan Cedex, France, e-mail : guégan@ecogest.ens-cachan.fr
are more convenient for this kind of work (see Harding and Pagan, 2001, or Anas and Ferrara, 2002b, for a discussion on this issue). Real-time detection refers mainly to short-term economic analysis, which is not an easy task for practitioners. Indeed, several economic indicators are released on a regularly monthly basis, or even on a daily basis as regards the financial sector, adding volatility to the existing volatility and thus leading to an inflation of the available information set. Moreover, the data are often strongly revised and the diverse statistical methods, such as seasonal adjustment or filtering techniques, lead to edge effects. In this framework, the statistician has a crucial role to play which consists in extracting the right signal to help the short-term economic diagnosis. The too often quoted word "data miner" seems to be here well appropriate. Therefore, the real-time economic analysis asks for methods with strong statistical content.

In this respect, Markov-switching models have shown their interest in real-time business cycle analysis. Besides this well known approach, other parametric models have been proposed in the statistical literature to allow for different regimes. For instance, probit and logit models have been used by Estrella and Mishkin (1998) to predict US recessions. The threshold autoregressive (TAR) model, proposed first by Lim and Tong (1980), is able to produce limit cycle, time irreversibility and asymmetry behavior of a time series. TAR models have been used to describe the asymmetry observed in the quarterly US real GNP by different authors, such as Tiao and Tsay (1994), Potter (1995) and Proietti (1998) for instance, and using US unemployment monthly data by Hansen (1997). With the TAR model the transition variable is observed: it may be either an exogenous variable, such as a leading index for example, or a linear combination of lagged values of the series. In this latter case, the model is referred to as a self-exciting threshold autoregressive (SETAR) model. This is the main difference with the Markov-Switching model whose parameters of the autoregressive data generating process vary according to the states of the latent Markov chain. These two approaches are complementary because the notion captured under investigation is not exactly the same. Nevertheless, one of the interest of SETAR processes lies on their predictability, see for instance De Googier and De Bruin (1999) and Clements and Smith (1999, 2000). When dealing with SETAR models, the transition is discrete, but smooth transition is also chosen to study the business cycle by some authors. Then, we get the so-called STAR model, see for instance Terasvirta and Anderson (1992) and van Dijk, Terasvirta and Franses (2002).

In this paper, we focus on real-time detection of business cycle turning points. Our aim is rather to point out some thresholds under (over) which a signal of turning point could be given in real-time. We prefer the SETAR approach because a threshold model seems to be attractive in terms
of business cycle analysis. Here, we propose a prospective approach as an alternative to other approaches, including, for instance, the use of switching models to detect the business cycle. Thus, in the following, we assume that it is possible to adjust a SETAR on the considered data and we do not test this assumption. This way is based on the following intuition: the existence of two (or more) states inside the data and the possible distinction of these states using the structure of the data, without imposing the existence of another series to explain the split from one state to another one.

This paper is split into two parts. In a first step (Section two), we introduce the various threshold models and we discuss their statistical properties. Especially, we recall the classical techniques to estimate the number of regimes, the threshold, the delay and the parameters of the model as the forecasting method (Section three). In a second step, we apply these models to the Eurozone industrial production index to detect in real time the dates of peaks and troughs for the business cycle (Section four). By using a dynamic simulation approach, we also provide a measure of performance of our model by comparison to a benchmark dating chronology (Section four). Lastly, some conclusions and further research directions are proposed in Section six.

2 Description of models which capture states

In this section, we specify some of the models which permit to take into account the existence of various states inside real data. For sake of simplicity, we describe the models only with two regimes, but they can be easily generalised to more regimes.

1. The mean stationary process \((Y_t)_t\) follows a SETAR process if it verifies the following equation:

\[
Y_t = (\phi_{0,1} + \phi_{1,1} Y_{t-1})(1 - I_{[Y_{t-d}>c]}) + (\phi_{0,2} + \phi_{1,2} Y_{t-1})I_{[Y_{t-d}>c]} + \varepsilon_t. \tag{1}
\]

For a given threshold \(c\) and the position of the random variable \(Y_{t-d}\) with respect to this threshold \(c\), the process \((Y_t)_t\) follows here a particular AR(1) model: \(\phi_{0,1} + \phi_{1,1} Y_{t-1} + \varepsilon_t\) or : \(\phi_{0,2} + \phi_{1,2} Y_{t-1} + \varepsilon_t\). The model parameters are \(\phi_{i,j}\), for \(i = 0, 1\) and \(j = 1, 2\), the threshold \(c\) and the delay \(d\). This model has been introduced by Lim et Tong (1980), see also Tong (1990). On each state, it is possible to propose more complex stationary models like ARMA\((p,q)\) processes (Brockwell and Davis, 1988), bilinear models (Guégan, 1994) or GARCH\((p,q)\) processes (Bollerslev, 1986).
If we denote \( \pi \) the non conditional stationary distribution of the process \((Y_t)_t\), to get its analytical form is a non trivial problem. However an implicit solution is always available if \((Y_t)_t\) can be considered as an ergodic Markov chain over \( R^n \), which is given by:

\[
\pi(A) = \int_{-\infty}^{\infty} P(A|x) \pi(dx),
\]

where \( \pi \) denotes the limiting distribution of \((Y_t)_t\). For SETAR processes introduced in (1), different numerical solutions have been proposed to solve this problem, see Jones (1978) and Pemberton (1985). Practically we will obtained an approximation of this distribution, computing empirically the percentage of points belonging to the first regime or to the second one. This will give an estimation of the non conditional probability (\( \pi_1 \) or \( \pi_2 \)) to be in a given regime. The theoretical approach is still an opened problem.

