An Unobserved-Component Model with Switching Permanent and Transitory Innovations

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: October 10, 2004

† This paper was originally entitled "A component-driven model for regime switching and its empirical evidence." 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 on early drafts of this paper. We are also indebted to the associate editor and two anonymous referees whose comments lead to a much improved version of the paper. All remaining errors are ours. 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 an unobserved-component model in which component innovations are governed by a state variable that follows a Markov process. The proposed model is capable of describing both stationary and non-stationary behaviors of real data and allows the random innovations to have permanent and transitory effects in different periods. The model also permits a deterministic trend with or without breaks and hence bridges the gap between the trend-stationary model and a random walk with drift. For ease in presentation and in application, our discussion focuses on the model consisting of a random walk component and a stationary ARMA component. However, the proposed model is much more flexible. We investigate properties of the proposed model and derive an estimation algorithm. We also propose a simulation-based test to distinguish between the proposed model and an ARIMA model. In empirical application, we apply the 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 expansion (recession) are more likely to have a permanent (transitory) effect.

Keywords: Innovation regime-switching model, Markov trend, permanent innovation, transitory innovation, trend stationarity, unit-root nonstationarity, unobserved-component model.

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. Econometric models that can 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. A more general 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 has not received much attention in the literature.

In this paper, we follow the latter approach and propose an unobserved-component model with regime switching. 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 while the other component continues to evolve without any innovation. As such, there is only a single innovation for the observed process at each time point. The prevailing component in turn determines the characteristics of the innovation and, hence, 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 helpful to consider a model that can accommodate both dynamic patterns.

For simplicity, we focus on the special model consisting of a random walk component and a stationary autoregressive and moving-average (ARMA) component with a Markov process governing the state transition. This simple model permits a deterministic trend with or without breaks and hence can serve as an intermediate case between the trend-stationary model and a random walk with drift. It can also be a model that 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. This 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. Details are given in Section 2. 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).

We also derive an estimation algorithm for the proposed model using a state-space representation and propose a simulation-based test to distinguish between the model and a pure ARIMA process. As an application, we apply the proposed model to U.S. quarterly real gross domestic product (GDP), and find that unit-root nonstationarity is likely to be the prevailing dynamic pattern for about 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 inference drawn from either a unit-root or a trend-break model. Furthermore, we observe that the nonstationarity (stationarity) periods match the NBER dating of expansions (recessions) closely. These results suggest that the innovations in expansion (recession) are more likely to have a permanent (transitory) effect. That the shocks in expansion are more persistent than those in recession 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 unobserved-component 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. The empirical analysis of U.S. real GDP based on the proposed model is presented in Section 5. Section 6 concludes the paper.

2 An Innovation Switching Model

Consider the simple case of two components. Suppose that the observed process y_t is the sum of two components — namely,

$$y_t = y_{1,t} + y_{0,t} \tag{1}$$

where

$$y_{1,t} = g(y_{1,t-1}, \dots, y_{1,t-p}; \boldsymbol{\theta}_1) + s_t \upsilon_t,$$

$$y_{0,t} = h(y_{0,t-1}, \dots, y_{0,t-q}; \boldsymbol{\theta}_0) + (1 - s_t)\upsilon_t,$$

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 $\{v_t\}$ is a sequence of independent and identically distributed random variables. The model is readily generalized to have more than two components.

The proposed model in Eq. (1) has some unique characteristics. First, each innovation v_t excites only one component determined by the state variable s_t. As such, s_t determines the dynamic effect of v_t on the observed series. When $s_t = 1$, the first component $y_{1,t}$ is activated and the effect of v_t will be propagated through the function $g(.)$ while the second component $y_{0,t}$ evolves according to h without any innovation. When $s_t = 0$, v_t excites $y_{0,t}$ but does not enter the first component y_{1t}. Furthermore, since both $s_t v_t$ and $(1 - s_t)v_t$ are present in Eq. (1), the state variable s_t does not affect y_t. But it determines how the innovation v_t affects the subsequent y_{t+j} for $j > 0$. Second, the dynamic of the observed process y_t is a mixture of two components governed by the state variable s_t. For distinct functions $g(.)$ and $h(.)$, the dynamic of each component evolves over time with input $\{y_{1,t-j}\}$ and $\{y_{0,t-j}\}$, respectively. On the other hand, the regime-switching models of Hamilton (1989, 1994) and Tong (1990) typically contain a single component with state-dependent parameter $\boldsymbol{\theta}_{s_t}$:

