The Semi-Nonstationary Process: Model and Empirical Evidence

Chung-Ming Kuan

Institute of Economics
Academia Sinica

Yue-Lieh Huang

Department of Economics
National Taiwan University

Ruey S. Tsay

Graduate School of Business
University of Chicago

This version: September 21, 2001

Author for correspondence: Chung-Ming Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; ckuan@ieas.econ.sinica.edu.tw; phone: +886 2 2782–2791 ext. 646, fax: +886 2 2651–7428.

† We thank Y.-T. Chen, R. Engle, C. Granger, C.-C. Hsu, C.-S. Mao 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 H-001-071). 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 class of models in which the effect of an innovation may be permanent or transitory, depending on an unobservable state variable that follows a first order Markov chain. This model can describe stationary and non-stationary characteristics at different time periods. It also permits shifts in the deterministic trend such that different trending patterns are associated with innovations having distinct effects. The process generated from the proposed model is referred to as the *semi-nonstationary* process. We first show some properties of this process and derive an estimation algorithm. A simulation-based test is then proposed to distinguish between the proposed model and a random walk. In empirical application, we apply the proposed model to U.S. real GDP and find that 83 percent of the shocks to real GDP are likely to have permanent effects. Moreover, the periods that are likely to be in the state of permanent (transitory) shocks match closely to the expansion (recession) periods identified by NBER.

Keywords: Markov trend, permanent shock, regime switching, semi-nonstationary process, transitory shock, trend stationarity, unit root

JEL Classification: C22. C51

1 Introduction

In analyzing the behavior of economic time series, it is important to determine whether the underlying random shocks (innovations) of an economic system have permanent effects. While the impact of a transitory shock on the distant future eventually dies out, that of a permanent shock remains constant through time. Thus, transitory innovations induce only short-run fluctuations, and permanent innovations are able to influence the long-run equilibrium level. This topic has attracted much attention in the literature. For example, persistent output fluctuations were examined in different ways by Blanchard (1981), Beveridge and Nelson (1981), Nelson and Plosser (1982), Watson (1986), Campbell and Mankiw (1987), Clark (1987), Cochrane (1988), Hamilton (1989), Lam (1990), and Beaudry and Koop (1993), among many others.

Based on the dichotomous view of random shocks, the commonly used time-series models can be classified into two categories — namely, the models whose innovations all have permanent or transitory effects. In particular, the innovations of a trend stationary model result in temporary deviations from the trend line, whereas those of a unit-root model all have a permanent effect. It is, however, conceivable that the shocks of a system in the periods of no significance are fundamentally different from those in turbulent periods. 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). Thus, it may not be necessary to restrict *a priori* that all innovations must have the same effect. From an econometric point of view, such a restriction simplifies the model structure but limits the dynamic patterns that can be described by the existing models. These considerations motivate us to propose models that are able to accommodate both permanent and transitory shocks.

Researchers have proposed various approaches to incorporating different shocks into one model (e.g., Evans and Wachtel, 1993; McCabe and Tremayne, 1995; Granger and Swanson, 1997). In the Evans-Wachtel model, a process is driven by two sets of innovations, and there is a switching mechanism to determine which set should come into force. A problem with this model is that it is not easy to explain why one set of innovations dominates in some periods but plays absolutely no role in the other periods. The random-coefficient models of McCabe and Tremayne (1995) and Granger and Swanson (1997) are also difficult to interpret when the coefficient switches between the stable and unstable regions. On the other hand, Perron (1989) argues that only the events that change the deterministic trend have a permanent effect and that such events occur infrequently; see also Rappaport and Reichlin (1989) and Balke and Fomby (1991). Note, however, that for trend-break models, it is the exogenous event that has permanent effects; all the innovations are still transitory.

In this paper, we propose a class of models in which the effect of an innovation may be transitory or permanent, depending on an unobservable state variable that follows a first-order Markov chain. The process generated from the proposed model contains a flexible stochastic trend and exhibits stationary and non-stationary characteristics at different time periods. When innovations are state independent, this process simply reduces to the one with a single type of innovations. Hence, we refer to such a process as the *semi-nonstationry* process. These new processes constitute intermediate cases between (trend) stationary and unit-root processes. The proposed model also permits shifts in the deterministic trend while permanent shocks are prevailing. This differs from the existing models, such as the regime switching model of Hamilton (1989), in that there are smooth transitions between trend segments and that different trending features are directly linked to the innovations with distinct effects.

In empirical study, we apply the proposed model to U.S. real GDP. It is found that approximately 83 percent of the shocks have permanent effects. This result suggests that the innovations with permanent effect occur more frequently than those asserted by trend-break models. Moreover, the periods that are likely to be in the state of permanent (transitory) shocks match closely to the expansion (recession) periods identified by NBER. The expected durations of expansion and recession are also close to the averages of NBER dating. By contrast, Hamilton's regime switching model fails to provide reasonable parameter estimates for this data set. The finding that the effects of shocks are different over time is in fact consistent with that of Beaudry and Koop (1993). These empirical results thus provide us an alternative view of U.S. real GDP.

This paper is organized as follows. In Section 2 we illustrate the basic idea of the semi-nonstationary process using two simple examples and compare them with existing models. The general process is then analyzed in Section 3. We discuss an estimation algorithm and hypothesis testing in Sections 4 and 5, respectively. The empirical analysis of U.S. real GDP based on the proposed model is presented in Section 6. Section 7 concludes the paper.

2 Examples of the Semi-Nonstationary Process

The effects of innovations of most well-known econometric models are restricted to be of the same type. For an ARIMA(m, d, n) model, for instance, all the innovations have permanent effects when $d \geq 1$ and transitory effects when $d = 0$. Similarly, trend-stationary processes are trend reverting and involve only transitory shocks, whereas difference-stationary processes are trend perverting and have permanent shocks. To introduce the proposed models, we begin with two simple processes that allow their innovations to exhibit different effects over different time periods.

2.1 Two Simple Processes

Let $\{\upsilon_t\}$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance σ_υ^2. Suppose that the process y_t is the sum of two components — namely, $y_t = y_{1,t} + y_{0,t}$. The first simple process is such that

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

and the second simple process is such that

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + s_t \upsilon_t, \\ y_{0,t} &= (1 - s_t)\upsilon_t, -\varphi_1(1 - s_{t-1})\upsilon_{t-1}, \qquad |\varphi_1| < 1, \end{aligned} \tag{2}$$

where s_t is the unobserved state variable at time t that assumes the value one or zero. In these processes, the component $y_{1,t}$ is of a random-walk type but accumulates υ_i only when $s_i = 1$, whereas $y_{0,t}$ has an AR(1) or an MA(1) structure driven by those υ_i such that $s_i = 0$. Thus, $y_{1,t}$ and $y_{0,t}$ can be viewed as the long-run and short-run components of y_t, respectively; the state variable s_t determines which component an innovation should enter. The processes (1) and (2) are in effect two intermediate cases between a random walk and a weakly stationary process.