2. We can use a smooth transition variable to characterize the states of the model and we get thus the smooth transition autoregressive (STAR) process. In that case the process \((Y_t)_t\) follows the recursive scheme:

\[
Y_t = \left( \phi_{0,1} + \phi_{1,1} Y_{t-1} \right) \left( 1 - G(Y_{t-d}, \gamma, c) \right) + \left( \phi_{0,2} + \phi_{1,2} Y_{t-1} \right) G(Y_{t-d}, \gamma, c) + \varepsilon_t,
\]

where \( G \) is some continuous function, for instance the logistic one:

\[
G(Y_{t-d}, \gamma, c) = \frac{1}{1 + \exp\left(-\gamma (Y_{t-d} - c)\right)}.
\]

Note that the transition function \( G \) is bounded between 0 and 1. The parameter \( c \) can be interpreted as the threshold between the two regimes in the sense that the logistic function changes monotonically from 0 to 1 with respect to the value of the lagged endogenous variable \( Y_{t-d} \). The parameter \( \gamma \) determines the smoothness of the change in the value of the logistic function, and thus, the smoothness of the transition of one regime to the other. As \( \gamma \) becomes very large, the logistic function (5) approaches the indicator function \( I_{Y_{t-d} \leq c} \), defined as \( I_A = 1 \) if \( A \) is true and \( I_A = 0 \) otherwise. Consequently, the change of \( G(Y_{t-d}, \gamma, c) \) from 0 to 1 becomes instantaneous at \( Y_{t-d} = c \). Then we find the SETAR model as a particular case of this STAR model. When \( \gamma \to 0 \), the logistic function approaches a constant (equal to 0.5) and when \( \gamma = 0 \), the STAR model reduces to a linear AR model. This model has been described by Terasvirta and Anderson (1992), see also Van Dijk, Franses and Paap (2002).
3. Stationary SETAR models with changes in the variance may also be considered. Then, the process \((Y_t)_t\) is such that:

\[
Y_t = \begin{cases} 
\phi_{0,1} + \phi_{1,1} Y_{t-1} + \sigma_{1,1} \varepsilon_t & \text{if } Y_{t-1} < c_1 \text{ and if } \sigma_{1,i} < c_2,; i = 1,2 \\
\phi_{0,2} + \phi_{1,2} Y_{t-1} + \sigma_{1,2} \varepsilon_t & \text{if } Y_{t-1} \geq c_1 \text{ and if } \sigma_{1,i} \geq c_2,; i = 1,2.
\end{cases}
\]

(6)

Here, the threshold \(c_1\) characterizes the level of the process and the threshold \(c_2\) the noise variance, see Pfann, Schotman and Tchernig (1996).

4. A SETAR process \((Y_t)_t\) with long memory dynamics can be defined as follows:

\[
\begin{cases} 
(1 - B)^d Y_t = \varepsilon_t^{(1)}, & \text{if } Y_{t-l} \leq c : \text{ regime 1} \\
Y_t = \varepsilon_t^{(2)}, & \text{if } Y_{t-l} > c : \text{ regime 2.}
\end{cases}
\]

This model can be complexified in different ways. Here we assume that we have a long memory behavior on only one state and a particular short memory behavior on the other state. We can, for instance, consider two different long memory behaviors on each state, then we get the following representation:

\[
\begin{cases} 
(1 - B)^d Y_t = \varepsilon_t^{(1)}, & \text{if } Y_{t-l} \leq c : \text{ regime 1} \\
(1 - B)^d Y_t = \varepsilon_t^{(2)}, & \text{if } Y_{t-l} > c : \text{ regime 2,}
\end{cases}
\]

(8)

This model has been introduced and discussed by Dufrénot, Guégan and Peignu-Féissolle (2003).

5. A generalization of the STAR model with long memory dynamics can also be considered, then the process \((Y_t)_t\) defined in (9) is extended in the following way:

\[
(I - B)^d Y_t = (\phi_{0,1} + \phi_{1,1} Y_{t-1})(1 - G(Y_{t-\delta}, \gamma, c)) + (\phi_{0,2} + \phi_{1,2} Y_{t-1})G(Y_{t-\delta}, \gamma, c) + \varepsilon_t,
\]

(9)

where \(G\) is for instance the logistic function introduced in (5). This model has been introduced by van Dijk, Franses and Paap (2002).

6. The mean stationary switching model has been introduced, first, by Quandt (1958), then reconsidered by Neftci (1982, 1984) and popularized later in economics by Hamilton in 1988. It is defined by the following equations:

\[
Y_t = \phi_{0,s_t} + \phi_{1,s_t} Y_{t-1} + \varepsilon_t,
\]

(11)
where the non-observed process \((s_t)_t\) is an ergodic Markov chain and 
\((\varepsilon_t)_t\) is a classical noise. The associated probability transition to the
process \((s_t)_t\) is defined by:

\[
P[s_t = j | s_{t-1} = i] = p_{ij},
\]

with \(0 < p_{ij} < 1 \in N\) and \(i, j = 1, 2\). We can assume that the process
is characterized by two states, then in that latter case, if \(s_t = 1\), the process \((Y_t)_t\)
follows the regime \(\phi_{0,1} + \phi_{1,1}Y_{t-1} + \varepsilon_t\) and if the variable \(s_t = 2\), the process \((Y_t)_t\) follows the regime \(\phi_{0,2} + \phi_{1,2}Y_{t-1} + \varepsilon_t\).

