A Component-Driven Model for Regime Switching and Its Empirical Evidence

Chung-Ming Kuan
Institute of Economics
Academia Sinica

Yu-Lieh Huang
Department of Quantitative Finance
National Tsing-Hua University

Ruey S. Tsay
Graduate School of Business
University of Chicago

This version: April 24, 2003

† We thank Y.-T. Chen, R. Engle, C. W. J. Granger, C.-C. Hsu, C.-H. Huang, C.-J. Kim, C.-S. Mao, K. Tanaka and T. Teräsvirta for valuable suggestions and comments. C.-M. Kuan gratefully acknowledges research support of the National Science Council of the Republic of China (NSC-89-2415-H-001-037 and -071). The work of RST is supported by the National Science Foundation and the Graduate School of Business, University of Chicago.

Abstract

In this paper we propose a general component-driven model to analyze economic data with different characteristics (or regimes) in different time periods. Motivated by empirical data characteristics, our discussion focuses on a simple model driven by a random walk component and a stationary ARMA component that are governed by a Markovian state variable. The proposed model is capable of describing both stationary and non-stationary behaviors of data and allows its random innovations to have both permanent and transitory effects. This model also permits a deterministic trend with or without breaks and hence constitutes intermediate cases between the trend-stationary model and a random walk with drift. We investigate properties of the proposed model and derive an estimation algorithm. A simulation-based test is also proposed to distinguish between the proposed model and an ARIMA model. In empirical application, we apply this model to U.S. quarterly real GDP and find that unit-root nonstationarity is likely to be the prevailing dynamic pattern in more than 80 percent of the sample periods. As nonstationarity (stationarity) periods match the NBER dating of expansions (recessions) closely, our result suggests that the innovations in expansions (recessions) are more likely to have a permanent (transitory) effect.

Keywords: component-driven model, Markov trend, permanent shock, regime switching, transitory shock, trend stationarity, unit root

JEL Classification: C22. C51

1 Introduction

It has been well documented in the literature that many economic and financial data exhibit different characteristics (or regimes) over time. The econometric models that are able to accommodate multiple regimes include structure-change models, threshold models, and Markov switching models. In these models, it is typical to postulate one function for the data and characterize different regimes by distinct parameter values of the function. While this modeling strategy is convenient and useful in various applications, it is rather restrictive because it only permits similar dynamic patterns in different regimes. An alternative approach is to model data regimes using different functions. Although there are some models constructed along this line (e.g., Evans and Wachtel, 1993, and Engle and Smith, 1999), this approach did not receive much attention in the literature.

In this paper, we follow the latter approach and propose a general component-driven model to characterize data in different regimes. The model consists of two (or more) unobserved components and a state variable linked specifically to innovations. Depending on the value of the state variable, each innovation excites only one of the components, and the prevailing component in turn determines the new data dynamics. By assigning distinct functions to different components, the proposed model is able to characterize completely different dynamic structures in different regimes. Our choice of functions in the paper is motivated by two leading models in time series econometrics, viz., the unit-root model and the trend-stationary model. Since the seminal work of Nelson and Plosser (1982), unit-root nonstationarity has been widely accepted as a stylized fact of many macroeconomic time series; see also Campbell and Mankiw (1987). Yet some researchers believe that trend-stationary models with or without breaks are better modeling tools; see, e.g., Blanchard (1981), Clark (1987), Perron (1989), and Rappoport and Richelin (1989). It is therefore natural to consider a model that can accommodate both dynamic patterns.

For simplicity, we specialize on the model driven by a random walk component and a stationary ARMA (autoregressive and moving average) component that are governed by a Markovian state variable. This model also permits a deterministic trend with or without breaks and hence constitutes intermediate cases between the trend-stationary model and a random walk with drift; otherwise, it is between the ARMA model and a random walk without drift. If the components are state independent, the model reduces to one of the two extreme cases. The proposed model is in sharp contrast with the commonly used time series models and the regime-switching model of Hamilton (1989) in that it is capable of describing both stationary and non-stationary dynamics and allows its random innovations to have both permanent and transitory effects. Moreover, trend breaks are endogenous in the sense that the trend function shifts when permanent shocks are present. Comparing to the model of Evans and Wachtel (1993) and McCulloch and Tsay (1994), the Markovian state variable here is attached to innovations and entails different dynamic behaviors. This model also results in an ARMA representation with random MA coefficients and hence is quite different from the more familiar, random AR-coefficient models, such as those of McCabe and Tremayne (1995) and Granger and Swanson (1997).

In this paper, we derive an estimation algorithm for the proposed model and propose a simulation-based test to distinguish between the model and an ARIMA process. As an application, we apply the proposed model to U.S. quarterly real GDP. We find that unit-root nonstationarity is likely to be the prevailing dynamic pattern for more than 80 percent of the sample periods, whereas the remaining periods are likely to be in the stationarity regime. This finding is quite different from the inferences drawn from unit-root and trend-break models. It is also observed that the nonstationarity (stationarity) periods match the NBER dating of expansions (recessions) very closely. These results suggest that the innovations in expansions (recessions) are more likely to have a permanent (transitory) effect. That the shocks in expansions are more persistent than those in recessions is compatible with the conclusion of Beaudry and Koop (1993). Our empirical results thus provide an alternative view of the characteristics of U.S. real GDP.

This paper is organized as follows. Section 2 introduces the proposed component-driven model and compares it with some existing models. Section 3 derives some statistical properties of the proposed model. We discuss the estimation algorithm and hypothesis testing in Section 4 and Section 5, respectively. The empirical analysis of U.S. real GDP based on the proposed model is presented in Section 6, and Section 7 concludes the paper.

2 The Proposed Component-Driven Model

Suppose that the process y_t can be modeled as the sum of two components — namely, $y_t = y_{1,t} + y_{0,t}$, with

$$\begin{aligned} y_{1,t} &= g(y_{1,t-1}, \dots, y_{1,t-p}; \boldsymbol{\theta}_1) + s_t \upsilon_t, \\ y_{0,t} &= h(y_{0,t-1}, \dots, y_{0,t-q}; \boldsymbol{\theta}_0) + (1 - s_t)\upsilon_t, \end{aligned} \tag{1}$$

where g and h are two possibly different functions, p and q are positive integers, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ are parameter vectors, s_t is the state variable at time t that assumes the value one or zero, and υ_t are random variables. This model is readily generalized to allow for more than two components. To ease exposition, we shall focus on the two-component model.

For the model in Eq. (1), each innovation excites only one component rather than the entire observation, and the state variable s_t determines the prevailing component and hence the effect of υ_t. When $s_t = 1$, the first component $y_{1,t}$ is activated and the effect of υ_t is propagated through g, while the second component $y_{0,t}$ evolves according to h without any innovation. When $s_t = 0$, υ_t excites $y_{0,t}$ but does not enter the first component y_{1t}. As both $s_t\upsilon_t$ and $(1-s_t)\upsilon_t$ are present in Eq. (1), s_t does not affect y_t but determines how υ_t affects subsequent y_t. The dynamics of this process is thus driven by the prevailing component and may change over time when s_t switches. When g and h are distinct functions, the dynamics described by these components may be very different. By contrast, standard regime-switching models, such as those of Hamilton (1989, 1994) and Tong (1990), contain one component with the state-dependent parameter vector $\boldsymbol{\theta}_{s_t}$:

$$y_t = g(y_{t-1}, \dots, y_{t-p}; \boldsymbol{\theta}_{s_t}) + \upsilon_t.$$

As there is only one function g in the model, the dynamic structures in different regimes are similar in essence. While the effect of υ_t on future y_t is controlled by the concurrent state s_t of the model in Eq. (1), it is s_{t+j} that determines how υ_t affects y_{t+j} in the model of Hamilton (1989, 1994).

2.1 The Proposed Model

Clearly, the model (1) is not operational unless the functions g and h are specified. Our choice of functions is motivated by two important data characteristics observed empirically, namely, unit-root nonstationarity and (trend) stationarity. There has been a long history of debate regarding whether an economic time series should be modeled using a unit-root model or a (trend-)stationary model. Distinguishing between these two behaviors is important in economics because their interpretations are different. When a time series contains a unit root, its innovations all have a permanent effect, and its time path exhibits large swings. When a time series is (trend) stationary, its innovations have a transitory effect and induce only short-run fluctuations around the trend. It is, however, conceivable that the random shocks in the periods of no significance are fundamentally different from those in turbulent periods. Restricting all the innovations of a model to have the same effect is convenient but may not always be appropriate. For example, Beaudry and Koop (1993) show that positive shocks to U.S. GDP are more persistent than negative shocks; see also Bradley and Jansen (1997) and Hess and Iwata (1997). A flexible model that allows for distinct dynamic patterns and innovations is thus desired.