To illustrate, Table 1 gives the moving-average representations of the two y_t processes for $t = 1, \ldots, 7$ with $\{s_1, \ldots, s_6\} = \{0, 0, 1, 1, 1, 0\}$ and $y_{i,t} = 0$ for $t \leq 0$ and $i = 1, 2$. In the table, the process on the left is based on Eq. (1), and the one on the right is due to Eq. (2). In both cases, it can be seen that the effects of some innovations last forever, but those of the other innovations die out eventually. For the process based on Eq. (1), the effects of υ_1, υ_2, and υ_6 on future y_t decay exponentially in a manner similar to that of a stationary AR(1) model, whereas the effects of υ_3, υ_4, and υ_5 are permanent and stay in the system eternally. By letting t tend to infinity, the short-run component vanishes, but the long-run component $y_{1,\infty}$ remains and includes only the innovations with permanent effect; see the last row of Table 1.

Table 1: Moving-average representations of the proposed processes.

	MA representation of Eq. (1)	MA representation of Eq. (2)
y_1	v_1	v_1
y_2	$\psi_1 v_1 + v_2$	$-\varphi_1 v_1 + v_2$
y_3	$\psi_1^2 v_1 + \psi_1 v_2 + v_3$	$0 \quad - \varphi_1 v_2 + v_3$
y_4	$\psi_1^3 v_1 + \psi_1^2 v_2 + v_3 + v_4$	$0 \quad + \quad 0 \quad + v_3 + v_4$
y_5	$\psi_1^4 v_1 + \psi_1^3 v_2 + v_3 + v_4 + v_5$	$0 \quad + \quad 0 \quad + v_3 + v_4 + v_5$
y_6	$\psi_1^5 v_1 + \psi_1^4 v_2 + v_3 + v_4 + v_5 + v_6$	$0 \quad + \quad 0 \quad + v_3 + v_4 + v_5 + v_6$
y_7	$\psi_1^6 v_1 + \psi_1^5 v_2 + v_3 + v_4 + v_5 + \psi_1 v_6 + v_7$	$0 \quad + \quad 0 \quad + v_3 + v_4 + v_5 - \varphi_1 v_6 + v_7$
	$\vdots$	$\vdots$
$y_{1,\infty}$	$0 + 0 + v_3 + v_4 + v_5 + 0 + \cdots$	$0 + 0 + v_3 + v_4 + v_5 + 0 + \ldots$

These two processes are conceptually different from that of Evans and Wachtel (1993) which allows for switching between a random walk and an AR(1) process and can be written as

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

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 problem with this process is that its switching mechanism, by construction, changes the entire process and hence all the underlying innovations. Thus, one set of innovations may prevail in some periods but does not play any role in the other periods. When a switching takes place, in particular, two consecutive observations are resulted from completely different sets of innovations. This is difficult to interpret in practice and creates drastic changes of the time path. Clearly, this problem is due to the random mechanism which changes the effects of all previous innovations. This is somewhat ironic because such a switching mechanism is supposed to affect what will happen in the future but turns out to influence only the past. By contrast, the state variable s_t of the proposed process determines only the subsequent effects of v_t and has no influence whatsoever on the previous innovations v_i with $i < t$.

Let B denote the back-shift operator. We can write the model in Eq. (1) as

$$(1 - B)(1 - \psi_1 B) y_t = (1 - \psi_1 B) s_t v_t + (1 - B)(1 - s_t) v_t, \tag{3}$$

which implies the following random-coefficient representation for y_t:

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

with

$$\xi_{s_{t-1}} = \begin{cases} \psi_1, & \text{if } s_{t-1} = 1, \\ 1, & \text{otherwise.} \end{cases}$$

Similarly, we can also write Eq. (2) as

$$(1 - B)y_t = s_t v_t + (1 - B)(1 - \varphi_1 B)(1 - s_t)v_t, \tag{4}$$

which yields the random-coefficient representation:

$$y_t = y_{t-1} + \xi_{1,s_{t-1}} v_{t-1} + \xi_{2,s_{t-2}} v_{t-2} + v_t,$$

with

$$\xi_{1,s_{t-1}} = \begin{cases} 0, & \text{if } s_{t-1} = 1, \\ -1 - \varphi_1, & \text{otherwise,} \end{cases} \qquad \xi_{2,s_{t-2}} = \begin{cases} 0, & \text{if } s_{t-2} = 1, \\ \varphi_1, & \text{otherwise.} \end{cases}$$

Note that in these (reduced-form) representations, only the MA coefficients are random, and only the previous (but not the current) state variables may affect y_t.

These two simple processes are different from the standard random coefficient processes that involve random AR coefficients only (e.g., McCabe and Tremayne, 1995). A simple AR(1) process with random coefficient is

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

where a_t are random variables, typically assumed to be exogenous. The stochastic unit-root process 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 process, such processes 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 process is the STOPBREAK process proposed by Engle and Smith (1999). The simplest STOPBREAK process is

$$y_t = \sum_{i=1}^{\infty} q_{t-i} v_{t-i} + v_t, \tag{5}$$

where $q_{t-i} = v_{t-i}^2 / (\gamma + v_{t-i}^2)$ 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 this process is not much different from a pure random walk. When υ_t has a continuous distribution, $q_t = 0$ occurs only with probability zero. Thus, this process 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.

To illustrate, we simulate the random walk, the STOPBREAK process (5) with $\gamma = 1$, and the process (1) with $\psi_1 = 0$. In our simulations, these processes are generated from the same innovations, and the process (1) 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 (1). It is easy to see that the STOPBREAK process always mimics the random walk, yet the process (1) exhibits flexible dynamic patterns.

2.2 Properties of the Simple Processes

Setting $(\upsilon_0, \upsilon_{-1}, \dots)$ to zero, we can easily derive properties of the simple processes introduced in the preceding subsection. First, by Eq. (3), we have

$$y_t = \sum_{i=1}^{t} s_i \upsilon_i + \sum_{i=1}^{t} \psi_1^{t-i}(1 - s_i)\upsilon_i. \tag{6}$$

Thus, y_t becomes a pure AR(1) process when $s_i = 0$ with probability one for all i, and it is a pure random walk when $s_i = 1$ with probability one for all i. Other than these two extreme cases, y_t behaves like an AR(1) process for certain time periods and like a random walk for the other periods, when s_i assumes different values across i. Similarly, it follows from Eq. (4) that

$$y_t = \sum_{i=1}^{t} s_i \upsilon_i + (1 - \varphi_1 B)(1 - s_t)\upsilon_t, \tag{7}$$

which differs from Eq. (6) by the second term on the right-hand side. This process may exhibit the behavior of an MA(1) process and a random walk; it is a pure MA(1) series when $s_i = 0$ with probability one for all i and a pure random walk when $s_i = 1$ with probability one for all i. As long as s_i assumes different values over time, these two processes contain a flexible stochastic trend, in the sense that only those υ_i with $s_i = 1$ enter the stochastic trend component. From Eqs. (6) and (7) we can also see that the impulse response function of the model depends on the realization of s_i. The function is a constant when $s_i = 1$ but becomes the same as that of an AR(1) or MA(1) process when $s_i = 0$.