In the two-regime case, it is possible to compute the non-conditional probabilities associated to the process \((Y_t)_t\). They are equal to:

\[
P[s_t = 1] = \frac{1 - p_{22}}{1 - p_{11} + 1 - p_{22}} = \pi,
\]

and

\[
P[s_t = 2] = 1 - \pi.
\]

7. We can generalize these processes, imposing the states on the vari-
ances, then we get the AR-SWGARCH processes introduced by Hamilton
and Susmel (1994). These processes are defined by the following
equations (using the same notations as above):

\[
Y_t = a_0 + \phi Y_{t-1} + u_t
\]

(13)

\[
u_t = \sigma_t \varepsilon_t
\]

(14)

\[
\sigma_t^2 = a_{0,ss} + a_{1,ss-1} u_{t-1}^2 + \delta_{ss-1} \sigma_{t-1}^2.
\]

(15)

Here, \((\nu_t)_t\) is a white noise process. Other extensions permit to intro-
duce different scale parameters, for instance with the following represen-
tation introduced by Krolzig and Toro (2000). The process \((Y_t)_t\)
follows an AR-SWGARCH process with level parameter if it follows
the recursive scheme:

\[
Y_t = a_0 + \phi Y_{t-1} + u_t
\]

(16)

\[
u_t = \sqrt{g_{ss}} \varepsilon_t
\]

(17)

\[
\varepsilon_t = \sigma_t \nu_t
\]

(18)

\[
\sigma_t^2 = a_{0,ss} + a_{1,ss-1} \varepsilon_{t-1}^2 + \xi \varepsilon_{t-1}^2 + \delta_{ss-1} \sigma_{t-1}^2.
\]

(19)

Here, \((\nu_t)_t\) is a white noise process and \(g_{ss}\) is a scale function allowing
the parameters to move from one state to the other. The variable \(d_t\)
is such that:

\[
d_{t-1} = 1 \quad \text{if} \quad \varepsilon_{t-1} \leq 0
\]

(20)

\[
d_{t-1} = 0 \quad \text{if} \quad \varepsilon_{t-1} > 0,
\]

(21)
and $\xi$ is the level parameter. The error process $(u_t)_t$ changes with the
regime in which the process $(Y_t)_t$ is. The noise's variance is given by:

$$E[u_t^2|s_t, s_{t-1}, ..., u_{t-1}, ...] = g_{s_t}[a_0 + a_1 \frac{\varepsilon^2_{t-1}}{g_{s_{t-1}}} + \xi d_{t-1} \frac{\varepsilon^2_{t-1}}{g_{s_{t-1}}} + \delta_{s_{t-1}} \frac{\sigma^2_{t-1}}{g_{s_{t-1}}}].$$

The five first models belong to the class of SETAR models: the first three
assume the presence of two or more regimes within which the time series
requires different short memory behavior for description. The fourth and
the fifth models permit to capture both features of long memory and nonlinearity,
combining the concepts of fractional integration and transition nonlinearity (discrete or smooth). The last two ones belong to another class of
models called the switching models. We do not consider the use of these
later models in this paper. For a review on all these models, we refer to

3 Inference for SETAR models

The TAR model introduced in the eighties' has not been widely used in applications until recently, primarily because it was hard in practice to identify
the threshold variable and to estimate the associated values and secondly
because there was no simple modeling procedure available. Recently some
authors have proposed different ways to bypass this problem. First, we
present in this section a classical way to estimate the parameters of the
SETAR models and we specify some recent literature which permits to implem
quickly the procedure described here. Then, in a second step, we recall how we can forecast with these models in a Gaussian context.

3.1 Estimation theory

Here, we assume first that the model available for our purpose is a SETAR-
$(2, p_1, p_2)$ model with two regimes and we assume that it is possible to adjust
an $AR(p_i, i = 1, 2)$ process on each regime. We do not test this assumption
and we refer to Tsay (1989) for a test procedure, see other references below.
The autoregressive lag $p_1$ in the first regime may also be different from the
lag $p_2$ in the other regime.

As noted above, a major difficulty in applying TAR models is the specifi-
cation of the threshold variable, which plays a key role in the non-linear
structure of the model. Since there is only a finite number of choices for the
parameters $c$ and $d$, the best choice can be done using the Akaike Informa-
tion Criterion (AIC), see Akaike (1974). This procedure has been proposed
by Tong and Lim (1980) and is used by a large part of the practitioners.
A SETAR \((2,p_1,p_2)\) model can be written in the following form:

\[
Y_t = (1 - I(Y_{t-d} > c))((\phi_{0,0} + \sum_{i=1}^{p_1} \phi_{0,i} Y_{t-i} + \sigma_0 \varepsilon_t)) + I(Y_{t-d} > c)((\phi_{1,0} + \sum_{i=1}^{p_2} \phi_{1,i} Y_{t-i} + \sigma_1 \varepsilon_t)),
\]

where \(I(Y_{t-d} > c) = 1\) if \(Y_{t-d} > c\) and zero otherwise. Now, using some algebraic notations, the model (22) can be rewritten as a regression model.

Denote \(I_d(c) = I(Y_{t-d} > c)\), \(\Phi_0 = [\phi_{0,0}, \cdots, \phi_{0,p_1}]^T\), \(\Phi_1 = [\phi_{1,0}, \cdots, \phi_{1,p_2}]^T\) and \(\mathbf{Y}_{t-1} = [Y_t, Y_{t-1}, \cdots, Y_{t-p_1}]\), then, we get, for the process \((Y_t)_t\), the following representation:

\[
Y_t = (1 - I_d(c))\mathbf{Y}_{t-1} \Phi_0 + I_d(c)\mathbf{Y}_{t-1} \Phi_1 + ((1 - I_d(c))\sigma_0 + I_d(c)\sigma_1)\varepsilon_t.
\]

Now, we assume that we observe a sequence of data \((Y_1, \cdots, Y_n)\) from the model (24). The equation (24) is a regression equation (albeit nonlinear in parameters) and an appropriate estimation method is least squares (LS). Under the auxiliary assumption that the noise \((\varepsilon_t)_t\) is a strong Gaussian white noise, the least squares estimation is equivalent to the maximum likelihood estimation.