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

Given that there is only one function $g(.)$ in the model, the dynamic structures in different regimes are similar in essence. Also, the input of the $g(.)$ function is $\{y_{t-j}\}$ for both regimes. Third, for the proposed model, the state variable s_{t+j} $(j > 0)$ determines the allocation of the innovation v_{t+j}, but the effect of v_t on y_{t+j} is controlled by the state s_t. This special characteristic enables use to classify the innovation v_t as a permanent or transitory shock. On the other hand, many econometric models available in the literature classify the impact of v_t on y_{t+j} based on the state variable s_{t+j}, and can lead to different classifications for the same innovation v_t when j changes.

Although it is possible to express Eq. (1) as a special case of a general model with state-dependent coefficients (for example, the ARMA model of Eq. (4) and the state-space model of Eq. (8) below), this two-component structure with the state variable linked to innovations is convenient for researchers to specify the dynamic patterns they would like to characterize.

2.1 The Proposed Model

Clearly, the model in Eq. (1) is not operational unless the functions g and h are specified. Many selections are possible. Our choice of the functions is motivated by two important empirical characteristics commonly observed in application, 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. These two models generate distinct dynamic behaviors and bear quite different economic interpretations. 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. In this paper we propose a model based on Eq. (1) that can accommodate these two distinctive dynamic patterns.

We propose an unobserved-component model that consists of a random walk component and a stationary ARMA component. That is,

$$y_t = y_{1,t} + y_{0,t} \tag{2}$$

with

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

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, and $v_t = \alpha_1 + \varepsilon_t$ with $\{\varepsilon_t\}$ a white noise with mean zero. Note that allowing v_t to have a (possibly) non-zero mean adds more flexibility to the model; see further discussion below. To avoid identification problem, we do not include a state-independent constant term in the second component $y_{0,t}$. In what follows, we shall postulate that s_t follows a first-order Markov chain, as in Hamilton (1989).

Similar to Clark (1987) and Cochrane (1988), y_t in the proposed model is the sum of a random walk and a stationary component. A major difference is that the innovation in our model excites only one component in a given time period. When $s_t = 1$, the first component is excited and evolves like a random walk; the associated innovation ε_t has a permanent effect on y_{t+j} ($j > 0$). When $s_t = 0$, ε_t excites the stationary ARMA component and has a transitory effect on future y_{t+j}. Hence, the effect of an innovation may alternate from time to time, depending on the state variable s_t. Many existing models, on the other hand, postulate that there are permanent and transitory innovations at each time index. As such, these innovations together still have a permanent effect and generate unit-root nonstationarity for all time indices. This setup may not always be appropriate, however. First, it is conceivable that random shocks may be fundamentally different in different periods. 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). Second, a time series does not always exhibit unit-root nonstationarity. For example, an asset price may have large changes and behave like a unit-root process when the market is inundated with vital information, yet it may fluctuate only slightly and behave like a stationary process when the market is tranquil. By allowing the innovation to excite only one component, the proposed model can easily distinguish between the innovation effects and allows for distinct (unit-root and stationary) dynamics in different periods.

Setting $y_0 = 0$ and $\varepsilon_i = 0$ for $i \leq 0$, 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 the deterministic trend and does not depend on the state variable. This trend function is altered by the second component $\alpha_1 \sum_{i=1}^{t} s_i$ if $\alpha_1 \neq 0$. In particular, the deterministic trend is subject to 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. Of course, there will be no trend break if $\alpha_1 = 0$. It can also be seen that, when s_t follows the first-order Markov chain, the first two components of Eq. (3) are just the "Markov trend" of Hamilton (1989). While the Markov trend is kinked, the third component involves a weighted sum of $1 - s_{t-j}$ and induces smooth transitions between trend segments. Figure 1 illustrates a Markov trend and a smooth trend function of Eq. (3).[1] The difference between these trend behaviors is evident. 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. As different regimes in this model is linked to the effects of innovations, Eq. (2) with $\alpha_1 \neq 0$ and s_t a first-order Markov chain will be referred to as an Innovation Regime-Switching model of orders 1 and (m, n), or IRS$(1; m, n)$ for short, with a smooth Markov trend.