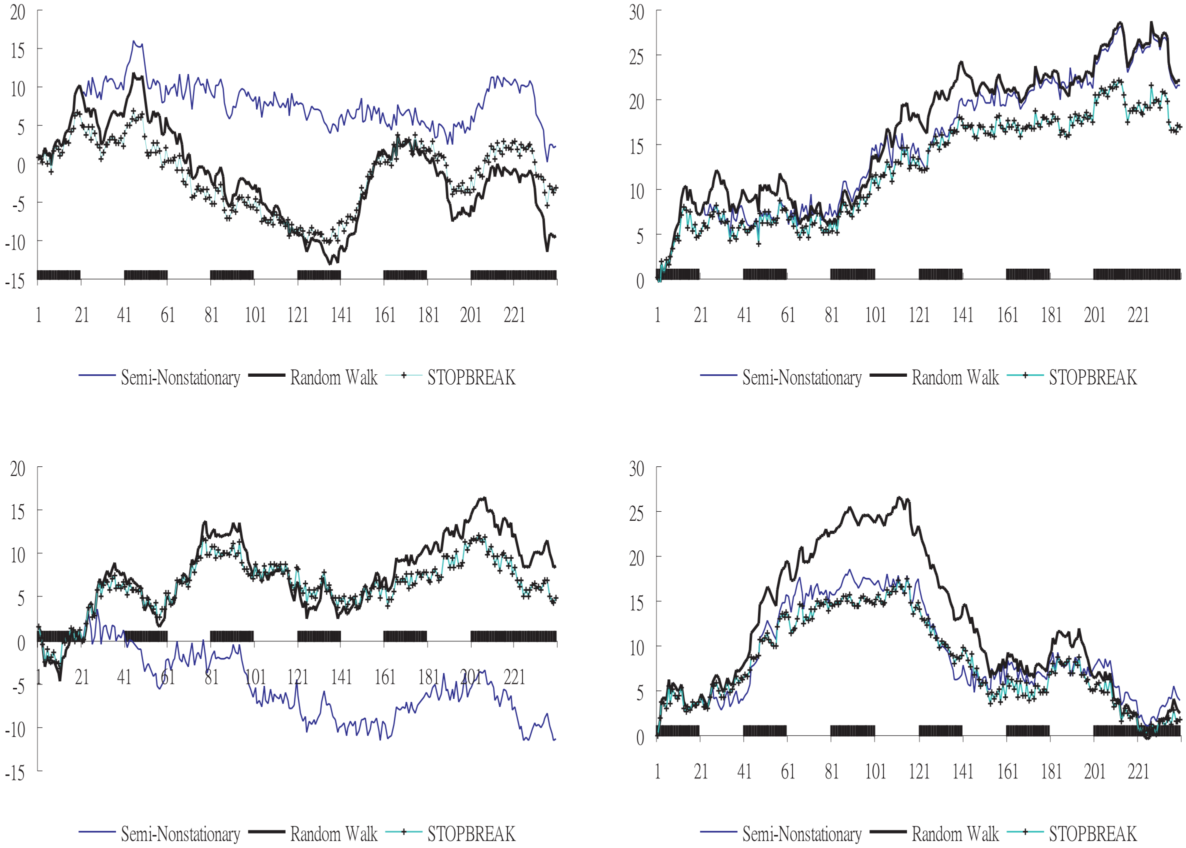

Figure 1: Simulated random walk process, STOPBREAK process and the first simple process in Eq. (1).

Let $S^t = \{s_t, s_{t-1}, \dots\}$ denote the collection of all state variables up to time t. Suppose that $\{\upsilon_t\}$ is a sequence of random variables such that $\mathbb{E}(\upsilon_t|S^t) = 0$, $\operatorname{var}(\upsilon_t|S^t) = \sigma_\upsilon^2$ and $\mathbb{E}(\upsilon_t\upsilon_{t-i}|S^t) = 0$, for all $i > 0$. By invoking the law of iterated expectations, it is easy to verify that $\{\upsilon_t\}$ is a white noise. Moreover, $\mathbb{E}(s_t\upsilon_t) = \mathbb{E}[s_t\,\mathbb{E}(\upsilon_t|S^t)] = 0$,

$$\operatorname{var}(s_t\upsilon_t) = \mathbb{E}[s_t^2\,\mathbb{E}(\upsilon_t^2|S^t)] = \sigma_\upsilon^2\,\mathbb{P}(s_t = 1),$$

and for $i > 0$, $\operatorname{cov}(s_t\upsilon_t, s_{t-i}\upsilon_{t-i}) = \mathbb{E}[s_ts_{t-i}\,\mathbb{E}(\upsilon_t\upsilon_{t-i}|S^t)] = 0$. Thus, $\{s_t\upsilon_t\}$ is also a white noise when $\mathbb{P}(s_t = 1)$ is a constant. Similarly,

$$\operatorname{cov}(s_t\upsilon_t, (1 - s_{t-i})\upsilon_{t-i}) = \mathbb{E}[s_t(1 - s_{t-i})\,\mathbb{E}(\upsilon_t\upsilon_{t-i}|S^t)] = 0,$$

for $i \geq 0$.

In view of Eq. (6), it is now straightforward to show the process (3) has the following properties: $\mathbb{E}(y_t) = 0$ and

$$\operatorname{var}(y_t) = \sigma_v^2 \sum_{i=1}^{t} \mathbb{P}(s_i = 1) + \sigma_v^2 \sum_{i=1}^{t} (\psi_1^{t-i})^2 [1 - \mathbb{P}(s_i = 1)].$$

When y_t are generated by Eq. (4), it follows from Eq. (7) that $\mathbb{E}(y_t) = 0$ and

$$\operatorname{var}(y_t) = \sigma_v^2 \sum_{i=1}^{t} \mathbb{P}(s_i = 1) + (1 + \varphi_1^2)\sigma_v^2 [1 - \mathbb{P}(s_t = 1)].$$

Consequently, $\operatorname{var}(y_t)$ varies with t for both cases. In particular, when $\mathbb{P}(s_i = 1) = \pi$ is a constant between zero and one, $\operatorname{var}(y_t)$ grows linearly with t, but its magnitude is proportional to that of a random walk. These processes would have a bounded variance provided that $\sum_{t=1}^{\infty} \mathbb{P}(s_t = 1) < \infty$. By the Borel-Cantelli lemma, this condition implies $\mathbb{P}(s_t = 1 \text{ infinitely often}) = 0$, i.e., $s_t = 1$ for at most finitely many t with probability one. In this case, this process is eventually an AR(1) (or an MA(1)) process. In view of these variance properties, we refer to the proposed process as a *semi-nonstationary process*, so that it can be distinguished from the unit-root nonstationary processes.

When t tends to infinity, the first term on the right-hand side of Eqs. (6) and (7) is

$$y_{1,\infty} = \sum_{i=1}^{\infty} s_i v_i,$$

which may be interpreted as the "long-run stochastic trend." The presence of a particular innovation v_t in $y_{1,\infty}$ depends on the state variable s_t. Let $\mathcal{F}^t$ denote the information set up to time t and

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

denote the long-run effect of v_t on the optimal forecast of y_{t+k}. Recall that the long-run effect δ_t is one for a random walk and zero for a weakly stationary process. For the simplest STOPBREAK process,

$$\delta_t = q_t + \frac{\partial q_t}{\partial v_t} v_t = q_t + \frac{2\gamma v_t^2}{(\gamma + v_t^2)^2}.$$

When v_t has a continuous distribution, v_t is zero with probability zero. It follows that δ_t is positive with probability one. For the proposed processes, we can see from Eqs. (6) and (7) that $\delta_t = s_t$ which may be one or zero.