Since the regression equation (24) is nonlinear and discontinuous, the easiest method to obtain the LS estimates is to use sequential conditional LS. We will use this approach here. Recall that conditional least squares lead to the minimization of:

\[
\sum_{Y_{t-d} < c, t=1} Y_t - \phi_{0,0} + \phi_{0,1} Y_{t-1} - \cdots - \phi_{0,p_1} Y_{t-p_1})^2 + \sum_{Y_{t-d} > c, t=1} (Y_t - \phi_{0,0} + \phi_{0,1} Y_{t-1} - \cdots - \phi_{0,p_1} Y_{t-p_1})^2 \quad \text{min},
\]

with respect to \(\Phi_0, \Phi_1, c, d, p_1, p_2\). Generally, we first assume that the parameters \(p_1\) and \(p_2\) are known.

Recall that Chan (1993) proves that, under geometric ergodicity and some other regularity conditions for the process (24), that the LS parameters estimates of this process have good properties. The threshold parameter is consistent, tends to the true value at rate \(n\) and suitably normalized, follows asymptotically a Compound Poisson process. The other parameters of the
model are $n^{-1/2}$ consistent and are asymptotically distributed. The limitation of the theory of Chan (1993) concerns the construction of confidence intervals for the threshold $c$. Indeed, if we denote $\hat{c}$ the LS estimate for $c$, Chan (1993) finds that $(\hat{c} - c)$ converges in distribution to a functional of a Compound Poisson process and unfortunately, this representation depends upon a host of nuisance parameters, including the marginal distribution of $(Y_t)_t$ and all the regression coefficients. Hence, this theory does not yield a practical method to construct confidence intervals. Some discussions and extensions to this work can be found in Hansen (2000), Clements and Smith (2001) and Gonzalo and Pitarakis (2002), for instance.

The method used here to estimate all the parameters follows Lim and Tong (1980) and Tsay (1989). We need to determine the parameters $c, d, p_1,p_2$. We assume $P$ the maximal possible order of the two subregimes and $D$ the greatest possible delay. The threshold parameter $c$ is chosen by grid search. The grid points are obtained using the quantiles of the sample under investigation. We use equally spaced quantiles from the 10 (percent) quantiles and ending at the 90 (percent) quantiles. Now, for each fixed pair $(d, c_i)$, $0 < d < D$, $i = 1, \ldots, s$, the appropriate TAR model is to be identified. The AIC criterion is used for selection of the orders $p_1$ and $p_2$. In this context, it becomes:

$$AIC(p_1, p_2, d, c) = \ln\left(\frac{1}{n} \sum \varepsilon_t^2 \right) + 2 \frac{p_1 + p_2 + 2}{n},$$

(27)

where $\varepsilon_t$ denotes the residuals.

Finally the model with the parameters $p_1^*, p_2^*, d^*$ and $c^*$ that minimize the AIC criterion can be chosen. Since for different $d$ there are different numbers of values that can be used for estimation, the following adjustment should be done. With $n_d = \max(d, P)$ it is:

$$AIC(p_1^*, p_2^*, d^*, c^*) = \min_{p_1, p_2, d, c} \frac{1}{n - n_d} AIC(p_1, p_2, d, c).$$

(28)

The fitting procedure used in this paper follows the suggestion of Tong (1990) and we refer to his book for the details. Different algorithms have been proposed to improve the properties of the estimates and the speed of the algorithms: we specify now some references. Although we do not consider here a econometrician approach using tests to justify the use of SETAR models, we also recall some of the tests for different classes of STAR and SETAR processes.

Concerning the algorithms of the parameters’ estimation for SETAR models we recall the most important approaches. Bayesian inference via the Gibbs sampler is developed in Tiao and Tsay (1994). Several graphical procedures
are proposed by Chen (1995) classifying the observations without knowing the threshold variable to estimate the parameters. Different numerical approaches to making the estimation of the threshold autoregressive time series more efficient is developed by Coakley, Fuertes and Pérez (2003). Adopting the Markov chain Monte Carlo techniques, So and Chen (2003) identify what they call the best subset model characterizing a SETAR model. A review on classical and Bayesian estimation techniques is done by Potter (1999).

Tsay (1989) proposes a statistic to test the threshold nonlinearity and specify the threshold variable. This test statistic is derived by simple linear regression and its performance is evaluated by simulation. Hansen (1997) considers a likelihood ratio statistic for testing SETAR hypotheses. A Lagrange Multiplier test is proposed by Proietti (1998).

Concerning SETAR-ARCH models, Li and Li (1996), provide model identification, estimation and diagnostic checking techniques: the maximum likelihood estimation is achieved via an easy-to-use iteratively weighted least squares algorithm. Maximum likelihood and least squares estimation is considered for threshold heteroscedastic models in Zakoian (1994), see also Chan and McAl eer (2002). A Bayesian approach is proposed for the Threshold Stochastic Volatility model by So, Li and Lam (2002) and genetic algorithms are developed by Wu and Chang (2002). Wong and Li (2000) study the asymptotic null distribution of the Lagrange Multiplier test statistic in the context of SETAR-ARCH model.

Linearity testing against smooth transition autoregression, determining the delay parameter and choosing different STAR models are discussed in Terasvirta (1994). A Lagrange Multiplier test for STAR models is developed in Eitrheim and Terasvirta (1996). A pseudo-true score encompassing test to distinguish between competing STAR models is proposed by Chen (2003). Misspecification tests are developed in Lundberg and Terasvirta (1998) for STAR-STGARCH models. Test for unit root in the nonlinear STAR framework is derived by Kapetanios, Shin and Snell (2003).

We can remark that no test is exhibited to decide between SETAR models and switching models. This seems very difficult to settle.