[1]The smooth trend function is generated by setting $\alpha_0 = 1$, $\alpha_1 = 2$, $\Psi(B) = 1$, $\Phi(B) = 1{+}0.4B{+}0.2B^2$, $s_t = 1$ for $t = 1, \ldots, 5, 16, \ldots, 25, 36, \ldots, 40$ (as indicated by the shaded boxes on the horizontal axis), and $s_t = 0$ otherwise.

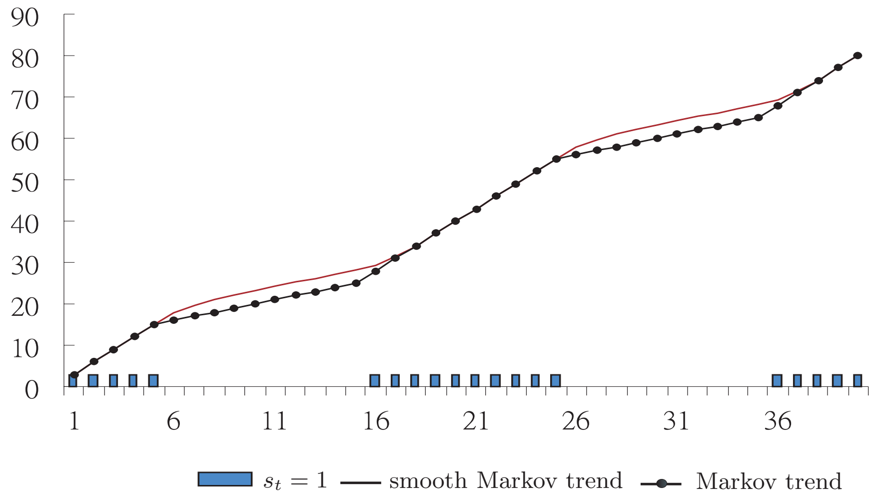

Figure 1: Simulated Markov trend and smooth Markov trend.

When $\alpha_1 = 0$, it is an IRS$(1; m, n)$ model with a linear trend; when both $\alpha_0 = \alpha_1 = 0$, it is simply an IRS$(1; m, n)$ model without trend.

Note that if $s_t = 1$ for all t, then

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

which is a random walk with the drift term $\alpha_0 + \alpha_1$. If $s_t = 0$ 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 (with or without breaks) and a random walk with drift.

The model (2) can also be written as a special case of a general dynamic model with state-dependent coefficients. In particular, it can be shown that (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 $i = 2, \ldots, r$, 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 $i > n$. From (4) we can see that only the past state variables $s_{t-1}, \ldots, s_{t-r-1}$ can affect y_t and, as stated before, the concurrent state s_t is irrelevant to y_t, because both $s_t\varepsilon_t$ and $(1 - s_t)\varepsilon_t$ are present at time t. Thus, under the proposed model, the effect of a state variable s_t can only be revealed in subsequent periods $t+j$ with $j > 0$. Moreover, the model (2) has a state-space representation with switching parameters (see Eq. (8) below) and hence is a special case of the Markov-switching state space model considered by Kim (1994) and Kim and Nelson (1999).

Although Eq. (2) is a specialized model, its two-component structure is useful in practice, because it allows researchers to explicitly specify the dynamic patterns of their choice. Note that Eq. (2) is just one possibility of modeling economic time series under the general framework of Eq. (1). Many other interesting models can also be constructed as special cases of Eq. (1). For example, one may set y_t as the sum of a fractionally integrated component and a weakly stationary ARMA component, so that y_t may exhibit both long- and short-range dependence. It is worth noting that when a fractionally integrated component is present, there will be no finite-dimensional state-space representation. It is also possible to specify a different switching mechanism for s_t, cf. the threshold-disturbance moving-average model of Elwood (1998).

2.2 Comparison with Existing Models

To compare with the existing models that allow for different structures, we consider the IRS$(1; 1, 0)$ model without trend, which consists of a random walk component and a stationary AR(1) component with $\alpha_0 = \alpha_1 = 0$. That is, $y_t = y_{1t} + y_{0t}$ with

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

Table 1 gives the moving-average representation of $y_{1,t}$ and $y_{0,t}$ for $t = 1, \ldots, 7$ when $\{s_1, \ldots, s_7\} = \{0, 0, 1, 1, 1, 0, 0\}$ 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 to evolve like a random walk when $s_t = 1$. It can be seen that when s_t switches, the classification 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 IRS$(1; 1, 0)$ model (5) is conceptually different from that of Evans and Wachtel (1993) and McCulloch and Tsay (1994) which also allows for switching between a

Table 1: Moving-average representation of the IRS(1; 1, 0) model in Eq. (5) with $(s_1, \cdots, s_7) = (0, 0, 1, 1, 1, 0, 0)$.