Let $z_t = (1 - B)y_t$. If y_t are generated from Eq. (1) and $\mathbb{P}(s_t = 1)$ is a constant π, it is easily verified from Eq. (3) that

$$z_t = \sum_{i=0}^{t-2} \psi_1^i (\psi_1 - 1)(1 - s_{t-1-i})\upsilon_{t-1-i} + \upsilon_t. \tag{8}$$

Thus, z_t has mean zero and

$$\text{var}(z_t) = \sigma_\upsilon^2 + (1 - \pi)(1 - \psi_1^{2t-2})\frac{1 - \psi_1}{1 + \psi_1}\sigma_\upsilon^2.$$

The autocovariances are

$$\text{cov}(z_t, z_{t-1}) = -(1 - \pi)(1 + \psi_1^{2t-1})\frac{1 - \psi_1}{1 + \psi_1}\sigma_\upsilon^2,$$

$$\text{cov}(z_t, z_{t-i}) = \psi^{i-1}\ \text{cov}(z_t, z_{t-1}), \quad i \geq 2,$$

see the Appendix for details. These autocovariances depend only on i as t goes to infinity. Thus, the first difference of y_t is asymptotically a covariance stationary process. When $\pi = 1$, z_t is simply a white noise, as originally postulated.

If y_t are generated by Eq. (2) and $\mathbb{P}(s_t = 1) = \pi$, it follows from Eq. (4) that

$$z_t = \upsilon_t - (\varphi_1 + 1)(1 - s_{t-1})\upsilon_{t-1} + \varphi_1(1 - s_{t-2})\upsilon_{t-2}. \tag{9}$$

Therefore, z_t has mean zero and

$$\text{var}(z_t) = \sigma_\upsilon^2 + (1 - \pi)[\varphi_1^2 + (\varphi_1 + 1)^2]\sigma_\upsilon^2.$$

The autocovariances are

$$\text{cov}(z_t, z_{t-1}) = -(1 - \pi)(\varphi_1 + 1)^2\sigma_\upsilon^2,$$

$$\text{cov}(z_t, z_{t-2}) = (1 - \pi)\varphi_1\sigma_\upsilon^2,$$

$$\text{cov}(z_t, z_{t-i}) = 0, \qquad i \geq 3.$$

The derivation of these results are also given in the Appendix. In this case, z_t is covariance stationary, and it is again a white noise when $\pi = 1$. Moreover, these autocovariances agree with those of a non-invertible MA(2) process if and only if $\pi = 0$. Thus, z_t is invertible provided that $\pi > 0$.

3 General Semi-Nonstationary Processes

The simple processes (1) and (2) can easily be extended to allow for more general dynamic behavior. We first consider the following generalization:

$$\begin{aligned} y_t &= y_{1,t} + y_{0,t}, \\ (1-B)y_{1,t} &= \alpha_0 + s_t\upsilon_t, \\ \Psi(B)y_{0,t} &= \Phi(B)(1-s_t)\upsilon_t, \end{aligned} \tag{10}$$

where s_t are state variables taking the value one or zero, $\Psi(B) = 1-\psi_1 B - \cdots - \psi_m B^m$ 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. Setting $(\upsilon_0, \upsilon_{-1}, \ldots)$ to zero, it can be seen that

$$(1-B)y_t = \alpha_0 + s_t\upsilon_t + (1-B)\Psi(B)^{-1}\Phi(B)(1-s_t)\upsilon_t,$$

or equivalently,

$$y_t = \alpha_0\, t + \sum_{i=1}^{t} s_i\upsilon_i + \Psi(B)^{-1}\Phi(B)(1-s_t)\upsilon_t. \tag{11}$$

This process now contains three components − namely, a deterministic trend $\alpha_0 t$, a long-run stochastic trend

$$y^*_{1,t} = \sum_{i=1}^{t} s_i\upsilon_i,$$

and a weakly stationary process $\Psi(B)^{-1}\Phi(B)(1-s_t)\upsilon_t$ that gives rise to short-run fluctuations. The process (10) is referred to as an SN$(1; m, n)$ process with a linear trend. When $\alpha_0 = 0$, it is an SN$(1; m, n)$ process without a linear trend. Clearly, the models in Eqs. (1) and (2) are, respectively, SN$(1; 1, 0)$ and SN$(1; 0, 1)$ processes without a linear trend.

Similar to the simple processes discussed before, not all υ_i of this process enter its stochastic trend component. In view of Eq. (11), the long-run effect of υ_t on y_{t+k} is also $\delta_t = s_t$ which may be one or zero. The general process thus constitutes intermediate cases between a random walk with drift (when $s_t = 1$ with probability one for all t) and a trend-stationary process (when $s_t = 0$ with probability one for all t). This process exhibits large swings and tends to drift away from the trend line during the periods that $s_t = 1$, but it induces only temporary fluctuations around the trend while $s_t = 0$.

The process (10) has an ARMA representation with MA random coefficients:

$$\Psi(B)(1-B)y_t = \alpha_0\Psi(1) + \sum_{i=1}^{r+1} \xi_{i,s_{t-i}} \upsilon_{t-i} + \upsilon_t, \tag{12}$$

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 $i = 2, \ldots, r$, 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 $i > n$. For the dynamic behavior of y_t, only the past state variables $s_{t-1}, \ldots, s_{t-r-1}$ are relevant; the current state s_t does not affect y_t because $s_t\upsilon_t$ and $(1-s_t)\upsilon_t$ are both present at time t.

We again assume that $\{\upsilon_t\}$ is a sequence of random variables such that $\mathbb{E}(\upsilon_t|S^t) = 0$, $\text{var}(\upsilon_t|S^t) = \sigma_\upsilon^2$, and $\mathbb{E}(\upsilon_t\upsilon_{t-i}|S^t) = 0$ for all $i > 0$. For simplicity, we maintain the assumption that $\mathbb{P}(s_t = 1) = \pi$, a non-zero constant. Thus, $\text{var}(s_t\upsilon_t) = \pi\sigma_\upsilon^2$ and $\text{var}((1-s_t)\upsilon_t) = (1-\pi)\sigma_\upsilon^2$. Moreover, $s_t\upsilon_t$ and $(1-s_t)\upsilon_t$ are serially uncorrelated and pairwise uncorrelated at all leads and lags. It is then easy to verify that $\mathbb{E}(y_t) = \alpha_0\, t$ and

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

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, $\text{var}(y_t)$ grows linearly with t as $\sigma_\upsilon^2\pi t$ which is proportional to that of a pure random walk. The autocovariance generating function of $z_t = (1-B)y_t$ is

$$g(a) = \pi\,\sigma_\upsilon^2 + (1-\pi)(1-a)(1-a^{-1})\Psi(a)^{-1}\Psi(a^{-1})^{-1}\Phi(a)\Phi(a^{-1})\sigma_\upsilon^2;$$

see the Appendix. From this generating function, all the autocovariances of z_t obtained in the preceding section can also be derived straightforwardly. Note that when $\pi > 0$, z_t is invertible.