3.2 Forecasts for a SETAR

To make forecast with a general SETAR model stays until now an open problem. Some works have been done in a Gaussian context. We specify the more classical method used by the practitioners. We assume that the observations are explained by a process such that (1). Say, we assume that \((Y_t)_t\) is a linear autoregression within a regime, but may move between
regimes depending on the values taken by a lag of $Y_t$, say $Y_{t-d}$, where $d$ is assumed to be known as the length of the delay. We assume also that the disturbances $(\varepsilon_t)_t$ are Gaussian processes.

Assume that we observe $(Y_1, \cdots, Y_n)$ from (1), one way to forecast using such a SETAR model is the following: one takes a weighted average of the forecast from the regime 1 and the regime 2. At time $n$, for an horizon $k$, these forecasts are denoted $\hat{Y}_{1,n+k}$ for the regime 1 and $\hat{Y}_{2,n+k}$ for the regime 2 and equal to:

$$\hat{Y}_{1,n+k} = \phi_{0,1} + \phi_{1,1}\hat{Y}_{1,n+k-1},$$  \hspace{1cm} (29)

and

$$\hat{Y}_{2,n+k} = \phi_{0,2} + \phi_{1,2}\hat{Y}_{2,n+k-1}.$$ \hspace{1cm} (30)

The forecast of $Y_{n+k}$, denoted $\hat{Y}_{n+k}$, is then:

$$\hat{Y}_{n+k} = p_{k-1}\hat{Y}_{1,n+k} + (1-p_{k-1})\hat{Y}_{2,n+k}$$ \hspace{1cm} (31)

$$+ (\phi_{1,2} - \phi_{1,1})\hat{\sigma}_{n+k-1} \phi\left(\frac{c - \hat{Y}_{n+k-1}}{\hat{\sigma}_{n+k-1}}\right),$$ \hspace{1cm} (32)

for $k = 2, \cdots$. The weight $p_{k-1}$ is the probability of the process being in the lower regime at time $n + k - 1$ assuming normality is $\phi\left(\frac{c - \hat{Y}_{n+k-1}}{\sigma_{n+k-1}}\right)$, where $\phi(.)$ is the probability density function of the Standard Gaussian $\mathcal{N}(0,1)$ and $\hat{\sigma}_{n+k-1}$ is the variance of the residuals at time $n + k - 1$.

Now, substituting $\hat{Y}_{1,n+k}$, $\hat{Y}_{2,n+k}$ and $p_{k-1}$ in (31), we obtain the following recursive relation for obtaining approximate $k$-step ($k > 1$) forecasts:

$$\hat{Y}_{n+k} = \Phi\left(\frac{c - \hat{Y}_{n+k-1}}{\hat{\sigma}_{n+k-1}}\right)(\phi_{0,1} + \phi_{1,1}\hat{Y}_{n+k-1})$$ \hspace{1cm} (33)

$$+ \Phi\left(\frac{c - \hat{Y}_{n+k-1}}{\hat{\sigma}_{n+k-1}}\right)(\phi_{0,2} + \phi_{1,2}\hat{Y}_{n+k-1})$$ \hspace{1cm} (34)

$$+ \phi\left(\frac{c - \hat{Y}_{n+k-1}}{\hat{\sigma}_{n+k-1}}\right)(\phi_{1,2} - \phi_{1,1})\hat{\sigma}_{n+k-1},$$ \hspace{1cm} (35)

where $\Phi(.)$ is the distribution function for the standard Gaussian distribution $\mathcal{N}(0,1)$.

For more details, we refer to AlQassam and Lane (1989) and to Clements and Smith (1999) for the Monte Carlo method of calculating these SETAR forecasts, see also de Gooijer and de Bruin (1998).
4 Empirical results

In this section, our aim is to apply a SETAR model to the Euro-zone Industrial Production Index in order to detect the low phases of the industrial business cycle referred to as the industrial recessions. The application is done in two steps: first we try to find the best SETAR model according to the AIC criterion presented in the previous section and second we use this model to detect the periods of each regime. By comparing the results to reference recession dates, we can assess the ability of the model to reproduce the industrial business cycle features.

4.1 Data description

The analysis is carried out on the IPI series considered in the paper of Anas et al. (2003). This series is a proxy of the monthly aggregate Euro-zone IPI for the 12 countries, beginning in January 1970 and ending in December 2002. The data are working day adjusted and seasonally adjusted by using the Tramo-SEATS methodology implemented in the Demetra software. Moreover, the irregular part including outliers has been removed.

Figure 1: Euro12 IPI (top) and its monthly growth rate (bottom), as well as the reference industrial recession periods (shaded areas), from January 1970 to December 2002.
The original series \((X_t)_t\) is presented in figure 1 as well as its monthly growth rate \((Y_t)_t\) defined by \(Y_t = \log(X_t) - \log(X_{t-1})\). In Figure 1, the shaded areas represent the reference industrial recession dates. Several authors have proposed a turning point chronology for the Euro-zone industrial business cycle, by using different statistical techniques and economic arguments. For example, we refer to Anas et al. (2003), who propose a classical NBER-based non-parametric approach, and to Artis et al. (2003), Krolzig (2003) or Anas and Ferrara (2002b) who apply parametric Markov-Switching models. Generally, the industrial recession dates are more or less similar. In fact, it turns out that the Euro-zone experienced five industrial recessions: in 1974-75 and 1980-81 due to the first and second oil shocks, in 1981-82, in 1992-93, due to the American recession and the Gulf war, and lastly in 2000-2001 because of the global economic slowdown caused itself by the US recession from March 2001 to November 2001. It is noteworthy that, contrary to a common belief among economists, the Asian crisis in 1997-98 has not caused an industrial recession in the whole Euro-zone, but only a slowdown of the production. Finally, we retain as a benchmark for our study the dates proposed by Anas et al. (2003) and summarized in the first column in table 4.