	$y_{1,t}$	$y_{0,t}$
$t = 1$	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

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 white-noise random variables. This model differs from the proposed model in its switching mechanism. Here, y_t switches between two processes so that all past innovations must change accordingly. Such a switching mechanism affects the past but not future dynamics. It also induces drastic changes (jumps) in the time path of y_t, especially when t gets larger, given that a random walk wanders off quickly. It seems implausible to have a time series exhibiting frequent jumps while the magnitude of jump is increasingly large. By contrast, s_t of 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 the Evans-Wachtel 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.

From the random MA-coefficient representation (4), Eq. (5) can be expressed as

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

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) assumes $a_t = \exp(\alpha_t)$ 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, this 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 of 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 close to 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.

To illustrate, we simulate a random walk, the STOPBREAK process (6) with $\gamma = 1$, and the process (5) with $\psi_1 = 0$.[2] The simulated paths are plotted in four panels of Figure 2, 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. In particular, the process (5) may be quite different from the random walk and the STOPBREAK process, or it may move closely with the random walk and the STOPBREAK process.

[2] In our simulations, these processes are generated from the same set of innovations; the process (5) is such that $s_t = 1$ for those t indicated by the shaded boxes on the horizontal axis, and $s_t = 0$ otherwise.

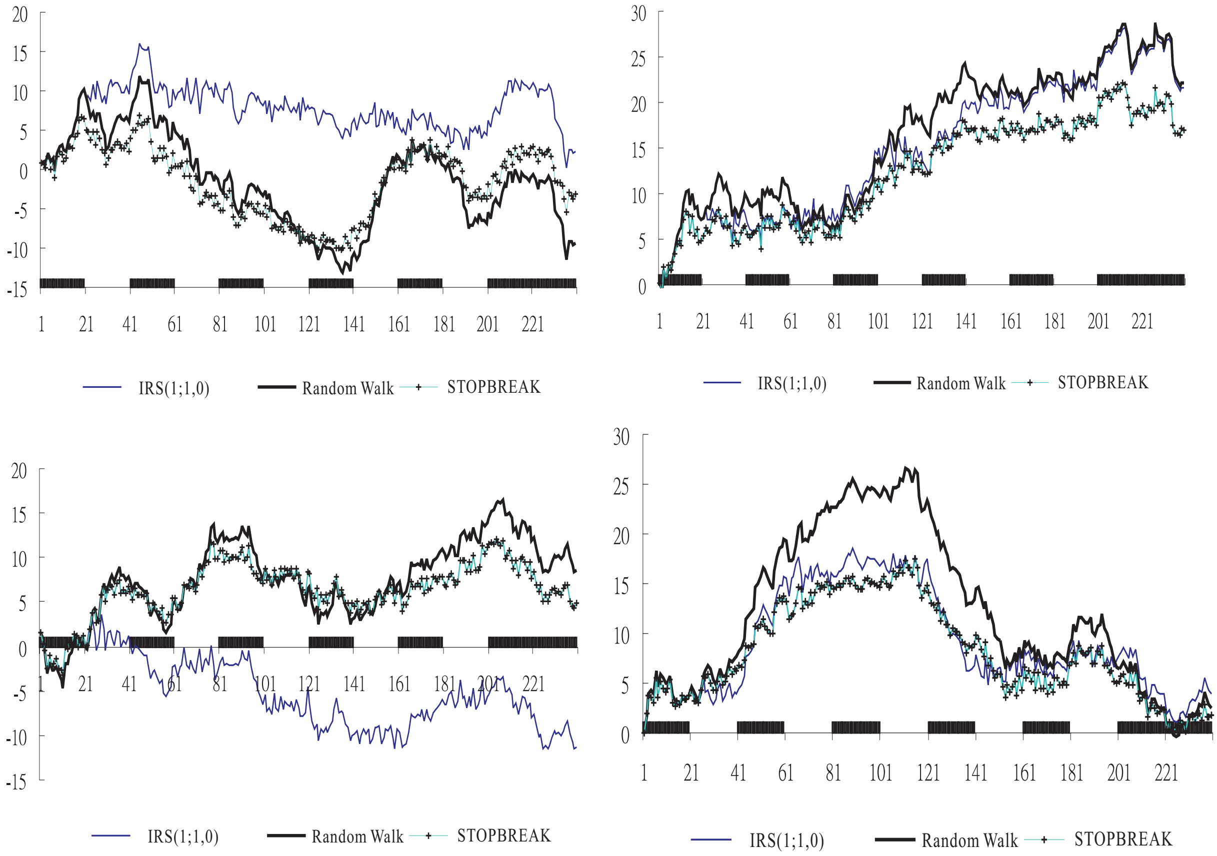

Figure 2: Examples of simulated random walk, STOPBREAK process, and IRS(1; 1, 0) process.

3 Properties of the Proposed IRS 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 $S^t = \{s_t, s_{t-1}, \ldots\}$ 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 $\operatorname{cov}(s_t\varepsilon_t, s_{t-i}\varepsilon_{t-i}) = \mathbb{E}[s_t s_{t-i}\,\mathbb{E}(\varepsilon_t\varepsilon_{t-i}|S^t)] = 0$. 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,

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

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

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

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, $\operatorname{var}(y_t)$ 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 $\pi_0>0$, $\operatorname{var}(y_t)$ 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 $\operatorname{var}(y_t)$ because the components of (3) are no longer mutually uncorrelated. To see this, first note that