Let $\upsilon_t = \alpha_1 + \varepsilon_t$, where $\{\varepsilon_t\}$ is a white noise with mean zero. In view of the motivation above and model (1), the proposed component-driven model consists of a random walk component and a stationary ARMA component. That is, $y_t = y_{1,t} + y_{0,t}$ with

$$\begin{aligned}(1-B)y_{1,t} &= \alpha_0 + s_t v_t = (\alpha_0 + s_t\alpha_1) + s_t\varepsilon_t,\\ \Psi(B)y_{0,t} &= \Phi(B)(1-s_t)v_t = \Phi(B)(1-s_t)\alpha_1 + \Phi(B)(1-s_t)\varepsilon_t,\end{aligned} \tag{2}$$

where $\Psi(B) = 1 - \psi_1 B - \cdots - \psi_m B^m$ and $\Phi(B) = 1 - \varphi_1 B - \cdots - \varphi_n B^n$ are finite-order polynomials of the back-shift operator B such that they have no common factors and their roots are all outside the unit circle. As in Clark (1987) and Cochrane (1988), y_t in the proposed model is also the sum of a random walk and a stationary component, yet its dynamic structures alternate from time to time, owing to the presence of s_t with the innovations. This model is able to accommodate both unit-root and stationary dynamics. When $s_t = 1$, the first component evolves like a random walk, and the corresponding innovation ε_t has a permanent effect on future y_t. When $s_t = 0$, ε_t excites the stationary ARMA component and has only a transitory effect. To avoid identification problem, we do not include a state-independent constant term in the second component. Note also that allowing v_t to have a non-zero mean adds more flexibility to this model, as can be seen in the following paragraph. In what follows, we shall postulate that s_t follows a first-order Markov chain, as in Hamilton (1989); other switching mechanisms, such as a threshold mechanism, may also be adopted.

Setting $y_0 = 0$ and $\varepsilon_i = 0$ for $i \le 0$, we have the following representation:

$$\begin{aligned}y_t = \alpha_0 t + \alpha_1 \sum_{i=1}^{t} s_i + \alpha_1 \Psi(B)^{-1}\Phi(B)(1-s_t) + \sum_{i=1}^{t} s_i\varepsilon_i\\ + \Psi(B)^{-1}\Phi(B)(1-s_t)\varepsilon_t.\end{aligned} \tag{3}$$

The first component of this representation, $\alpha_0 t$, is a "fundamental" trend because it does not depend on the state variable. This trend function may be altered by the second component $\alpha_1 \sum_{i=1}^{t} s_i$ if $\alpha_1 \neq 0$. In particular, the fundamental trend has a level shift when there is a one-time permanent shock, and its slope changes to $\alpha_0 + \alpha_1$ when permanent shocks are present consecutively. When there are trend breaks, the third component involves a weighted sum of $1 - s_{t-j}$ and induces smooth transitions between trend segments. It can also be seen that the fourth component, $\sum_{i=1}^{t} s_i\varepsilon_i$, admits only those ε_i with $s_i = 1$ (permanent shocks) and can be interpreted as a flexible stochastic trend. The last component, $\Psi(B)^{-1}\Phi(B)(1-s_t)\varepsilon_t$, is a weakly stationary process generated by transitory innovations and gives rise to short-run fluctuations. The features of this model are now clear. First, it admits both deterministic and stochastic trends. Second, apart from the deterministic trend, it exhibits both stationary and nonstationary patterns over different time periods. Third, it allows for endogenous trend breaks, in the sense that breaks are due to the presence of permanent shocks.

Note that when $s_t = 1$ with probability one for all t,

$$y_t = (\alpha_0 + \alpha_1)t + \sum_{i=1}^{t} \varepsilon_i, = (\alpha_0 + \alpha_1) + y_{t-1} + \varepsilon_t,$$

which is a random walk with the drift term $\alpha_0 + \alpha_1$. When $s_t = 0$ with probability one for all t, y_t is a trend-stationary process without break:

$$y_t = \alpha_1 \Psi(1)^{-1} \Phi(1) + \alpha_0 t + \Psi(B)^{-1} \Phi(B) \varepsilon_t.$$

The model (2) thus constitutes intermediate cases between the trend-stationary model and a random walk with drift. When s_t follows the first-order Markov chain, the first two components of Eq. (3) are the "Markov trend" of Hamilton (1989). Although the Markov trend is kinked, the trend function here is smooth due to the third component of (3). If $\alpha_1 = 0$, there will be no trend break. In what follows, the model in Eq. (2) with $\alpha_1 \neq 0$ ($\alpha_1 = 0$) will be referred to as a component-driven model of order 1 and (m, n), i.e. CD$(1; m, n)$, with a smooth Markov (linear) trend. When both $\alpha_0 = \alpha_1 = 0$, this is simply a CD$(1; m, n)$ model without trend.

It is also straightforward to show that the model (2) has an ARMA representation with random MA coefficients:

$$\Psi(B)(1 - B)y_t = \alpha_0 \Psi(1) + \sum_{i=1}^{r+1} \xi_{i,s_{t-i}} (\alpha_1 + \varepsilon_{t-i}) + (\alpha_1 + \varepsilon_t), \tag{4}$$

where $r = \max\{m, n\}$,

$$\xi_{1,s_{t-1}} = \begin{cases} -\psi_1, & \text{if } s_{t-1} = 1, \\ -1 - \varphi_1, & \text{otherwise,} \end{cases} \qquad \xi_{i,s_{t-i}} = \begin{cases} -\psi_i, & \text{if } s_{t-i} = 1, \\ \varphi_{i-1} - \varphi_i, & \text{otherwise,} \end{cases}$$

for $i = 2, \dots, r$, and the last coefficient is

$$\xi_{r+1,s_{t-r-1}} = \begin{cases} 0, & \text{if } s_{t-r-1} = 1, \\ \varphi_r, & \text{otherwise;} \end{cases}$$

$\psi_i = 0$ for $i > m$ and $\varphi_i = 0$ for $i > n$. From (4) we can see that only the past state variables $s_{t-1}, \dots, s_{t-r-1}$ can affect y_t; the current state s_t is irrelevant to y_t because $s_t \varepsilon_t$ and $(1 - s_t)\varepsilon_t$ are both present at time t. In fact, the effect of a state variable can only be revealed in subsequent periods.

We stress that (2) is just one possibility of modeling economic time series based on (1). Many other interesting models may also be constructed as special cases of (1). For example, one may set y_t as the sum of a fractionally integrated component and an ARMA component, so that y_t may exhibit both long- and short-range dependence. It is also possible to specify a different switching mechanism for s_t, cf. the threshold-disturbance moving-average model of Elwood (1998).

Table 1: Moving-average representation of the component-driven model (5).

	$y_{1,t}$	$y_{0,t}$
$t=1$	0	ε_1
$t=2$	0	$\psi_1\varepsilon_1+\varepsilon_2$
$t=3$	ε_3	$\psi_1^2\varepsilon_1+\psi_1\varepsilon_2$
$t=4$	$\varepsilon_3+\varepsilon_4$	$\psi_1^3\varepsilon_1+\psi_1^2\varepsilon_2$
$t=5$	$\varepsilon_3+\varepsilon_4+\varepsilon_5$	$\psi_1^4\varepsilon_1+\psi_1^3\varepsilon_2$
$t=6$	$\varepsilon_3+\varepsilon_4+\varepsilon_5$	$\psi_1^5\varepsilon_1+\psi_1^4\varepsilon_2+\varepsilon_6$
$t=7$	$\varepsilon_3+\varepsilon_4+\varepsilon_5$	$\psi_1^6\varepsilon_1+\psi_1^5\varepsilon_2+\psi_1\varepsilon_6+\varepsilon_7$
	$\vdots$	$\vdots$
$t=\infty$	$\sum_{i=1}^{\infty} s_i\varepsilon_i$	0

2.2 Comparison with Existing Models

To compare with the existing models that allow for different structures, we consider the CD(1; 1, 0) model consisting of a random walk component and a stationary AR(1) component with $\alpha_0 = \alpha_1 = 0$:

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + s_t\varepsilon_t, \\ y_{0,t} &= \psi_1 y_{0,t-1} + (1-s_t)\varepsilon_t, \qquad |\psi_1| < 1. \end{aligned} \tag{5}$$

Table 1 gives the moving-average representation of $y_{1,t}$ and $y_{0,t}$ for $t = 1, \ldots, 7$ when $\{s_1, \ldots, s_6\} = \{0, 0, 1, 1, 1, 0\}$ and $y_{i,t}$=0 for $t \leq 0$ and $i = 0$, 1. Here, y_t is a stationary AR(1) process at the beginning and starts evolving like a random walk when $s_t = 1$. It can be seen that when s_t switches, the natures of past innovations are not altered. By letting t tend to infinity, only the random walk component ($y_{1,\infty} = \sum_{i=1}^{\infty} s_i\varepsilon_i$) remains; see the last row of Table 1.