To ensure stationarity, we are going to deal with the monthly industrial growth rate \((Y_t)_t\). The unconditional empirical distribution of the IPI growth rate computed by using a non-parametric kernel estimate (with the Epanechnikov kernel) is presented in figure 2. There is a clear evidence of three peaks in the estimated distribution. The lowest peak is due to the negative growth rates during industrial recessions. The intermediate peak seems to be caused by periods of low, but positive, growth rates, experienced for example dur-
ing the eighties, while the peak corresponding to the highest value is related to periods of fast growth. It is noteworthy that, from 1970 to 2002, periods of low growth rates seem to appear more frequently than periods of high growth rates. Moreover, this empirical distribution is clearly asymmetric (skewness equal to -0.9315) and with heavy tails (excess kurtosis equal to 2.4850). Consequently, the unconditional Gaussian assumption is strongly rejected by a Jarque-Bera test.

4.2 Whole sample modelling

In this subsection we fit various SETAR models to the industrial growth rate series \((Y_t)_{1, t}\), that is, we model the speed of the Euro-zone industry. We consider first a two-regime model, the transition variable being successively the lagged series and the lagged differenced series. Then, we consider a multiple regime model by mixing the conditions on these previous series. For each model, we compare the estimated regimes with the reference recession phases in order to assess the ability of the model to reproduce business cycle features.

The first SETAR model uses the lagged series \(Y_{t-d}\) as transition variable. Thus, we model the speed of the industrial production according to the regimes of the lagged speed. The delay \(d\) and the threshold \(c\) are estimated by using the methodology presented in the previous section. However, the autoregressive lag \(p\) has to be determined \(a \text{ priori}\). We proceed by using a descendent stepwise approach by considering first \(p = 12\). For all estimated models, it turns out that the parameters corresponding to a lag greater than three are statistically not significant by the usual Student test. Therefore, we impose the choice \(p = 3\) for all the models. We get the following estimates for \(c\) and \(d\): \(\hat{c} = -0.0024\) and \(\hat{d} = 1\). The full estimated model is as follows (estimates and their standard errors are given in table 1):

\[
Y_t = (-0.0047 + 1.0843Y_{t-1} - 0.4055Y_{t-2} + 0.0859Y_{t-3})(1 - I_{|Y_{t-1}| > 0.0024})
\]
\[
+ (0.0025 + 1.3950Y_{t-1} - 0.8742Y_{t-2} - 0.3318Y_{t-3})I_{|Y_{t-1}| > 0.0024} + \varepsilon_t.
\]

We note from table 1 that, in the high regime, the persistence is stronger, because the parameters corresponding to \(p = 2\) and \(p = 3\) in the low regime are not statistically significant, and the variance is smaller, which are expected results in business cycle analysis. The empirical unconditional probabilities of being in each regime are \(\pi_1 = 0.11\) and \(\pi_2 = 0.89\), which is consistent with the usual probabilities of being in recession and expansion in business cycle analysis. As regards the estimated recession dates, we get them by assuming that the low regime matches with the recession regime. The results are presented in figure 3 (top graph) and table 4 along with the two other dating
chronologies stemming from the models described below. By comparison with the reference dating chronology, we can observe that the results are basically identical, except that we get a supplementary of recession in 1977, lasting only three months. If we had to establish a dating chronology, this period would not be retained as an industrial recession insofar as its duration is too short in comparison with the minimum duration of a business cycle phase, which generally of six months. However in this paper, to avoid non-persistent signals, we adopt the censoring rule saying that a signal must stay at least three months to be recognized as an estimated recession phase. Thus, this supplementary recession in 1977 is interpreted as a false signal of recession. In the remaining of this paper, a recession phase detected by the model but not present in the reference chronology is interpreted as a false signal of recession. Regarding the last industrial recession, the model estimates a recession period cut into two parts. This can be interpreted as a false signal of recovery. We also note that the other estimated industrial recessions are shorter, especially the 1982 recession but we get a first signal of recession in January 1982 which was not persistent. Otherwise, this model does not provide any other false signal for industrial recession.

The second SETAR model uses the differenced lagged series as transition variable, that is we try to model the speed of the industrial production according the regimes of its acceleration. We note this series $Z_{t-d}$, defined such as $Z_{t-d} = Y_{t-1} - Y_{t-d}$. Actually, this series can be considered as a proxy of the acceleration of the IPI over $d - 1$ months. It is interesting to investigate how the growth rate is related to the acceleration through a non-linear relationship. It turns out that the delay $d$ corresponding to the minimum AIC is equal to $d = 10$. That is, the acceleration over nine months seems to be the most significant. The estimated model is given by the following equation (estimates and their standard errors are given in table

<table>
<thead>
<tr>
<th></th>
<th>Low regime $[Y_{t-1} \leq -0.0024]$</th>
<th>High regime $[Y_{t-1} &gt; -0.0024]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>-0.0047 (0.0013)</td>
<td>0.0025 (0.0004)</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>1.0843 (0.1579)</td>
<td>1.3950 (0.0513)</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.4055 (0.2236)</td>
<td>-0.8742 (0.0782)</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.0859 (0.1582)</td>
<td>0.3318 (0.0513)</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>0.0020</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Table 1: Estimates and standard errors for model 1.
2):

\[
Y_t = (-0.0046 + 1.0998Y_{t-1} - 0.2073Y_{t-2} - 0.1916Y_{t-3})(1 - I_{[Z_{t-10} > -0.0061]}) \\
+ (0.0024 + 1.7444Y_{t-1} - 1.3796Y_{t-2} + 0.5567Y_{t-3})I_{[Z_{t-10} > -0.0061]} + \varepsilon_t.
\]

Here again, we observe that the persistence is stronger in the higher regime while the variance is smaller and the empirical unconditional probabilities of being in each regime are exactly equal to the previous ones. The estimated industrial recession dates, presented in figure 3 (middle graph) and table 4, slightly differ from the previous estimates. Indeed, we get another false signal of industrial recession in 1998 due to the impact of the Asian crisis. Moreover, we note that the 1977 recession lasts seven months, but the 1982 recession is only of two months. Therefore, by considering the censoring rule adopted above, this model does not recognize this period as a recession. We also note that a non-persistent signal of recession is given in September 1995. Thus, by comparison with the reference dating chronology, this model provides two false signals of recession and misses the 1982 recession. Consequently, this model underperforms the previous one in detecting the industrial recession phases. This may be due to the fact that the acceleration, although computed over 9 months, appears to be too volatile.