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

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, $\operatorname{cov}(s_t, s_{t-j}\varepsilon_{t-j})$ are non-zero when $j \geq 1$. To ease our analysis of $\operatorname{var}(y_t)$, we assume for the time being that $\mathbb{E}(\varepsilon_t \mid S^\tau) = 0$ for $\tau > t$. Then, $\operatorname{cov}(s_t, s_{t-j}\varepsilon_{t-j}) = 0$ so that

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

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.

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 IRS(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

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

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 and Hypothesis Testing

4.1 Estimation of the State-Space Represntation

There are different ways to estimate the proposed IRS$(1; m, n)$ model. Following the suggestion of a referee, we adopt the approach of Kim (1994). We first put the proposed model into a Markov-switching state-space form and then derive an algorithm to compute the approximate quasi-maximum likelihood estimates (QMLE).

From the ARMA representation (4), we observe that the past $r+1$ state variables affect y_t, where $r = \max(m, n)$. Define s^*_{t-1} as the new state variable such that each of its values (in $\{1, 2, \ldots, 2^{r+1}\}$) represents a particular realization of $(s_{t-1}, \ldots, s_{t-r-1})$.[3] It can be seen that s^*_t also forms a first-order Markov chain with the transition matrix

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

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

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

To keep the exposition simple, we assume that the (i, j)th element of $\boldsymbol{P}^*$ is $p^*_{ij} = \mathbb{P}(s^*_t = j \mid s^*_{t-1} = i)$. Also let

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

denote the collection of the random MA coefficients in Eq. (4). When $s^*_{t-1} = j$, we write $\boldsymbol{\xi}_{t-1, s^*_{t-1}}$ as $\boldsymbol{\xi}_{t-1, j}$ so that

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

where $\mathbf{1}$ is the vector of ones, and the realizations of $s_{t-1}, \ldots, s_{t-r-1}$ are such that $s^*_{t-1} = j$. When $r = 2$ and $j = 3$, for example, the realization of $(s_{t-1}, s_{t-2}, s_{t-3})$ is $(0, 1, 0)$ so that $\boldsymbol{\xi}'_{t-1,3}\mathbf{1} = -(1 + \varphi_1) - \psi_2 + \varphi_3$.

[3]For example, when $r = 2$, s^*_{t-1} has 8 possible values: $s^*_{t-1} = 1$ if $s_{t-1} = s_{t-2} = s_{t-3} = 0$, $s^*_{t-1} = 2$ if $s_{t-1} = 0, s_{t-2} = 0$, and $s_{t-3} = 1$, …, $s^*_{t-1} = 8$ if $s_{t-1} = s_{t-2} = s_{t-3} = 1$.