The model (5) is conceptually different from that of Evans and Wachtel (1993) and McCulloch and Tsay (1994) which allows for switching between a random walk and an AR(1) process:

$$y_t = s_t y_{1,t} + (1-s_t)y_{0,t},$$

where $y_{1,t} = y_{1,t-1} + v_t$, $y_{0,t} = \psi_1 y_{0,t-1} + u_t$ with $|\psi_1| < 1$, and $\{u_t\}$ and $\{v_t\}$ are two sequences of random variables. A major difference between this model and the proposed model lies on the switching mechanism. Here, y_t switches between processes so that all the past innovations must change accordingly. Such a switching mechanism affects the past but not future dynamics. It also induces drastic changes in the time path of y_t, especially when t is large, given that a random walk wanders off quickly. By contrast, s_t in the proposed model determines the effect of ε_t on future observations but has no influence whatsoever on past innovations. Thus, y_t does not have sudden jumps because the new dynamics are resulted from new shocks, rather than from the entire history of past innovations. Note also that, as this model contains two sets of innovations, it is not easy to explain why one set prevails in some periods but does not play any role in the others. The proposed model does not have this problem.

By (4), the random MA-coefficient representation of the model (5) is

$$y_t = (1 + \psi_1)y_{t-1} - \psi_1 y_{t-2} + \xi_{s_{t-1}}\varepsilon_{t-1} + \varepsilon_t,$$

with $\xi_{s_{t-1}} = -\psi_1$ if $s_{t-1} = 1$ and $\xi_{s_{t-1}} = -1$ otherwise. This model is therefore different from the standard random AR-coefficient model, such as that of McCabe and Tremayne (1995). A simple random AR-coefficient model can be expressed as

$$y_t = a_t y_{t-1} + u_t,$$

where a_t are random variables, typically assumed to be exogenous. The stochastic unit-root model of Granger and Swanson (1997) is such that $a_t = \exp(\alpha_t)$ with α_t being a weakly stationary process with zero mean. Setting the initial value $y_1 = u_1$, we can write

$$y_t = u_t + \sum_{i=1}^{t-1}\Big(\prod_{j=0}^{i-1} a_{t-j}\Big)u_{t-i}.$$

Similar to the Evans-Wachtel model, such a model may not be easy to interpret when the product $\prod_{j=0}^{i-1} a_{t-j}$ switches from the stable region to the explosive region.

A similar but fundamentally different model is the STOPBREAK model proposed by Engle and Smith (1999). The moving-average representation of the simplest STOPBREAK model is

$$y_t = \sum_{i=1}^{\infty} q_{t-i}\varepsilon_{t-i} + \varepsilon_t, \tag{6}$$

where $q_{t-i} = \varepsilon_{t-i}^2/(\gamma + \varepsilon_{t-i}^2)$ with the parameter $\gamma \geq 0$. When γ is very small, as in the empirical example of Engle and Smith (1999), $q_t \approx 1$ so that the resulting process is not much different from a pure random walk. When ε_t has a continuous distribution, $q_t = 0$ occurs only with probability zero. That is, the STOPBREAK model cannot exhibit stationary behavior. As $0 < q_t \leq 1$ with probability one, this process is in effect an $I(1)$ process with nonlinear moving-average terms.

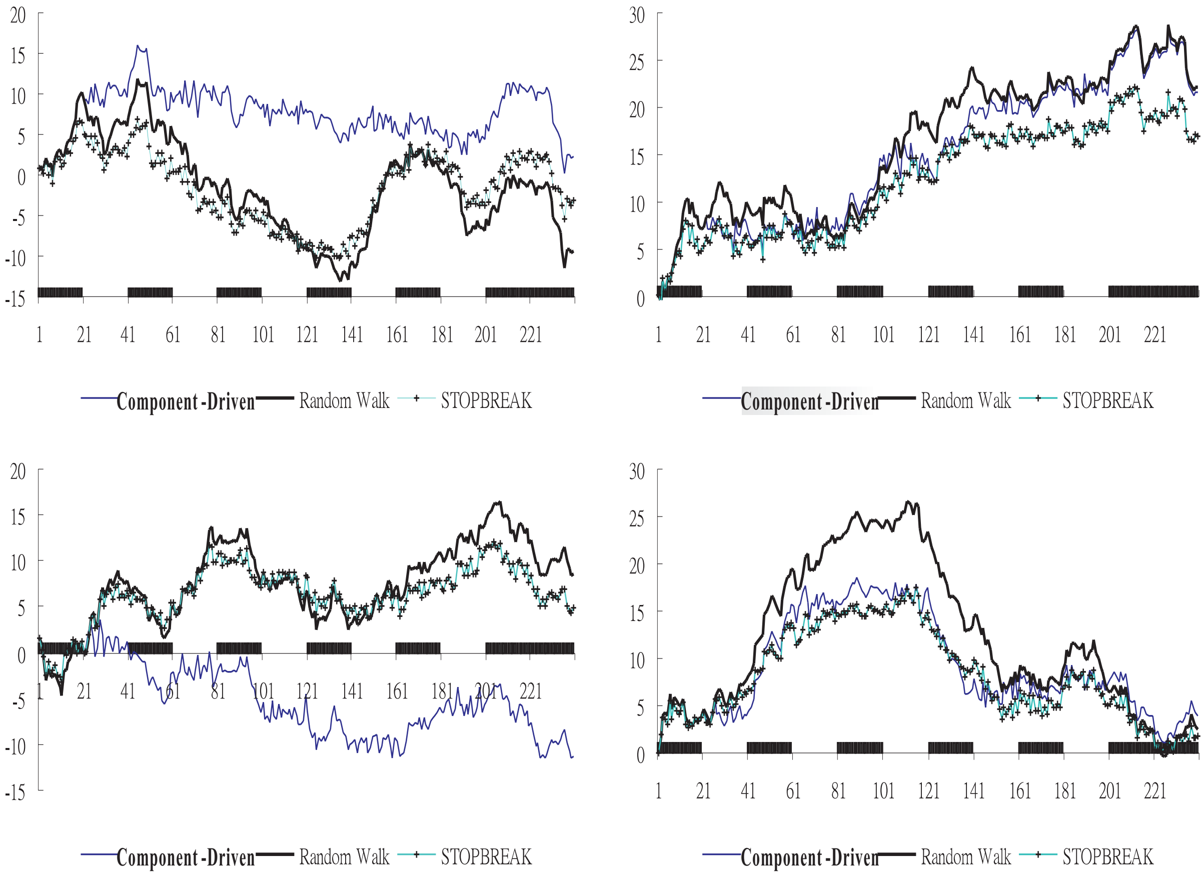


Figure 1: Simulated random walk, STOPBREAK process, and CD(1; 1, 0) process of Eq. (5).

To illustrate, we simulate a random walk, the STOPBREAK process (6) with $\gamma = 1$, and the process (5) with $\psi_1 = 0$. In our simulations, these processes are generated from the same innovations, and the process (5) is such that $s_t = 0$ for $t \in [21, 40]$, $[61, 80]$, $[101, 120]$, $[140, 160]$ and $[180, 200]$ and $s_t = 1$ otherwise. The simulated paths are plotted in four panels of Figure 1, where the thick line is the random walk, the line with "+" is the STOPBREAK process, and the thin line is the process (5). These figures show that the STOPBREAK process always mimics the random walk, yet the process (5) exhibits flexible dynamic patterns which may or may not be similar to the random walk.