Lastly, the idea which appears to be natural is to combine the two previous SETAR models in a single model with two transition variables: the lagged growth rate and the acceleration. Therefore, the model possesses four regimes and two thresholds \(c_1\) and \(c_2\) have to be estimated. The estimated model which minimizes the AIC is given by the following equations (estimates and their standard errors are given in table 3):

- **Regime 1**: if \(Y_{t-1} < -0.00148\) and \(Z_{t-10} < -0.00076\), then

\[
Y_t = -0.0040 + 1.1454Y_{t-1} - 0.3712Y_{t-2} - 0.0359Y_{t-3} + \varepsilon_t,
\]

<table>
<thead>
<tr>
<th>(Z_{t-10} \leq -0.0061)</th>
<th>(Z_{t-10} &gt; -0.0061)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_0)</td>
<td>(0.0024)</td>
</tr>
<tr>
<td>(0.0046)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>(1.0998)</td>
</tr>
<tr>
<td>(0.1568)</td>
<td>(0.0456)</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>(-0.2073)</td>
</tr>
<tr>
<td>(0.2349)</td>
<td>(0.0742)</td>
</tr>
<tr>
<td>(\phi_3)</td>
<td>(-0.1916)</td>
</tr>
<tr>
<td>(0.1584)</td>
<td>(0.0458)</td>
</tr>
<tr>
<td>(\sigma_\varepsilon)</td>
<td>(0.0019)</td>
</tr>
</tbody>
</table>

Table 2: Estimates and standard errors for model 2.
• Regime 2: if \( Y_{t-1} < -0.00148 \) and \( Z_{t-10} \geq -0.00076 \)
  \[
  Y_t = -0.0070 + 1.2803Y_{t-1} - 0.0180Y_{t-2} - 0.0001Y_{t-3} + \varepsilon_t,
  \]

• Regime 3: if \( Y_{t-1} \geq -0.00148 \) and \( Z_{t-10} < -0.00076 \)
  \[
  Y_t = 0.0010 + 0.7106Y_{t-1} - 0.0750Y_{t-2} - 0.0331Y_{t-3} + \varepsilon_t,
  \]

• Regime 4: if \( Y_{t-1} \geq -0.00148 \) and \( Z_{t-10} \geq -0.00076 \)
  \[
  Y_t = 0.0036 + 1.3005Y_{t-1} - 0.7883Y_{t-2} - 0.3321Y_{t-3} + \varepsilon_t.
  \]

The two thresholds are estimated by using a double loop, but the delays of the model are fixed \textit{a priori} according the two previous estimated models. Both estimated threshold are negative but very close to zero. The

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Regime 2</th>
<th>Regime 3</th>
<th>Regime 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_{t-1} &lt; -0.0015 )</td>
<td>( Y_{t-1} &lt; -0.0015 )</td>
<td>( Y_{t-1} \geq -0.0015 )</td>
<td>( Y_{t-1} \geq -0.0015 )</td>
</tr>
<tr>
<td>( Z_{t-10} &lt; -0.0008 )</td>
<td>( Z_{t-10} &lt; -0.0008 )</td>
<td>( Z_{t-10} \geq -0.0008 )</td>
<td>( Z_{t-10} \geq -0.0008 )</td>
</tr>
<tr>
<td>( \phi_0 )</td>
<td>-0.0040</td>
<td>-0.0070</td>
<td>0.0010</td>
</tr>
<tr>
<td>(0.0010)</td>
<td>(NA)</td>
<td>(0.0004)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>1.1454</td>
<td>1.2803</td>
<td>0.7106</td>
</tr>
<tr>
<td>(0.1408)</td>
<td>(NA)</td>
<td>(0.0995)</td>
<td>(0.0660)</td>
</tr>
<tr>
<td>( \phi_2 )</td>
<td>-0.3712</td>
<td>-0.0180</td>
<td>-0.0750</td>
</tr>
<tr>
<td>(0.2076)</td>
<td>(NA)</td>
<td>(0.1216)</td>
<td>(0.0974)</td>
</tr>
<tr>
<td>( \phi_3 )</td>
<td>-0.0359</td>
<td>-0.0001</td>
<td>-0.0331</td>
</tr>
<tr>
<td>(0.1408)</td>
<td>(NA)</td>
<td>(0.1003)</td>
<td>(0.0662)</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>0.0019</td>
<td>NA</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