It is now easy to show that the proposed IRS$(1;m,n)$ model can be expressed as a state-space model with the following measurement and transition equations:

$$\begin{aligned} z_t &= \mu_0 + \mu_1 \boldsymbol{\xi}'_{t-1,s^*_{t-1}} \mathbf{1} + \boldsymbol{H}_{s^*_{t-1}} \boldsymbol{\Gamma}_t, \\ \boldsymbol{\Gamma}_t &= \boldsymbol{F} \boldsymbol{\Gamma}_{t-1} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{8}$$

where $\mu_0 = \alpha_0 + \alpha_1 \Psi(1)^{-1}$ and $\mu_1 = \alpha_1 \Psi(1)^{-1}$ are two constant terms, $\boldsymbol{H}_{s^*_{t-1}} = (1\ \boldsymbol{\xi}'_{t-1,s^*_{t-1}})'$ is a $(r+2)$-dimensional row vector,

$$\boldsymbol{F} = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_r & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

is a $(r+2) \times (r+2)$ matrix such that $\psi_i = 0$ for $i > m$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)'$ denotes the shock of the model. This is also a Markov-switching state-space model discussed in Kim (1994).

We therefore follow Kim (1994) to derive an estimation algorithm which involves the Kalman filter and the Hamilton filter; see also Kim and Nelson (1999) for a thorough discussion. The derivation of our algorithm is similar, but not identical, to that of Kim (1994) because y_t here depends only on the past (but not the current) state variables; a detailed description of this algorithm is given in the Appendix. From the recursions of this algorithm we obtain the filtering probabilities $\mathbb{P}(s_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta})$, the smoothing probabilities $\mathbb{P}(s_t = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta})$, and an approximate quasi-log-likelihood function:

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

where $z_t = (1-B)y_t$ and $\mathcal{Z}^t = \{z_1, \ldots, z_t\}$. The approximate QMLE,

$$\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\psi}_1, \ldots, \hat{\psi}_m, \hat{\varphi}_1, \ldots, \hat{\varphi}_n, \hat{\sigma}_\varepsilon, \hat{p}_{00}, \hat{p}_{11})',$$

can then be found using a numerical-search method. Our program is written in GAUSS which employs the BFGS (Broyden-Fletcher-Goldfarb-Shanno) search algorithm. Plugging $\hat{\boldsymbol{\theta}}$ into the formulae of $\mathbb{P}(s_t = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta})$, we obtain the estimated smoothing probabilities.

4.2 Hypothesis Testing

It is well-known that many economic and financial time series contain a unit root. However, an interesting hypothesis is whether the data follow a random walk. This amounts to testing the null hypothesis $H_o : p_{11} = 1$ under the proposed model, i.e. $s_t = 1$ almost surely. 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, Lagrange multiplier (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 simulation, 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 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 3, 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 employ a simulation-based test to distinguish an ARIMA model from the proposed IRS model. For a given series, we first estimate an array of ARIMA$(p, 1, q)$ models and choose 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 IRS$(1; m, n)$ models and denote the selected model as IRS$(1; m^*, n^*)$ 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

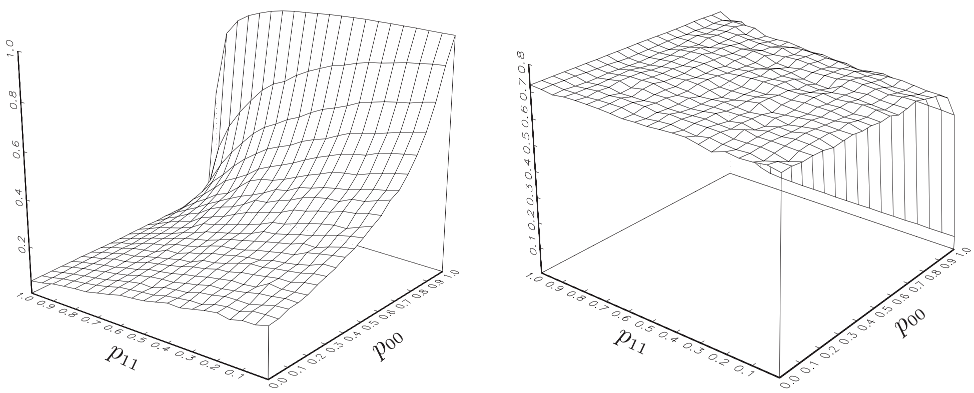

Figure 3: Empirical powers of the Dickey-Fuller test (left) and KPSS test (right).

simulated sample, we re-estimate the IRS$(1; m^*, n^*)$ model and obtain an estimate of p_{11}, denoted by $\hat{p}_{11}$. Replicating this procedure many times yields a finite-sample reference distribution of $\hat{p}_{11}$ on which we can compute the p-value 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%.