3 Properties of the Component-Driven Model

To derive properties of the y_t process generated from Eq. (2), we maintain the assumption that s_t follows a first-order Markov chain with the transition matrix

$$\begin{bmatrix} \mathbb{P}(s_t = 0 \mid s_{t-1} = 0) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 0) \\ \mathbb{P}(s_t = 0 \mid s_{t-1} = 1) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 1) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}.$$

Let $S^t = \{s_t, s_{t-1}, \dots\}$ denote the collection of all state variables up to time t. We also assume that $\{\varepsilon_t\}$ is a sequence of random variables such that $\mathbb{E}(\varepsilon_t \mid S^t) = 0$, $\mathrm{var}(\varepsilon_t \mid S^t) = \sigma_\varepsilon^2$, and $\mathbb{E}(\varepsilon_t \varepsilon_{t-i} \mid S^t) = 0$ for all $i > 0$. By invoking the law of iterated expectations, it is easy to verify that $\{\varepsilon_t\}$ is a white noise. Also, $\mathbb{E}(s_t\varepsilon_t) = \mathbb{E}[s_t\,\mathbb{E}(\varepsilon_t \mid S^t)] = 0$,

$$\mathrm{var}(s_t\varepsilon_t) = \mathbb{E}[s_t^2\,\mathbb{E}(\varepsilon_t^2 \mid S^t)] = \sigma_\varepsilon^2\ \mathbb{P}(s_t = 1),$$

and $\mathrm{cov}(s_t\varepsilon_t, s_{t-i}\varepsilon_{t-i}) = \mathbb{E}[s_t s_{t-i}\,\mathbb{E}(\varepsilon_t\varepsilon_{t-i}|S^t)] = 0$. Similarly, $(1 - s_t)\varepsilon_t$ also has mean zero and variance $[1 - \mathbb{P}(s_t = 1)]\sigma_\varepsilon^2$ and are serially uncorrelated. These two series are white noise when $\mathbb{P}(s_t = 1)$ is a constant π_0. Moreover,

$$\mathrm{cov}(s_t\varepsilon_t, (1 - s_{t-i})\varepsilon_{t-i}) = \mathbb{E}[s_t(1 - s_{t-i})\,\mathbb{E}(\varepsilon_t\varepsilon_{t-i}|S^t)] = 0,$$

for $i \geq 0$, so that $s_t\varepsilon_t$ and $(1 - s_t)\varepsilon_t$ are mutually uncorrelated at all leads and lags.

We first consider the model (2) with $\alpha_1 = 0$. By (3),

$$y_t = \alpha_0 t + \sum_{i=1}^{t} s_i\varepsilon_i + \Psi(B)^{-1}\Phi(B)(1 - s_t)\varepsilon_t, \tag{7}$$

which is the sum of uncorrelated components. Then, $\mathbb{E}(y_t) = \alpha_0\, t$, and

$$\mathrm{var}(y_t) = \sigma_\varepsilon^2 \sum_{i=1}^{t} \mathbb{P}(s_i = 1) + \sigma_\varepsilon^2 \sum_{i=1}^{t} (\psi_i^*)^2[1 - \mathbb{P}(s_i = 1)],$$

where ψ_i^* are the coefficients of $\Psi(B)^* = \Psi(B)^{-1}\Phi(B) = (1 - \psi_1^* B - \psi_2^* B^2 - \cdots)$. When ψ_i^* are square summable, the second term on the right-hand side converges. Thus, $\mathrm{var}(y_t)$ would grow without bound if the partial sum $\sum_{i=1}^t \mathbb{P}(s_i = 1)$ diverges. In particular, when $\mathbb{P}(s_i = 1)$ is a constant $\pi_0 > 0$, $\mathrm{var}(y_t)$ would grow linearly with t as $\sigma_\varepsilon^2\pi_0 t$, which is proportional to that of a pure random walk. On the other hand, when $\sum_{i=1}^t \mathbb{P}(s_i = 1)$ converges, the Borel-Cantelli lemma implies $\mathbb{P}(s_i = 1 \text{ infinitely often}) = 0$, i.e., the event $\{s_i = 1\}$ occurs for at most finitely many i with probability one. It follows that y_t is eventually a stationary ARMA process, even though it may be nonstationary during any finite time period.

When $\alpha_1 \neq 0$, $\mathbb{E}(y_t) = \alpha_0 t + \alpha_1 \sum_{i=1}^t \mathbb{P}(s_i = 1) + \alpha_1 \sum_{i=1}^t \psi_i^* [1 - \mathbb{P}(s_i = 1)]$. It is, however, more cumbersome to calculate $\mathrm{var}(y_t)$ because the components of (3) are no longer mutually uncorrelated. To see this, first note that

$$\begin{aligned}\mathrm{cov}(s_t, s_{t-j}) &= \mathbb{P}(s_t = 1 \text{ and } s_{t-j} = 1) - \mathbb{P}(s_t = 1)\,\mathbb{P}(s_{t-j} = 1) \\ &= \mathbb{P}(s_{t-j} = 1)[\mathbb{P}(s_t = 1 \mid s_{t-j} = 1) - \mathbb{P}(s_t = 1)].\end{aligned}$$

By the Markovian property, $\mathbb{P}(s_t = 1 \mid s_{t-j} = 1)$ is the (2,2) element of the jth power of the transition matrix for $j \geq 1$. Thus, s_t are serially correlated, and s_t and $(1 - s_\tau)$ are not mutually uncorrelated. Moreover, $\mathrm{cov}(s_t, s_{t-j}\varepsilon_{t-j})$ are non-zero when $j \geq 1$. To ease our analysis of $\mathrm{var}(y_t)$, we assume for the time being that $\mathbb{E}(\varepsilon_t \mid S^\tau) = 0$ for $\tau > t$. Then, $\mathrm{cov}(s_t, s_{t-j}\varepsilon_{t-j}) = 0$ so that

$$\begin{aligned}\mathrm{var}(y_t) &= \alpha_1^2 \,\mathrm{var}\Big(\sum_{i=1}^t s_i\Big) + 2\alpha_1^2 \,\mathrm{cov}\Big(\sum_{i=1}^t s_i, \sum_{i=1}^t \psi_i^*(1 - s_i)\Big) \\ &\quad + \alpha_1^2 \,\mathrm{var}\Big(\sum_{i=1}^t \psi_i^*(1 - s_i)\Big) + \sigma_\varepsilon^2 \sum_{i=1}^t \mathbb{P}(s_i = 1) + \sigma_\varepsilon^2 \sum_{i=1}^t (\psi_i^*)^2 [1 - \mathbb{P}(s_i = 1)].\end{aligned}$$

Note that $\mathrm{var}(\sum_{i=1}^t s_i)$ is bounded when the partial sum $\sum_{i=1}^t \mathbb{P}(s_i = 1)$ converges; the same conclusion also holds for the second and third terms on the right-hand side of the prior equation. This shows that the stationarity of y_t still depends essentially on the convergence of $\sum_{i=1}^t \mathbb{P}(s_i = 1)$, as in the case that $\alpha_1 = 0$.

From Eq. (3), it is clear that the impulse response function of the proposed model depends on the realization of s_i. Let $\mathcal{F}^t$ denote the information set up to time t and

$$\delta_t \equiv \lim_{k \to \infty} \frac{\partial\, \mathbb{E}(y_{t+k} | \mathcal{F}^t)}{\partial \varepsilon_t}$$

denote the long-run effect of ε_t on the optimal forecast of y_{t+k}. Then, $\delta_t = s_t$, which is one or zero and changes from time to time. Recall that the long-run effect δ_t is one for a random walk and zero for a weakly stationary process. Note also that for the simplest STOPBREAK process (6),

$$\delta_t = q_t + \frac{\partial q_t}{\partial \varepsilon_t}\, \varepsilon_t = q_t + \frac{2\gamma\varepsilon_t^2}{(\gamma + \varepsilon_t^2)^2},$$

where $q_t = \varepsilon_t^2/(\gamma + \varepsilon_t^2)$. When ε_t has a continuous distribution, ε_t is zero with probability zero. Thus, δ_t is positive with probability one, showing that the innovations of the STOPBREAK process must have a permanent effect.