Table 3: Estimates and standard errors for model 3.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>m4 1974</td>
<td>m6 1974</td>
<td>m6 1974</td>
</tr>
<tr>
<td>Trough</td>
<td>m5 1975</td>
<td>m5 1975</td>
<td>m6 1975</td>
</tr>
<tr>
<td>Peak</td>
<td>-</td>
<td>m3 1977</td>
<td>m12 1976</td>
</tr>
<tr>
<td>Trough</td>
<td>-</td>
<td>m6 1977</td>
<td>m7 1977</td>
</tr>
<tr>
<td>Peak</td>
<td>m2 1980</td>
<td>m4 1980</td>
<td>m3 1980</td>
</tr>
<tr>
<td>Trough</td>
<td>m1 1981</td>
<td>m10 1980</td>
<td>m10 1980</td>
</tr>
<tr>
<td>Peak</td>
<td>m10 1981</td>
<td>m5 1982</td>
<td>m6 1982</td>
</tr>
<tr>
<td>Trough</td>
<td>m12 1982</td>
<td>m12 1982</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>m1 1992</td>
<td>m4 1992</td>
<td>m7 1992</td>
</tr>
<tr>
<td>Trough</td>
<td>m5 1993</td>
<td>m5 1993</td>
<td>m1 1993</td>
</tr>
<tr>
<td>Peak</td>
<td>-</td>
<td>-</td>
<td>m7 1998</td>
</tr>
<tr>
<td>Trough</td>
<td>-</td>
<td>-</td>
<td>m1 1998</td>
</tr>
<tr>
<td>Peak</td>
<td>m12 2000</td>
<td>m2 2001</td>
<td>m1 2001</td>
</tr>
<tr>
<td>Trough</td>
<td>m12 2001</td>
<td>m12 2001</td>
<td>m10 2001</td>
</tr>
</tbody>
</table>

Table 4: Reference and estimated dating chronologies stemming from the 3 considered SETAR models.
first regime has an empirical unconditional probability of 0.15 and should be considered at a first sight as a period of recession because the estimated recession dates match the reference recession dates. However, the second regime is also meaningful. Indeed, this second regime possesses an unconditional probability of 0.02: only 7 observations over 385 belong to this state. This is the reason why standard errors of estimates are not available in this regime. Although the frequency of this second regime is very low, this regime is persistent and appear in clusters. In fact, this regime is very interesting because it corresponds to the end of a recession phase when the economy is accelerating again. This regime was detected twice: at the end of the 1974-75 recession and at the end of the 1992-93 recession. Thus, the sum of regime 1 and regime 2 corresponds to the industrial recession phase.
The third regime can be considered as a slowdown of the industrial production, that is the industry is below its trend growth rate without being in recession. Lastly, when the series is in the high regime, we can deduce that the industrial growth rate is over its trend growth rate. Actually, regime 3 and regime 4 correspond to the high phase of the industrial business cycle. It appears that only three regimes would be sufficient to describe the industrial business cycle. However, we decide to keep four states because it gives a deeper understanding of the industrial business cycle features. As regards the dating results, the model provides almost the same results than the first model, the last recession period being not cut into two parts (see figure 3, bottom graph, and table 4). However, this model presents some non-persistent signals of recession.

After this whole sample analysis, we retain the third SETAR model with four regimes for the dynamic real-time analysis, because it provides the more accurate description of the industrial business cycle.

4.3 Dynamic real-time analysis

In real-time analysis, an economic indicator requires at least two qualities: it must be reliable and must provide a readable signal as soon as possible. Thus, there is a well known trade-off between advance and reliability for the economic indicators. By using the previous 4-regime SETAR model, we assess if it is possible to have a clear and timely signal for the turning points of the industrial business cycle in a dynamic analysis.

In this part, we consider the previous IPI series from January 1970 to December 2002.

![Figure 4: Euro12 IPI and the real-time estimated recession period (shaded area), from January 2000 to December 2002.](image)
cember 1999, and we add progressively a monthly data until December 2002. For each step, we re-estimate the model and we classify the series into one of the four regimes. Thus, by using the conclusions of the whole-sample analysis, if the series lies into regime 1 or regime 2, we can conclude that the industry is in a recession phase. We are aware that a true real-time time analysis should be done by using historically released data (see for instance Chauvet and Piger, 2003) in order to take the revisions and the edge-effects of the statistical treatments of the raw data into account. However, such series are very difficult to find in economic data bases.

The results of the real-time estimated recession period are presented in figure 4. We observe that these results match with the 2001 recession period estimated in the whole-sample analysis. This fact points out the stability of the model. Indeed, we detect a peak in the business cycle in February 2001 and a trough in December 2001. However, it must be noted that a false signal of a change in regime is emitted in August 2001 but it lasts only one month. Knowing that a signal must be persistent to be reliable, we have to propose an ad hoc real-time decision rule. Thus, it is advocated to wait at least two months before sending a signal of a change in regime. We also note that the exit of the recession is very fast, because the series goes directly from regime 1 in December 2001 to regime 4 in January 2002. Moreover, we observe that the series falls into regime three in December 2002.

![Threshold $\hat{C}_1$](image1)

![Threshold $\hat{C}_2$](image2)

**Figure 5:** Evolution of the real-time estimated thresholds of the 4-regime SETAR from January 2000 to December 2002.
It is also interesting to consider the evolution of the parameters in a dynamic analysis. In figure 5, the evolution of the thresholds \( \hat{c}_1 \) and \( \hat{c}_2 \) is presented. It is striking to observe the change in level of both thresholds during the recession phase. During this phase, thresholds tend to become closer to zero. It is also noteworthy that \( \hat{c}_1 \) increases slowly from May 2000 to June 2001 but decreases suddenly, while, conversely, \( \hat{c}_2 \) increases suddenly in March 2001 but decreases slowly. This feature indicates perhaps an asymmetry between the start and the end of recession and may be exploited later to get a more advanced signal. Lastly, we note that both thresholds are remarkably stable since the end of the recession. Unfortunately, as noted in section 3, there is no practical way to test a change in the thresholds.

**Conclusion**

This paper is an exploratory analysis of the ability of SETAR models to reproduce the business cycle stylised facts. The results are promising. It appears that the model allows to identify the turning points of the industrial cycle and can thus be useful for real-time detection. However, a true real-time analysis should be extended by using historically released data, as used in the recent paper of Chauvet and Piger (2003) as regards the US GDP and employment. Unfortunately, such data are not systematically stored in data bases and are therefore very difficult to get. As another example, business surveys seem to be good candidates for real-time analysis through SETAR models because they are timely released and are not generally revised.

**References**