5 Empirical Study

To demonstrate the applicability of the proposed model, we apply the model (2) with a smooth Markov trend to U.S. quarterly 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 2002:I with 221 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 IRS$(1; m, n)$ models with a Markov trend and $0 \leq m, n \leq 4$. The parameters are estimated using the algorithm described in Section 4.1 and the Appendix. This algorithm is initialized by a broad range of random initial values. The covariance matrix of $\hat{\boldsymbol{\theta}}$ 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 IRS model.

Estimate	Estimate	Standard error	t-statistic
$\hat{\alpha}_0$	-0.00832	0.00306	-2.718^*
$\hat{\alpha}_1$	0.01989	0.00216	9.200^*
$\hat{\psi}_1$	0.24265	0.24931	0.973
$\hat{\psi}_2$	-0.51115	0.19204	-2.662^*
$\hat{\varphi}_1$	-0.28466	0.13282	2.143^*
$\hat{\varphi}_2$	-0.63296	0.22149	2.858^*
$\hat{\sigma}_\varepsilon$	0.00815	0.00244	3.340^*
$\hat{p}_{00}$	0.60647	0.11364	
$\hat{p}_{11}$	0.90882	0.03479	
Log-Likelihood=700.336			
AR roots: $0.121 \pm 0.704i$		AIC=-1382.67	
MA roots: $-0.142 \pm 0.783i$		SIC=-1352.13	

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

Among all the models considered, both AIC and SIC select the IRS$(1;2,2)$ model. The estimation results are summarized in Table 2; the estimated transition probabilities are $\hat{p}^*_{11} \approx 0.91$ and $\hat{p}^*_{00} \approx 0.61$. We first apply the simulation-based approach in Section 4.2 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:

$$\Delta y_t = 0.008 + 0.347\Delta y_{t-1} + u_t, \tag{9}$$

with $\sigma_u = 0.0097$, where $\Delta y_t = y_t - y_{t-1}$. We then re-estimate the IRS$(1;2,2)$ model using the data generated from Eq. (9) and obtain $\hat{p}_{11}$. Using 1000 replications we obtain a finite-sample reference distribution of $\hat{p}_{11}$. The p-value of $\hat{p}^*_{11} = 0.91$ based on this simulated distribution is about 0.034 and hence we reject the model in Eq. (9) at the 5% significance level.

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 29.77, 30.88, 32.07 and 35.91, 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 18.79 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. We also conduct some diagnostic checks on the estimated model, including the Q test of Ljung-Box (1978) on serial correlations and the LM test of Engle (1982) for ARCH effects. These tests are applied to the residual series $\hat{\varepsilon}_t$. The statistics are $Q(20) = 17.244$, $Q(30) = 25.490$, ARCH(2) $= 2.049$ and ARCH(4) $= 2.882$. They are all insignificant even at the 10% level, under the $\chi^2(20)$, $\chi^2(30)$, $\chi^2(2)$ and $\chi^2(4)$ distributions, respectively. These tests suggest that there are no significant serial correlations or conditional heteroskedasticity in residuals. The proposed model, thus, fits the GDP data well.

In Figure 4 we plot the estimated smoothing probabilities of $s_t = 0$, where the shaded areas signify the recession periods identified by NBER, and the solid (dashed) lines denote the peaks (troughs). We find that there are 33 periods (about 15.2% of the sample) with the estimated smoothing probability $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T; \hat{\boldsymbol{\theta}}) > 0.5$. This shows that unit-root nonstationarity is more likely to prevail in about 85% 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) closely. It is also worth mentioning that our model successfully identifies the recession period started in March 2001 which is at the end of the sample span. Hence, the innovations in expansion (recession) are more likely to have a permanent (transitory) effect. These results together 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 $\alpha_0 = -0.83\%$ during the state of transitory shocks (recessions) and $(\alpha_0 + \alpha_1) = 1.15\%$