As for the property of $z_t = (1 - B)y_t$, we consider the case that $\alpha_1 = 0$. Write z_t as $z_t = z_{1,t} + z_{2,t}$, where $z_{1,t} = \alpha_0 + s_t\varepsilon_t$ and $z_{2,t} = (1 - B)\Psi(B)^{-1}\Phi(B)(1 - s_t)\varepsilon_t$. Then z_t

is the sum of two uncorrelated components. Let g_{z_1} and g_{z_2} denote the autocovariance generating functions of z_1 and z_2, respectively. When $\mathbb{P}(s_t = 1)$ is a constant π_0, it is now easy to see that the autocovariance generating function of z_t is

$$\begin{aligned} g(a) &= g_{z_1}(a) + g_{z_2}(a) \\ &= \pi_0\,\sigma_\varepsilon^2 + (1-\pi_0)(1-a)(1-a^{-1})\Psi(a)^{-1}\Psi(a^{-1})^{-1}\Phi(a)\Phi(a^{-1})\sigma_\varepsilon^2. \end{aligned}$$

For example, for the CD(1; 1, 0) model considered in Section 2.2, z_t has mean zero and

$$\mathrm{var}(z_t) = \sigma_\varepsilon^2 + (1-\pi_0)(1-\psi_1^{2t-2})\frac{1-\psi_1}{1+\psi_1}\sigma_\varepsilon^2.$$

The autocovariances are

$$\begin{aligned} \mathrm{cov}(z_t, z_{t-1}) &= -(1-\pi_0)(1+\psi_1^{2t-1})\frac{1-\psi_1}{1+\psi_1}\sigma_v^2, \\ \mathrm{cov}(z_t, z_{t-i}) &= \psi^{i-1}\,\mathrm{cov}(z_t, z_{t-1}), \quad i \geq 2. \end{aligned}$$

These autocovariances depend only on i as t goes to infinity, showing that z_t is asymptotically a covariance stationary process. When $\pi_0 = 1$, z_t is simply a white noise. Note also that these autocovariances agree with those of a non-invertible ARMA(1,1) process if and only if $\pi_0 = 0$. Thus, z_t would be invertible provided that $\pi_0 > 0$.

4 Model Estimation

The CD(1; m, n) model can be estimated either by the maximum likelihood method or Markov chain Monte Carlo method. We adopt the former in this paper. The derivation below is similar, but not identical, to those of Hamilton (1989, 1994) and Kim (1994) because y_t depends only on the past (but not the current) state variables.

From Eq. (4) we see that the past $r+1$ state variables affect y_t. Following Hamilton (1994), we define the new state variable $s_{t-1}^* = 1, 2, \ldots, 2^{r+1}$ such that each of these values represents a particular combination of the realizations of $(s_{t-1}, \ldots, s_{t-r-1})$. For example, when $r = 2$,

$$\begin{aligned} s_{t-1}^* &= 1 \text{ if } s_{t-1} = s_{t-2} = s_{t-3} = 0, \\ s_{t-1}^* &= 2 \text{ if } s_{t-1} = 0, s_{t-2} = 0, \text{ and } s_{t-3} = 1, \\ s_{t-1}^* &= 3 \text{ if } s_{t-1} = 0, s_{t-2} = 1, \text{ and } s_{t-3} = 0, \\ &\;\;\vdots \\ s_{t-1}^* &= 8 \text{ if } s_{t-1} = s_{t-2} = s_{t-3} = 1. \end{aligned}$$

It is easy to show that s_t^* also forms a first-order Markov chain with the transition matrix $\boldsymbol{P}^*$. This transition matrix can be expressed as

$$\boldsymbol{P}^* = \begin{bmatrix} \boldsymbol{P}_{00} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}_{10} \\ \boldsymbol{P}_{01} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}_{11} \end{bmatrix},$$

with $\boldsymbol{P}_{ji}$ $(j, i = 0, 1)$ being a $2^{r-1} \times 2^r$ block diagonal matrix given by

$$\boldsymbol{P}_{ji} = \begin{bmatrix} p_{ji} & p_{ji} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p_{ji} & p_{ji} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_{ji} & p_{ji} \end{bmatrix}.$$

Also let $\boldsymbol{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-r-1})'$ and for $s_{t-1}^* = \ell$, $\ell = 1, 2, \ldots, 2^{r+1}$, let

$$\boldsymbol{\xi}_{t-1,\ell} = (\xi_{1,s_{t-1}}, \xi_{2,s_{t-2}}, \ldots, \xi_{r+1,s_{t-r-1}})',$$

where the realizations of $s_{t-1}, \ldots, s_{t-r-1}$ are such that $s_{t-1}^* = \ell$. Then,

$$\boldsymbol{\xi}_{t-1,\ell}' \boldsymbol{\varepsilon}_{t-1} = \sum_{j=1}^{r+1} \xi_{j,s_{t-j}} \varepsilon_{t-j}.$$

When $r = 2$ and $\ell = 3$, for example, the realization of $(s_{t-1}, s_{t-2}, s_{t-3})$ is $(0, 1, 0)$ so that

$$\boldsymbol{\xi}_{t-1,3}' \boldsymbol{\varepsilon}_{t-1} = -(1 + \varphi_1)\varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} + \varphi_3 \varepsilon_{t-3}.$$

We first discuss the optimal forecasts of the state variable s_t based on the information up to time t: $\mathcal{Z}^t = \{z_t, z_{t-1}, \ldots, z_1\}$, where $z_t = (1 - B)y_t$. This amounts to calculating the filtering probabilities $\mathbb{P}(s_t \mid \mathcal{Z}^t)$. Under the normality assumption, the density of z_t conditional on $s_{t-1}^* = \ell$ and $\mathcal{Z}^{t-1}$ is

$$\begin{aligned} & f(z_t \mid s_{t-1}^* = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) \\ & \quad = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ \frac{-[z_t - \sum_{j=1}^m \psi_j(\alpha_0 + z_{t-j}) - \boldsymbol{\xi}_{t-1,\ell}'(\alpha_1 \mathbf{1} + \boldsymbol{\varepsilon}_{t-1})]^2}{2\sigma_\varepsilon^2} \right\}, \end{aligned} \tag{8}$$

where $\mathbf{1}$ is the vector of ones, $\ell = 1, 2, \cdots, 2^{r+1}$ and

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \psi_1, \ldots, \psi_m, \varphi_1, \ldots, \varphi_n, \sigma_\varepsilon, p_{00}, p_{11})'.$$

Although the innovations ε_t depend on s^*_{t-1}, $t = m+1, \ldots, T$, we, in accordance with Gray (1996), compute ε_t ($t = m+1, \ldots, T$) as

$$\begin{aligned} \varepsilon_t &= z_t - \mathbb{E}(z_t \mid \mathcal{Z}^{t-1}) \\ &= z_t - \sum_{j=1}^{m} \psi_j(\alpha_0 + z_{t-j}) - \sum_{\ell=1}^{2^{r+1}} \mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, \boldsymbol{\xi}'_{t-1,\ell}(\alpha_1 \mathbf{1} + \boldsymbol{\varepsilon}_{t-1}), \end{aligned} \tag{9}$$

with the initial values $\varepsilon_m, \ldots, \varepsilon_1$ being zero, where $\mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ is the filtering probability of $s^*_{t-1} = \ell$ based on the information up to time $t-1$.

Given $\mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$, the density of z_t conditional on $\mathcal{Z}^{t-1}$ alone can be obtained via (8) as

$$f(z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}) = \sum_{\ell=1}^{2^{r+1}} \mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, f(z_t \mid s^*_{t-1} = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta}). \tag{10}$$

To compute $\mathbb{P}(s^*_t = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta})$, note that

$$\mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta}) = \frac{\mathbb{P}(s^*_{t-1} = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, f(z_t \mid s^*_{t-1} = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta})}{f(z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})}. \tag{11}$$

We also assume that the (j, i)th element of $\boldsymbol{P}^*$ is such that

$$p^*_{ji} = \mathbb{P}(s^*_t = i \mid s^*_{t-1} = j) = \mathbb{P}(s^*_t = i \mid s^*_{t-1} = j, \mathcal{Z}^t);$$

the second equality would hold if $\{s_t\}$ and $\{\varepsilon_t\}$ are independent. These in turn yield

$$\begin{aligned} \mathbb{P}(s^*_t = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta}) &= \sum_{j=1}^{2^{r+1}} \mathbb{P}(s^*_{t-1} = j \mid \mathcal{Z}^t; \boldsymbol{\theta})\ \mathbb{P}(s^*_t = \ell \mid s^*_{t-1} = j, \mathcal{Z}^t; \boldsymbol{\theta}) \\ &= \sum_{j=1}^{2^{r+1}} p^*_{j\ell}\ \mathbb{P}(s^*_{t-1} = j \mid \mathcal{Z}^t; \boldsymbol{\theta}). \end{aligned} \tag{12}$$

Thus, with the initial value $\mathbb{P}(s^*_m \mid \mathcal{Z}^m; \boldsymbol{\theta})$, we can iterate the equations (8)–(12) to obtain $\mathbb{P}(s^*_t = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta})$ for $t = m+1, \ldots, T$. Then for each t, the desired filtering probability is

$$\mathbb{P}(s_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta}) = \sum \mathbb{P}(s^*_t = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta}),$$

and $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}) = 1 - \mathbb{P}(s_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta})$, where the summation is taken over all ℓ that associated with $s_t = 1$.