We may also extend Eq. (10) to allow for breaking trends. Setting $\upsilon_t = \alpha_1 + \varepsilon_t$, Eq. (10) becomes

$$\begin{aligned} y_t &= y_{1,t} + y_{0,t}, \\ (1-B)y_{1,t} &= (\alpha_0 + s_t\alpha_1) + s_t\varepsilon_t, \\ \Psi(B)y_{0,t} &= \Phi(B)(1-s_t)\alpha_1 + \Phi(B)(1-s_t)\varepsilon_t. \end{aligned} \tag{13}$$

This process can be expressed as

$$\begin{aligned} y_t = \alpha_0 t + \alpha_1 \sum_{i=1}^{t} s_i + \alpha_1 \Psi(B)^{-1}\Phi(B)(1-s_t) \\ + \sum_{i=1}^{t} s_i \varepsilon_i + \Psi(B)^{-1}\Phi(B)(1-s_t)\varepsilon_t, \end{aligned} \tag{14}$$

which differs from Eq. (11) by its nonlinear deterministic trend component. The nonlinear trend component includes a "fundamental" linear trend $\alpha_0\, t$ which is not affected by s_t, a trend-shift component $\alpha_1 \sum_{i=1}^{t} s_i$ that alters the fundamental trend when $s_i = 1$, and a perturbation component $\alpha_1 \Psi(B)^{-1}\Phi(B)(1-s_t)$ based on those $s_i = 0$. When $\alpha_0 = 0$, there is no fundamental trend, so that a trend occurs when $s_t = 1$. When $\alpha_1 = 0$, there will be no trend shift, and the fundamental trend prevails. If s_t follows the first-order Markov chain, as we subsequently assume, the first two components of Eq. (14) form the "Markov trend" of Hamilton (1989). Thus, the process (14) will be referred to as the SN$(1; m, n)$ process with a Markov trend.

Note that Eq. (14) differs from Hamilton's regime switching model in several ways. First, the innovations with permanent (transitory) effect are associated with the shifted (fundamental) trend. This is because the state variables s_t in (14) simultaneously determine the effect of innovations and the pattern of the trend line. 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. Second, while Hamilton's model yields a kinked trend line, the perturbation component of (14) induces smooth transitions between trend segments. This can be seen from Figure 2, where the trend of (14) is simulated with $\Psi(B) = 1$, $\Phi(B) = 1 + 0.4B + 0.2B^2$, $\alpha_0 = 1$, $\alpha_1 = 2$ and $s_t = 0$ for $t \in [6, 15]$ and $[26, 35]$. This trend will then be referred to as the "smooth Markov trend." Third, a key condition for Hamilton's regime switching is $\alpha_1 \neq 0$, yet it is not a necessary condition for semi-nonstationarity.

Further generalizations of the model are possible. For example, we may postulate $y_{1,t}$ as an ARIMA process to allow for more general short-run dynamics when $s_t = 1$. We may also consider switching between long-memory and transitory shocks. This can be done by setting $y_{1,t}$ as a fractionally integrated process or a more complex ARFIMA process. The seasonal component can also be easily incorporated into the model. Moreover, it is possible to specify a different switching mechanism for s_t, e.g., a threshold mechanism resulting in a generalization of the threshold-disturbance moving-average (TDMA) model of Elwood (1998). We will not pursue these possibilities in this paper, however.

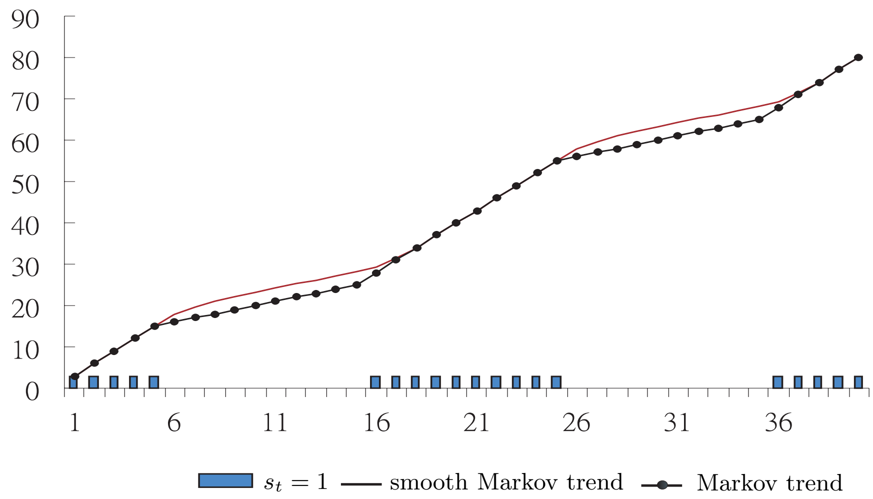

Figure 2: Simulated Markov trend and smooth Markov trend.

4 Model Estimation

The SN$(1; m, n)$ model in Eq. (12) can be estimated either by the maximum likelihood method or Markov chain Monte Carlo methods. We use the former in this paper. To allow for state persistence, we follow Hamilton (1989) and postulate that s_t follows a two-state Markov chain with the transition matrix

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

where the two-state structure is consistent with our classification of innovations. The derivation below is similar, but not identical, to those of Hamilton (1989, 1994) and Kim (1993) because y_t depends only on the past (but not the current) state variables, as can be seen from the ARMA representation in Eq. (12).

From Eq. (12) 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 $r = 2$,

$$s^*_{t-1} = 1 \text{ if } s_{t-1} = s_{t-2} = s_{t-3} = 0,$$

$$s^*_{t-1} = 2 \text{ if } s_{t-1} = 0, s_{t-2} = 0, \text{ and } s_{t-3} = 1,$$

$$s^*_{t-1} = 3 \text{ if } s_{t-1} = 0, s_{t-2} = 1, \text{ and } s_{t-3} = 0,$$

$$\vdots$$

$$s^*_{t-1} = 8 \text{ if } s_{t-1} = s_{t-2} = s_{t-3} = 1.$$

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

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

with $\boldsymbol{P}_{ji}$ $(j, i = 0, 1)$ being a $2^{r-1} \times 2^r$ block diagnoal 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 $\boldsymbol{v}_{t-1} = (v_{t-1}, \ldots, v_{t-r-1})'$ 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,

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

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

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

We first discuss the optimal forecasts of the state variable s_t based on the information up to time t: $\mathscr{Z}^t = \{z_t, z_{t-1}, \ldots, z_1\}$, 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

$$\begin{aligned} &f(z_t \mid s_{t-1}^* = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) \\ &\qquad = \frac{1}{\sqrt{2\pi\sigma_\upsilon^2}} \exp\left\{ \frac{-(z_t - \sum_{j=1}^m \psi_j z_{t-j} - \boldsymbol{\xi}_{t-1,\ell}' \boldsymbol{\upsilon}_{t-1})^2}{2\sigma_\upsilon^2} \right\}, \end{aligned} \tag{15}$$

where $\ell = 1, 2, \cdots, 2^{r+1}$ and

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

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

$$\begin{aligned} \upsilon_t &= z_t - \mathbb{E}(z_t \mid \mathcal{Z}^{t-1}) \\ &= z_t - \sum_{j=1}^m \psi_j z_{t-j} - \sum_{\ell=1}^{2^{r+1}} \mathbb{P}(s_{t-1}^* = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, \boldsymbol{\xi}_{t-1,\ell}' \boldsymbol{\upsilon}_{t-1}, \end{aligned} \tag{16}$$

with the initial values $\upsilon_m, \ldots, \upsilon_1$ 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 (15) as

$$f(z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}) = \sum_{\ell=1}^{2^{r+1}} \mathbb{P}(s_{t-1}^* = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, f(z_t \mid s_{t-1}^* = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta}). \tag{17}$$

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

$$\mathbb{P}(s_{t-1}^* = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta}) = \frac{\mathbb{P}(s_{t-1}^* = \ell \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})\, f(z_t \mid s_{t-1}^* = \ell, \mathcal{Z}^{t-1}; \boldsymbol{\theta})}{f(z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})}. \tag{18}$$

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 $\{\upsilon_t\}$ are independent. These results in turn yield

$$\begin{aligned} \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta}) &= \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t-1}^* = j \mid \mathcal{Z}^t; \boldsymbol{\theta})\ \mathbb{P}(s_t^* = \ell \mid s_{t-1}^* = j, \mathcal{Z}^t; \boldsymbol{\theta}) \\ &= \sum_{j=1}^{2^{r+1}} p_{j\ell}^*\ \mathbb{P}(s_{t-1}^* = j \mid \mathcal{Z}^t; \boldsymbol{\theta}). \end{aligned} \tag{19}$$

Thus, with the initial value $\mathrm{I\!P}(s^*_m \mid \mathcal{Z}^m; \boldsymbol{\theta})$, we can iterate the equations (15)–(19) to obtain $\mathrm{I\!P}(s^*_t = \ell \mid \mathcal{Z}^t; \boldsymbol{\theta})$ for $t = m+1, \ldots, T$. Then for each t, the desired filtering probability is

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

and $\mathrm{I\!P}(s_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}) = 1 - \mathrm{I\!P}(s_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta})$, 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 $\mathrm{I\!P}(s^*_m \mid \mathcal{Z}^m; \boldsymbol{\theta})$ 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 $\mathrm{I\!P}(s_t \mid \mathcal{Z}^T)$ for $t \leq T$, which are the optimal forecasts of s_t based on all the information in the sample. Observe that

$$\begin{aligned} &\mathrm{I\!P}(s^*_t = \ell \mid s^*_{t+1} = j, \mathcal{Z}^T) \\ &\qquad = \frac{\mathrm{I\!P}(s^*_t = \ell \mid s^*_{t+1} = j, \mathcal{Z}^{t+1})\ \mathrm{I\!P}(z_T, \ldots, z_{t+3} \mid s^*_t = \ell, s^*_{t+1} = j, \mathcal{Z}^{t+1})}{\mathrm{I\!P}(z_T, \ldots, z_{t+3} \mid s^*_{t+1} = j, \mathcal{Z}^{t+1})}. \end{aligned}$$

In the current context,

$$\mathrm{I\!P}(z_T, \cdots, z_{t+3} \mid s^*_t = \ell, s^*_{t+1} = j, \mathcal{Z}^{t+1}) = \mathrm{I\!P}(z_T, \cdots, z_{t+3} \mid s^*_{t+1} = j, \mathcal{Z}^{t+1}),$$

so that $\mathrm{I\!P}(s^*_t = \ell \mid s^*_{t+1} = j, \mathcal{Z}^T) = \mathrm{I\!P}(s^*_t = \ell \mid s^*_{t+1} = j, \mathcal{Z}^{t+1})$. Note, however, that the condition above does not hold in Kim (1993). It follows that

$$\begin{aligned}
&\mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^T) \\
&\qquad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \mathbb{P}(s_t^* = \ell \mid s_{t+1}^* = j, \mathcal{Z}^{t+1}) \\
&\qquad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \frac{\mathbb{P}(s_{t+1}^* = j \mid s_t^* = \ell, \mathcal{Z}^{t+1})\ \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^{t+1})}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})} \qquad (20) \\
&\qquad = \mathbb{P}(s_t^* = \ell \mid \mathcal{Z}^{t+1}) \sum_{j=1}^{2^{r+1}} \frac{p_{\ell j}^*\ \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})}.
\end{aligned}$$

Using the filtering probability $\mathbb{P}(s_T^* = \ell \mid \mathcal{Z}^T)$ as the initial value we can iterate the equations (18), (19) and (20) 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 of the proposed model is whether the data are indeed 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 (the parameters 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. A process may still be semi-nonstationary even when $\alpha_1 = 0$. Once we exclude the possibility that the process is a pure random walk, hypothesis testing of other parameters is standard and can be done using likelihood-based tests. Therefore, we focus on testing the null hypothesis of $p_{11} = 1$.

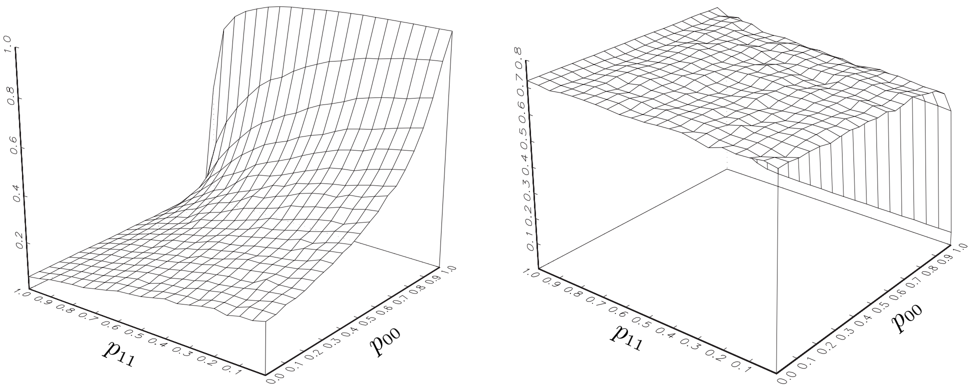

Figure 3: Empirical powers of the Dickey-Fuller and KPSS tests.

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. (14) with $\alpha_0, \alpha_1 = 0$, $\sigma_v^2 = 1$, $\Psi(B) = 1 - 0.5B$, $\Phi(B) = 1$, 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 3. We see that the DF test is not powerful against semi-nonstationary alternatives, 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.

On the other hand, the KPSS test of Kwiatkowski et al. (1992) is more powerful against semi-nonstationary data, as shown in the right panel of Figure 3. The rejection frequencies are quite stable and around 70% when p_{11} and p_{00} are between 0.1 and 0.9. This shows that the KPSS test has power against semi-nonstationarity. When the KPSS test rejects the null of stationarity, it is, however, difficult to decide whether the series being tested is a semi-nonstationary process or a pure random walk (a unit-root process). Further study is needed to justify semi-nonstationarity.

In this paper, we propose using a simulation-based test and a simple specification test to determine whether a process is a random walk. For the former test, we assume that the proposed semi-nonstationary model is the model under the null hypothesis. Given the empirical data, we first estimate an array of $\mathrm{SN}(1; m, n)$ models and choose an appropriate specification based on an information criterion (e.g., AIC or SIC). Denote the selected model as $\mathrm{SN}(1; m^*, n^*)$ and its estimated transition probability as $\hat{p}^*_{11}$. With the estimated parameters, the selected SN$(1; m^*, n^*)$ model is then taken as the data generating process to generate simulated samples. For each simulated sample, we re-estimate an SN$(1; m^*, n^*)$ 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 $p_{11} = \hat{p}_{11}^*$ if the p-value of $\hat{p}_{11}^*$ is small, say, less than 5%.

For the specification test, we simply take the random walk model as the null hypothesis and notice that $z_t = y_t - y_{t-1}$ is uncorrelated with all past z_{t-i} under the null. We can regress z_t on $z_{t-1}, \ldots z_{t-k}$ for any positive k and test the null hypothesis by invoking the standard Wald statistic for joint significance of the coefficients of z_{t-i}. 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.