From the recursions above we also obtain the quasi-log-likelihood function

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln f(z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}),$$

from which the quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$ may be found via a numerical-search algorithm. The estimation program is written in GAUSS which employs the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm. Following Hamilton (1989, 1994), we set the initial value $\mathbb{P}(s_m^* \mid \mathcal{Z}^m; \boldsymbol{\theta})$ to its limiting unconditional counterpart: the $(2^{r+1}+1)$th column of the matrix $(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} - \boldsymbol{P}^* \\ \boldsymbol{1}' \end{bmatrix},$$

$\boldsymbol{I}$ is the identity matrix and $\boldsymbol{1}$ is the 2^{r+1}-dimensional vector of ones; see Hamilton (1994, p. 684) for details.

We also follow the approach of Kim (1994) to calculate the smoothing probabilities $\mathbb{P}(s_t \mid \mathcal{Z}^T)$ for $t \leq T$, which are the optimal forecasts of s_t based on all the information in the sample. Observe that

$$\begin{aligned} &\mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^T) \\ &\qquad = \frac{\mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^{t+1})\ \mathbb{P}(z_T, \ldots, z_{t+2} \mid s_t^* = \ell, s_{t+1}^* = j, \mathcal{Z}^{t+1})}{\mathbb{P}(z_T, \ldots, z_{t+2} \mid s_{t+1}^* = j, \mathcal{Z}^{t+1})}. \end{aligned}$$

In the current context,

$$\mathbb{P}(z_T, \cdots, z_{t+2} \mid s_t^* = \ell, s_{t+1}^* = j, \mathcal{Z}^{t+1}) = \mathbb{P}(z_T, \cdots, z_{t+2} \mid s_{t+1}^* = j, \mathcal{Z}^{t+1}),$$

so that $\mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^T) = \mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^{t+1})$. Note, however, that the condition above does not hold in Kim (1994). It follows that

$$\begin{aligned} &\mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^T) \\ &\quad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^{t+1}) \\ &\quad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \frac{\mathbb{P}(s_{t+1}^* = j \mid s_t^* = \ell, \mathcal{Z}^{t+1})\ \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^{t+1})}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})} \qquad (13) \\ &\quad = \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^{t+1}) \sum_{j=1}^{2^{r+1}} \frac{p_{\ell j}^*\ \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})}. \end{aligned}$$

Using the filtering probability $\mathbb{P}(s_T^* = \ell \mid \mathcal{Z}^T)$ as the initial value we can iterate the equations (11), (12) and (13) backward for $t = T-1, \cdots, p+1$. Consequently, for each t, the desired smoothing probability is

$$\mathbb{P}(s_t = 1 \mid \mathcal{Z}^T) = \sum \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^T),$$

and $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T) = 1 - \mathbb{P}(s_t = 1 \mid \mathcal{Z}^T)$, where the summation is taken over all ℓ that associated with $s_t = 1$. Similar to the filtering probabilities derived earlier, the smoothing probabilities are also functions of $\boldsymbol{\theta}$. Plugging the quasi-maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ into these probabilities we obtain the estimated values for the filtering and smoothing probabilities.

5 Hypothesis Testing

As it is widely accepted that many economic and financial time series contain a unit root, an interesting hypothesis is whether the data follow a random walk. This amounts to testing $p_{11} = 1$. Under this null hypothesis, the stationary component does not enter the model so that its parameters (those of $\Psi(B)$ and $\Phi(B)$) are not identified. In this case, standard likelihood-based tests, such as the Wald, LM, and likelihood ratio tests, are not applicable; see Davies (1977, 1987) and Hansen (1996). The problem that certain parameters are not identified under the null hypothesis also arises in other regime switching models. In contrast with Hamilton's model, whether $\alpha_1 = 0$ is not of primary concern here. Once we exclude the possibility that the process is a random walk, hypothesis testing on other parameters is standard and can be done using likelihood-based tests. Therefore, we focus on the null hypothesis of $p_{11} = 1$.

Since the data are a random walk when $p_{11} = 1$, it is of interest to study the performance of the Dickey-Fuller (DF) test of Dickey and Fuller (1979). We simulate y_t according to Eq. (3) with $\alpha_0 = \alpha_1 = 0$, $\sigma_\varepsilon^2 = 1$, $\Psi(B) = 1 - 0.5B$, $\Phi(B) = 1$, and various combinations of the transition probabilities p_{11} and p_{00}. In the simulations, the nominal size is 5%, sample size is 120, and the number of replications is 5000. The resulting rejection frequencies of the DF test are plotted in the left panel of Figure 2. We see that the DF test is not powerful against the alternative (3), except when p_{11} is small and p_{00} is large. For example, given $p_{11} = 0.9$, when $p_{00} = 0.8$ and 0.2, the powers are 14.9% and 7.6%, respectively; given $p_{11} = 0.1$, when $p_{00} = 0.8$ and 0.2, the powers are 55.8% and 23.7%, respectively. A detailed table of rejection frequencies is available upon request.

As shown in the right panel of Figure 2, the KPSS test of Kwiatkowski et al. (1992) is more powerful against (3), except when p_{00} is close to one. The rejection frequencies are typically around 70% when p_{11} and p_{00} are between 0.1 and 0.9. When the KPSS test rejects the null of stationarity, it is still difficult to judge whether the series being tested is a random walk or a process generated from the proposed model. More testing results are needed to support the proposed model.

In this paper, we propose using a simulation-based test of an ARIMA model against the proposed model. We first estimate an array of ARIMA$(p, 1, q)$ models and choose

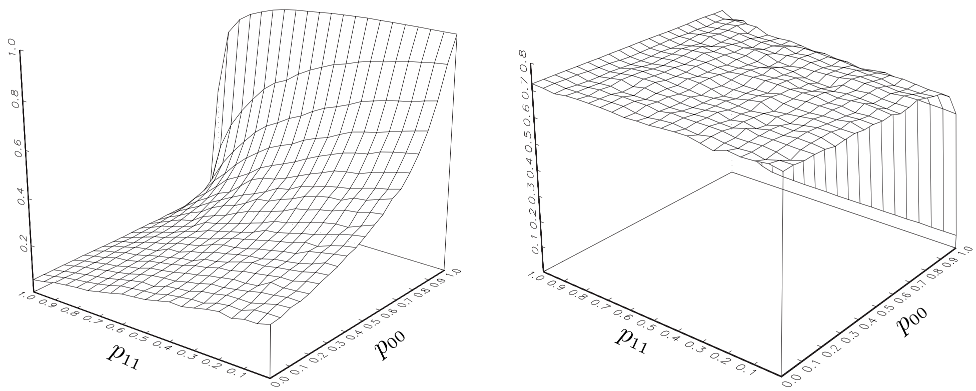


Figure 2: Empirical powers of the Dickey-Fuller test (left) and KPSS test (right).

an appropriate specification based on an information criterion (e.g., AIC or SIC). The selected model is denoted as ARIMA$(p^*, 1, q^*)$. Similarly, we also estimate an array of CD$(1; m, n)$ models and denote the selected model as CD$(1; m^*, n^*)$ and the estimated transition probability as $\hat{p}_{11}^*$. The selected ARIMA$(p^*, 1, q^*)$ model is then taken as the data generating process to generate simulated samples. For each simulated sample, we re-estimate the CD$(1; m^*, n^*)$ model and obtain an estimate of p_{11}, denoted by $\hat{p}_{11}$. Replicating this procedure many times yields a simulated distribution of $\hat{p}_{11}$. We reject the null hypothesis that the series follows the ARIMA$(p^*, 1, q^*)$ model if the p-value of $\hat{p}_{11}^*$ is small, say, less than 5%.

6 Empirical Study

To assess the empirical relevance of the proposed model, we apply the model (2) with a smooth Markov trend to U.S. real GDP. Leading models for GDP or GNP include the trend-stationary models, unit-root models, and regime switching models. For example, Blanchard (1981), Kydland and Prescott (1980) and Clark (1987) suggest that the logarithm of real GNP is trend stationary, whereas Nelson and Plosser (1982) and Campbell and Mankiw (1987) argue that real GNP contains a unit root. On the other hand, Hamilton (1989), Lam (1990), and Kim and Nelson (1999) adopt a Markov switching model to describe GNP or GDP. As the proposed model constitutes intermediate cases between these two models, it would be interesting to know if it is capable of accounting for the fluctuations of U.S. real GDP.