6 Empirical Study

To assess the empirical relevance of the proposed model, we apply the model (13) 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) uses the Markov trend and an ARIMA model to describe GNP; Lam (1990) employs the same model without imposing a unit root. Kim and Nelson (1999) also apply the regime switching model to GDP. As the proposed model constitutes intermediate cases between the trend-stationary and unit-root models while allowing for a smooth Markov trend, it would be interesting to know if this model 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 the variable y_t in the SN$(1; m, n)$ model (13) and estimate an array of SN$(1; m, n)$ models with m and n no greater than 4. The parameters,

$$\boldsymbol{\theta} = (\psi_1, \ldots, \psi_m, \varphi_1, \ldots, \varphi_n, \sigma_v, p_{00}, p_{11})',$$

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 $\hat{\boldsymbol{\theta}}$.

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_v	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.

Among all the models considered, both AIC and SIC select the SN$(1; 2, 2)$ 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 described in Section 5 to examine the parameter p_{11}. Simulated data are generated using the estimated parameters in Table 2. We then re-estimate the SN$(1; 2, 2)$ model using the simulated data and obtain a new estimate $\hat{p}_{11}$. With 5000 replications we obtain a simulated distribution of $\hat{p}_{11}$. The mean of this distribution is 0.938, and the p-value of 0.95 is 0.19606. We thus do not reject the null hypothesis that $p_{11} = 0.95$. We also regress z_t on $z_{t-1}, \dots, z_{t-k}$ for $k = 1, \cdots, 4$ with a constant term. The resulting Wald statistics for the joint significance of the associated coefficients 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 not a pure random walk, we now proceed to test other hypotheses using 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 justifies our Markovian specification of the state variable.

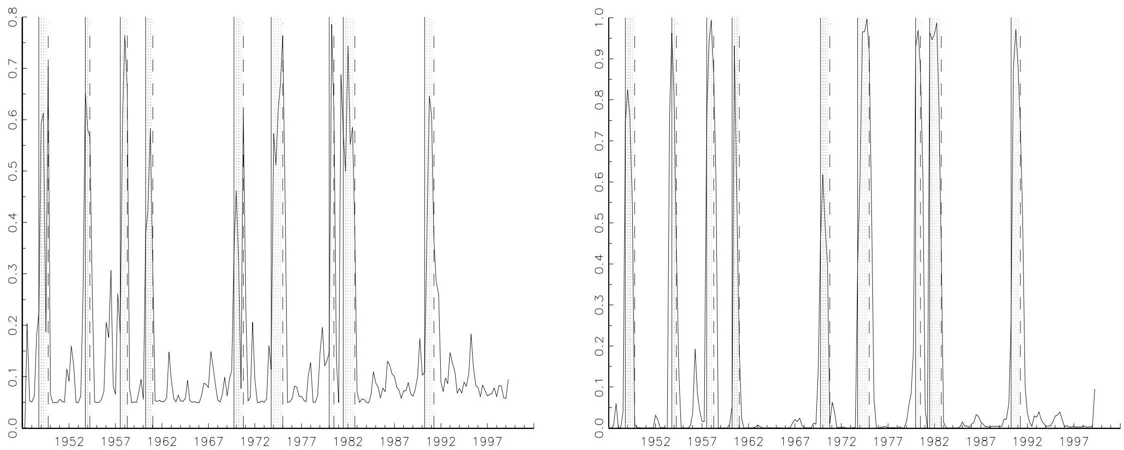

Figure 4: Estimated filtering and smoothing probabilities of $s_t = 0$.

In Figure 4 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). Among these probabilities, 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 $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta}) > 0.5$. That is, more than 80 percent (but not all) of the innovations are more likely to have permanent effects. The finding that permanent shocks occur quite frequently is different from the assertions of trend break models, such as Perron (1989) and Balke and Fomby (1991). Moreover, the periods that are more likely to be in the state of transitory (permanent) shocks match the NBER recession (expansion) periods very closely. The shocks in recessions (expansions) are therefore more likely to have transitory (permanent) effect. Beaudry and Koop (1993) found that positive shocks to GDP are more persistent than negative shocks. As negative (positive) shocks are closely related to the shocks in recessions (expansions), our result is compatible with that of Beaudry and Koop (1993).

From Table 2 we find 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

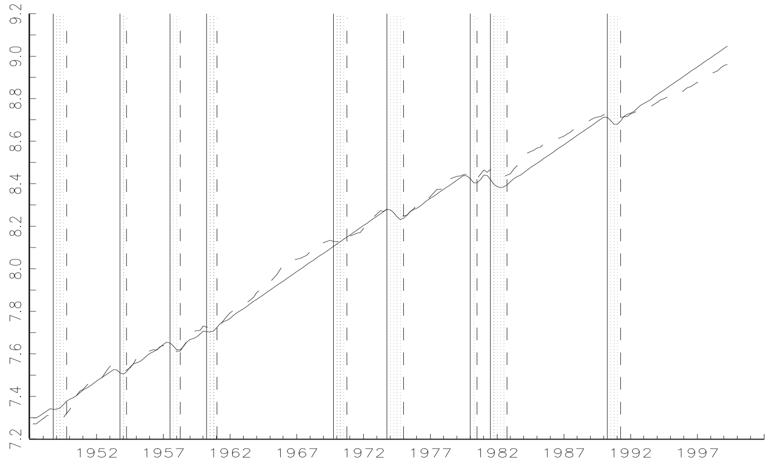

Figure 5: The expected trend line in U.S. real GDP.

for recession and $1/(1-0.95) = 20$ quarters for expansion. According to NBER 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. We also apply the model of Hamilton (1989) to the data set. The estimation Gauss program used is 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. Thus, the data cannot be properly classified into two regimes. This is in contrast with the result of Hamilton (1989).[2] For example, the estimated $\hat{p}_{11}$ is close to one so that almost all the filtering or smoothing probabilities of $s_t = 1$ are near unity. Kim and Nelson (1999, p. 78) reported a similar problem when a smaller data set was used; Boldin (1996) also 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 5.

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

[2] In Hamilton (1989), the estimated positive (negative) growth rate is 1.16% (−0.36%); the duration of high (low) growth periods lasts for about 10 (4) quarters. Our results thus suggest a longer duration for expansion and "deeper" recessions.

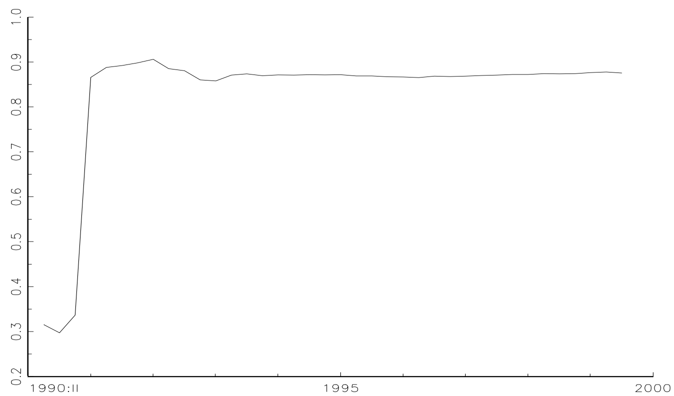

Figure 6: Dynamic path of $\mathbb{P}(s_{1990:\text{III}} = 0 \mid \mathcal{Z}^{\tau}; \boldsymbol{\theta})$.