Our data are seasonally adjusted, quarterly U.S. real GDP from 1947:I through 1999:II with 210 observations. The data set is taken from the AREMOS databank of the Ministry of Education in Taiwan. We take log GDP as y_t and estimate an array of CD$(1; m, n)$

Table 2: Quasi-maximum likelihood estimates of the proposed model.

Parameter	Estimate	Standard error	t-statistic
α_0	-0.01166	0.00118	-9.88135^*
α_1	0.02247	0.00153	14.68627^*
ψ_1	0.39963	0.07597	5.26036^*
ψ_2	-0.63773	0.08396	-7.59564^*
φ_1	0.05715	0.07920	0.72159
φ_2	0.43084	0.08585	5.01852^*
σ_ε	0.00813	0.00141	5.76595^*
p_{00}	0.80601	0.06084	
p_{11}	0.95053	0.01682	
AR roots: $0.19981 \pm 0.77318i$		MA roots: $-0.02857 \pm 0.65576i$	
Log-Likelihood=668.99		AIC=-1319.98	SIC=-1289.90

Note: t-statistics with an asterisk are significant at the 5% level.

models with a Markov trend and $0 \leq m, n \leq 4$. The parameters,

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \psi_1, \ldots, \psi_m, \varphi_1, \ldots, \varphi_n, \sigma_\varepsilon, p_{00}, p_{11})',$$

are estimated using the algorithm described in Section 4. This algorithm is initialized by a broad range of random initial values. The covariance matrix of $\boldsymbol{\theta}$ is $-H(\hat{\boldsymbol{\theta}})^{-1}$, where $H(\hat{\boldsymbol{\theta}})$ is the Hessian matrix of the log-likelihood function evaluated at the QMLE $\hat{\boldsymbol{\theta}}$.

Among all the models considered, both AIC and SIC select the CD$(1; 2, 2)$ model. The estimation results are summarized in Table 2; the estimated transition probabilities are $\hat{p}_{11}^* \approx 0.95$ and $\hat{p}_{00}^* \approx 0.80$. We first apply the simulation approach in Section 5 to test p_{11}. We estimate an array of ARIMA$(p, 1, q)$ models with p and q no greater than 4; both AIC and SIC select the ARIMA$(1, 1, 0)$ model:

$$\Delta y_t = 0.008 + 0.341 \Delta y_{t-1} + u_t, \tag{14}$$

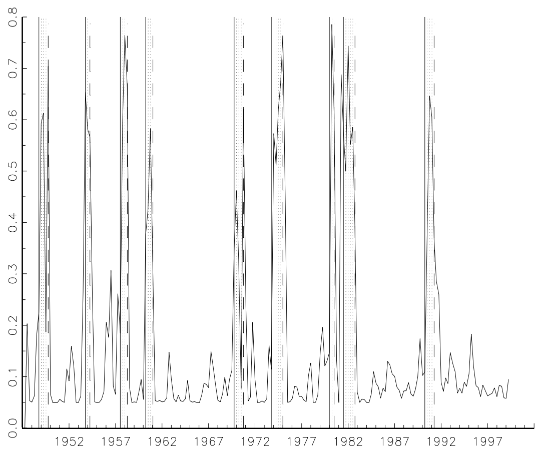
with $\sigma_u = 0.0098$, where $\Delta y_t = y_t - y_{t-1}$. We then re-estimate the CD$(1; 2, 2)$ model using the data generated from Eq. (14) and obtain $\hat{p}_{11}$. With 1000 replications we have a simulated distribution of $\hat{p}_{11}$. The p-value of $\hat{p}_{11}^* = 0.95$ based on this simulated distribution is 0.037 and hence rejects the model in Eq. (14).

In addition, we also take the random walk model as the null hypothesis and notice that $z_t = y_t - y_{t-1}$ should be uncorrelated with all past z_{t-i} under the null. We then regress z_t on $z_{t-1}, \ldots, z_{t-k}$ for $k = 1, \cdots, 4$ with a constant term and check the joint significance of the coefficients of z_{t-i} using the Wald test. Note that there will be no unidentified nuisance parameters under this framework. A similar approach is also taken by Tsay (1989) to test for threshold autoregressive models. The resulting Wald statistics are 90.96, 83.45, 93.49 and 114.32, respectively, which are all significant at the 1% level under $\chi^2(k)$ distribution. We thus reject the null hypothesis that the data series is a pure random walk.

Given that the data are neither an ARIMA process nor a pure random walk, we now proceed to test other hypotheses by the Wald test. In particular, as discussed in Engel and Hamilton (1990), the proposed model would be a simple mixture model if the probability that $s_t = 0$ or 1 is independent of the previous state. This amounts to testing the null hypothesis $p_{00} + p_{11} = 1$. The Wald statistic of this hypothesis is 148.794 which is also significant at the 1% level under the $\chi^2(1)$ distribution. The rejection of this hypothesis may justify our Markovian specification of the state variable.

In Figure 3 we plot the estimated filtering and smoothing probabilities of $s_t = 0$ in the left and right figures, respectively. The shaded areas denote the recession periods identified by NBER, where the solid (dashed) lines label the peaks (troughs). We find that there are 28 periods (about 14% of the sample) with the filtering probability $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}) > 0.5$ and 35 periods (about 17% of the sample) with the smoothing probability $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta}) > 0.5$. This shows that unit-root nonstationarity is more likely to prevail in more than 80 percent of the sample periods, yet stationarity dominates in the remaining periods. We also observe that the nonstationarity (stationarity) periods match the NBER dating of expansions (recessions) very closely. Hence, the innovations in expansion (recession) are more likely to have a permanent (transitory) effect. These results suggest the following features of real GDP. First, the nonstationary characteristic and permanent shocks do not appear all the time, in contrast with the result of unit-root models. Second, permanent shocks occur more frequently than the assertion of trend-break models, cf. Perron (1989) and Balke and Fomby (1991). Third, the shocks in the expansion periods generate nonstationary pattern and hence are more persistent than those in the recession periods. This is compatible with the conclusion of Beaudry and Koop (1993) who found that positive shocks to GDP are more persistent than negative shocks.

From Table 2 we see that the estimated quarterly growth rates of U.S. real GDP are $\alpha_0 = -1.16\%$ during the state of transitory shocks (recessions) and $(\alpha_0 + \alpha_1) = 1.08\%$ during the state of permanent shocks (expansions). The expected durations of recession and expansion can be calculated from the transition probabilities: $1/(1-0.8) = 5$ quarters for recession and $1/(1-0.95) = 20$ quarters for expansion. According to NBER



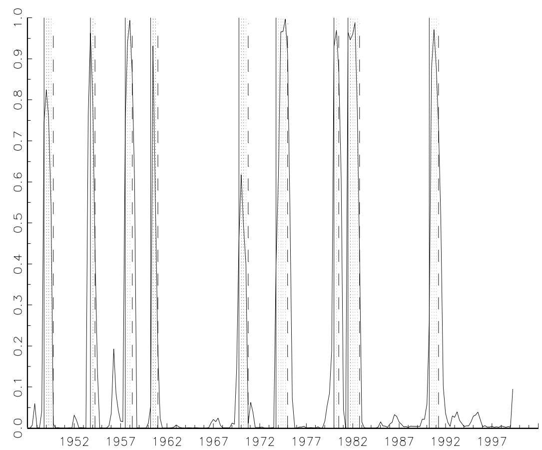


Figure 3: Estimated filtering and smoothing probabilities of $s_t = 0$ for U.S. quarterly real GDP from 1947.I to 1999.II.

dating, the average growth rates for recession and expansion are, respectively, -0.38% and 0.93%, and the average durations are, respectively, 3.6 and 19.3 quarters. Comparing with Hamilton (1989), our results lead to a longer expected duration for expansion and "deeper" recessions. We also apply the model of Hamilton (1989) to this data set. The estimation Gauss program used is taken from C. R. Nelson's web site and is initialized by 100 initial values.[1] Unfortunately, the estimation results fail to provide reasonable parameter estimates for the data; Kim and Nelson (1999, p. 78) also reported a similar problem when a different data set was used. This may not be very surprising because Boldin (1996) noticed that Hamilton's result is sensitive to the sample period.

We compute the expected trend line (smooth Markov trend) as

$$\hat{\alpha}_0\, t + \hat{\alpha}_1 \sum_{i=1}^{t} \mathrm{IP}(s_i = 1 \mid \mathcal{Z}^T; \hat{\boldsymbol{\theta}}) + \hat{\alpha}_1 \hat{\Psi}(B)^{-1} \hat{\Phi}(B)(1 - \mathrm{IP}(s_t = 1 \mid \mathcal{Z}^T; \hat{\boldsymbol{\theta}})),$$

where the stochastic trend component is weighted by the estimated smoothing probabilities. This trend line captures the trend behavior of log(GDP) quite well, as can be seen from Figure 4.

Examining the filtering and smoothing probabilities more carefully, we find that there are 14 periods with $\mathrm{IP}(s_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}) < 0.5$ but $\mathrm{IP}(s_t = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta}) > 0.5$. That is, the

[1] C. R. Nelson's web site is www.econ.washington.edu/user/cnelson/SSMARKOV.htm; the program is HMT4_KIM.OPT.

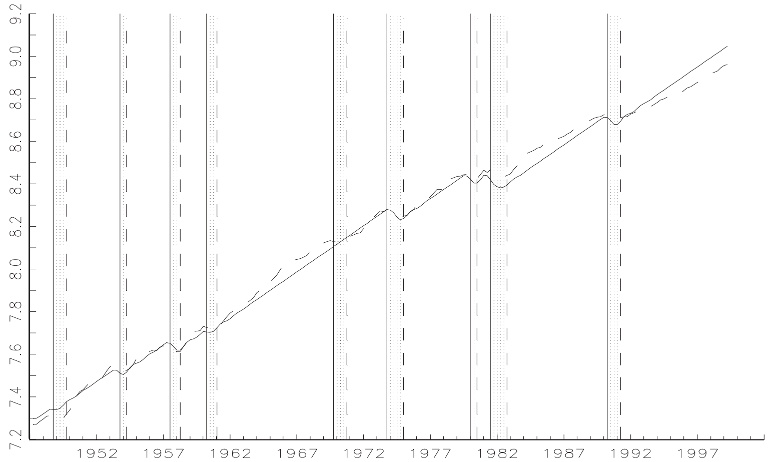


Figure 4: The expected trend line in U.S. real GDP.

likelihood of $s_t = 0$ may differ significantly when a information set changes. This is not unreasonable. A shock may seem very significant at time t, but its significance may subsequently diminish when more information is taken into account. To illustrate this point, we calculate the smoothing probabilities based on the filtration $\{\mathcal{Z}^\tau, \tau \geq t\}$, i.e., $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^\tau; \boldsymbol{\theta})$ for $\tau = t, t+1, \cdots, T$. Figure 5 shows the dynamic pattern of $\mathbb{P}(s_{1990:\text{III}} = 0 \mid \mathcal{Z}^\tau; \boldsymbol{\theta})$ for $\tau =$ 1990:III, 1990:IV, ..., 1999:II. This figure clearly shows how 1990:III changes from the state one to the state zero as information set expands.

7 Conclusions

In this paper we propose a class of component-driven models with several interesting features. First, it admits both deterministic and stochastic trends. Second, it can describe both stationary and nonstationary characteristics over different time periods. The effects of corresponding innovations thus may alternate from time to time. Third, it allows for endogenous breaks in the deterministic trend function such that different trending patterns are directly linked to the shocks with distinct effects. Fourth, when there are trend breaks, the transitions between trend segments are smooth. This class of models can be viewed as intermediate cases between trend-stationary and unit-root models and is able to accommodate both trend-reverting and trend-perverting behavior.

The empirical application of the proposed model to U.S. real GDP suggests that it is a useful analytical tool. It is shown that unit-root nonstationarity is more likely to prevail in more than 80 percent of the sample periods and that these periods match closely the expansion periods dated by NBER. Thus, the shocks in expansions are more likely to be permanent. This result differs from that of unit-root (trend-stationary) models in that

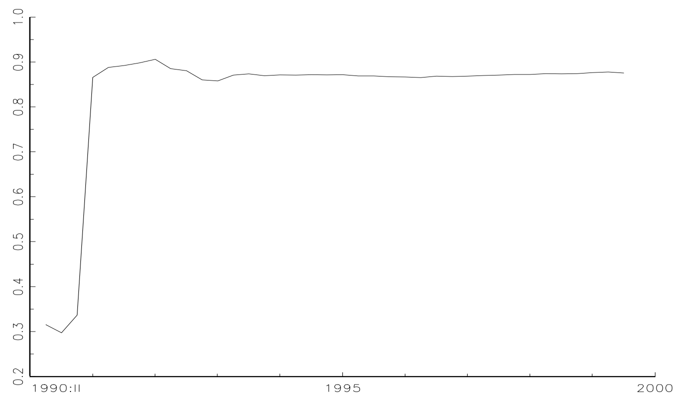


Figure 5: Dynamic path of $\mathbb{P}(s_{1990:\text{III}} = 0 \mid \mathcal{Z}^{\tau}; \boldsymbol{\theta})$ for U.S. quarterly real GDP data from 1947.I to 1999.II.

the shocks may not always be permanent (transitory). The fact that permanent shocks occur quite frequently is also different from the assertion of trend-break models, such as those of Perron (1989) and Balke and Fomby (1991). The proposed model may therefore serve as an alternative for modeling economic time series.

Reference

Balke, N. and T. B. Fomby (1991). Shifting trends, segmented trends, and infrequent permanent shocks, *Journal of Monetary Economics*, **28**, 61–85.

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

Blanchard, O. J. (1981). What is left of the multiplier-accelerator?, *American Economic Review Proceedings*, **71**, 150–154.

Boldin, M. D. (1996). A check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle, *Studies in Nonlinear Dynamics and Econometrics*, **1**, 35–46.

Bradley, M. D. and D. W. Jansen (1997). Nonlinear business cycle dynamics: Cross-country evidence on the persistence of aggregate shocks, *Economic Inquiry*, **35**, 495–509.

Campbell, J. H. and N. G. Mankiw (1987). Are output fluctuations transitory? *Quarterly Journal of Economics,* **102**, 857–880.

Clark, P. K. (1987). The cyclical component of U.S. economic activity, *Quarterly Journal of Economics*, **102**, 798–814.

Cochrane, J. H. (1988). How big is the random walk in GNP? *Journal of Political Economy*, **96**, 893–920.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **74**, 33–43.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Elwood, S. K. (1998). Is the persistence of shocks to output asymmetric?, *Journal of Monetary Economics*, **41**, 411–426.

Engel, C. and J. D. Hamilton (1990). Long swings in the dollar: Are they in the data and do markets know it?, *American Economic Review*, **80**, 689–713.

Engle, R. F. and A. D. Smith (1999). Stochastic permanent breaks, *The Review of*

Economics and Statistics, **25**, 475–511.

Evans, M. and P. Wachtel (1993), Inflation regimes and the sources of inflation uncertainty, *Journal of Money, Credit, and Banking*, **25**, 475–511.

Granger, C. and N. R. Swanson (1997). An introduction to stochastic unit-root processes, *Journal of Econometrics,* **80**, 35–62.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, **42**, 27–62.

Hamilton, J. D. (1989). A new approach to the Economic Analysis of Nonstationary time series and the business cycle, *Econometrica*, **57**, 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.

Hess, G. D. and S. Iwata (1997). Asymmetric persistence in GDP? A deeper look at depth, *Journal of Monetary Economics*, **40**, 535–554.

Kim, C. J. (1994). Dynamic linear models with Markov-switching, *Journal of Econometrics*, **60**, 1–22.

Kim, C..-J. and C. R. Nelson (1999). *State Space Models with Regime Switching, Classical and Gibbs Sampling Approaches with Applications*, Cambridge, MA: MIT Press.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**, 159–178.

Kydland, F. E. and E. C. Prescott (1980). A competitive theory of fluctuations and the feasibility and desirability of stabilization policy, in S. Fischer, ed., *Rational Expectations and Economic Policy*, pp. 169–198, Chicago: University of Chicago Press.

Lam, P. S. (1990). The Hamilton model with a general autoregressive component *Journal of Monetary Economics,* **26**, 409–432.

McCabe, B. P. M. and A. R. Tremayne (1995). Testing a time series for difference stationarity, *The Annals of Statistics*, **23**, 1015–1028.

McCulloch, R. E. and R. S. Tsay (1994). Bayesian inference of trend- and difference-stationarity, *Econometric Theory*, **10**, 596–608.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics*, **10**, 139–162.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, **57**, 1361–1401.

Rappoport, P. and L. Richelin (1989). Segmented trends and non-stationary time series, *Economic Journal* (supplement 1989), 168–177.

Tong, H. (1990). *Nonlinear Time Series*, Oxford, U.K.: Oxford University Press.

Tsay, R. S.̃(1989). Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association*, **84**, 231–240.