Examining the filtering and smoothing probabilities more carefully, we find that there are 14 periods with $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}) < 0.5$ but $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta}) > 0.5$. That is, the likelihood of $s_t = 0$ may differ significantly when information sets change. 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 6 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 of permanent shocks to the state of transitory shocks as information set expands.

7 Conclusions

In this paper we proposed a class of models with the many interesting features. First, the effect of each innovation in the model may be either permanent or transitory. This results in a flexible stochastic trend. Second, the model can generate a smooth Markov trend so that smooth transitions of the deterministic trend are allowed. Third, different trending characteristics are directly linked to the shocks with distinct effects. This class of models constitutes intermediate cases between (trend) stationary and unit-root models and therefore can accommodate both trend-reverting and trend-perverting behavior. The empirical application of the proposed model suggests that it is a useful analytical tool.

In particular, we find that for U.S. real GDP data, the shocks in recession (expansion) periods are more likely to be transitory (permanent) and that the shocks with permanent effects occur more frequently. The former result is consistent with that of Beaudry and Koop (1993), whereas the latter finding differs from the conclusions of trend-break models, such as Perron (1989) and Balke and Fomby (1991). The proposed model thus serves as a useful alternative to analyze economic and financial time series.

Appendix

The variance and autocovariances of $z_t = (1-B)y_t$.

When y_t is generated as Eq. (1), recall from Eq. (8) that $z_t = (1-B)y_t$ is

$$z_t = \sum_{i=0}^{t-2} \psi_1^i(\psi_1 - 1)(1 - s_{t-1-i})\upsilon_{t-1-i} + \upsilon_t.$$

Note that for each $i \geq 0$,

$$\mathrm{var}\big((1 - s_{t-1-i})\upsilon_{t-1-i}\big) = \sigma_\upsilon^2\ \mathbb{E}(1 - S_{t-1-i})^2 = \sigma_\upsilon^2\,[1 - \mathbb{P}(S_{t-1-i} = 1)],$$

and that $(1 - s_{t-1-i})\upsilon_{t-1-i}$ are uncorrelated with υ_t. It follows that $\mathbb{E}(z_t) = 0$ and for $\pi = \mathbb{P}(S_t = 1)$,

$$\begin{aligned}
\mathrm{var}(z_t) &= \sigma_\upsilon^2 + (1-\pi)(1-\psi_1)^2 \sum_{i=0}^{t-2}(\psi_1^i)^2\sigma_\upsilon^2 \\
&= \sigma_\upsilon^2 + (1-\pi)(1-\psi_1)^2 \frac{1-\psi_1^{2t-2}}{1-\psi_1^2}\sigma_\upsilon^2 \\
&= \sigma_\upsilon^2 + (1-\pi)(1-\psi_1^{2t-2})\frac{1-\psi_1}{1+\psi_1}\sigma_\upsilon^2.
\end{aligned}$$

Writing $z_t = \psi_1 z_{t-1} + \upsilon_t - \psi_1 s_{t-1}\upsilon_{t-1} - (1 - s_{t-1})\upsilon_{t-1}$, we have

$$\begin{aligned}
\mathrm{cov}(z_t, z_{t-1}) &= \psi_1\,\mathrm{var}(z_{t-1}) - \psi_1\pi\sigma_\upsilon^2 - (1-\pi)\sigma_\upsilon^2 \\
&= \psi_1(1-\pi)(1-\psi_1^{2t-2})\frac{1-\psi_1}{1+\psi_1}\sigma_\upsilon^2 + (1-\pi)(\psi_1 - 1)\sigma_\upsilon^2 \\
&= -(1-\pi)\frac{(1-\psi_1)}{1+\psi_1}\sigma_\upsilon^2 - (1-\pi)\psi_1^{2t-1}\frac{1-\psi_1}{1+\psi_1}\sigma_\upsilon^2 \\
&= -(1-\pi)(1+\psi_1^{2t-1})\frac{1-\psi_1}{1+\psi_1}\sigma_\upsilon^2.
\end{aligned}$$

Similarly, it can be seen that

$$\mathrm{cov}(z_t, z_{t-2}) = \mathrm{cov}(\psi_1 z_{t-1}, z_{t-2}) = \psi_1\,\mathrm{cov}(z_{t-1}, z_{t-2}).$$

Consequently, $\text{cov}(z_t, z_{t-i}) = \psi_1^{i-1} \text{cov}(z_t, z_{t-1})$ for $i \geq 2$.

When y_t is generated as (2), recall from (9) that

$$z_t = \upsilon_t - (\varphi_1 + 1)(1 - s_{t-1})\upsilon_{t-1} + \varphi_1(1 - s_{t-2})\upsilon_{t-2}.$$

It follows that z_t has mean zero and for $\pi = \mathbb{P}(s_t = 1)$,

$$\begin{aligned} \text{var}(z_t) &= \sigma_\upsilon^2 + (\varphi_1 + 1)^2(1 - \pi)\sigma_\upsilon^2 + \varphi_1^2(1 - \pi)\sigma_\upsilon^2 \\ &= \sigma_\upsilon^2 + (1 - \pi)[\varphi_1^2 + (\varphi_1 + 1)^2]\sigma_\upsilon^2. \end{aligned}$$

The autocovariance $\text{cov}(z_t, z_{t-1})$ is

$$\begin{aligned} &\text{cov}\big(-(\varphi_1 + 1)(1 - s_{t-1})\upsilon_{t-1} + \varphi_1(1 - s_{t-2})\upsilon_{t-2},\ \upsilon_{t-1} - (\varphi_1 + 1)(1 - s_{t-2})\upsilon_{t-2}\big) \\ &\quad = -(\varphi_1 + 1)(1 - \pi)\sigma_\upsilon^2 - \varphi_1(\varphi_1 + 1)(1 - \pi)\sigma_\upsilon^2 \\ &\quad = -(1 - \pi)(\varphi_1 + 1)^2\sigma_\upsilon^2. \end{aligned}$$

Similarly,

$$\begin{aligned} &\text{cov}(z_t, z_{t-2}) = (1 - \pi)\varphi_1\sigma_\upsilon^2, \\ &\text{cov}(z_t, z_{t-i}) = 0, \qquad i \geq 3. \end{aligned}$$

The autocovariance generating function of $z_t = (1 - B)y_t$.

We consider the process (10): $y_t = y_{1,t} + y_{0,t}$ with

$$\begin{aligned} (1 - B)y_{1,t} &= \alpha_0 + s_t\upsilon_t, \\ \Psi(B)y_{0,t} &= \Phi(B)(1 - s_t)\upsilon_t. \end{aligned}$$

Write z_t as the sum of two components:

$$z_t = z_{1,t} + z_{2,t},$$

where $z_{1,t} = \alpha_0 + s_t\upsilon_t$ and $z_{2,t} = (1-B)\Psi(B)^{-1}\Phi(B)(1-s_t)\upsilon_t$. Since $\text{cov}(s_t\upsilon_t, (1-s_t)\upsilon_t) = 0$, $z_{1,t}$ is uncorrelated with $z_{2,\tau}$ for all t and τ. Let g_{z_1} and g_{z_2} denote the autocovariance generating functions of z_1 and z_2, respectively. Thus, the autocovariance generating function of z_t is

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

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.

Beveridge S. and C.R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle", *Journal of Monetary Economics*, **7**, 151–174.

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 cylce, *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.

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.

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

Watson, M.W. (1986). Univariate detrending methods with stochastic trends, *Journal of Monetary Economics*, **18**, 49–75.