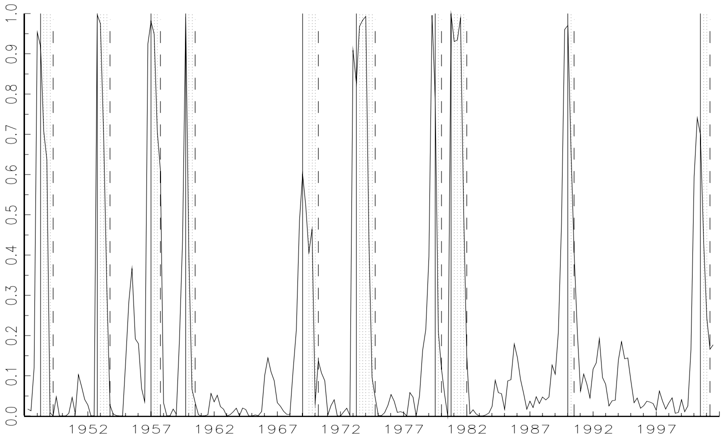

Figure 4: Estimated smoothing probabilities of $s_t = 0$ for U.S. quarterly real GDP from 1947.I to 2002.I.

during the state of permanent shocks (expansion). The expected durations of recession and expansion can be calculated from the transition probabilities: $1/(1-0.61) = 2.6$ quarters for recession and $1/(1-0.91) = 11$ quarters for expansion. According to NBER dating, the average growth rates for recession and expansion are, respectively, -0.35% and 1.10%, and the average durations are, respectively, 3.4 and 18.4 quarters. Comparing with NBER dating, our results lead to a shorter expected duration for expansion and "deeper" recessions. We also apply the model of Hamilton (1989) to the data set. The estimation Gauss program is taken from C. R. Nelson's web site and is initialized by 100 initial values.[4] 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 also compute the expected trend line (smooth Markov trend) as

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

and the expected stochastic trend component as $\sum_{i=1}^{t} \mathrm{I\!P}(s_i = 1 \mid \mathcal{Z}^T; \hat{\boldsymbol{\theta}})\hat{\varepsilon}_i$, both based on the estimated smoothing probabilities $\mathrm{I\!P}(s_i = 1 \mid \mathcal{Z}^T; \hat{\boldsymbol{\theta}})$. These two expected trend components together capture the behavior of log(GDP) quite well, as can be seen from Figure 5.

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

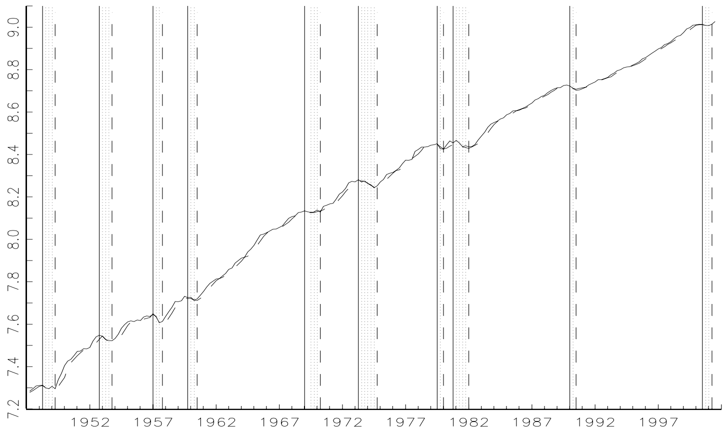

Figure 5: The expected trend line (dashed line) in U.S. real GDP.

6 Conclusions

In this paper, we proposed a class of unobserved-component models with switching permanent and transitory innovations. The model has 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 long-term effects of corresponding innovations thus may alternate from time to time. Third, it allows for breaks in the 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. Thus, the proposed models bridge the gap between trend-stationary and unit-root models and is able to accommodate both trend-reverting and trend-disturbing behaviors.

Application of the proposed model to U.S. quarterly real GDP suggests that it is a useful analytical tool in describing the characteristics of the data. In particular, it shows 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 expansion are more likely to be permanent. This result differs from that of unit-root (trend-stationary) models in that 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.

Appendix: Estimation Algorithm

To estimate the state-space representation (8) of the proposed model, we derive an algorithm following Kim (1994). We denote the expectation of the variable $\boldsymbol{X}_t$, conditional on the information available up to time s and the realized state values $s^*_{t-2} = i$ and $s^*_{t-1} = j$, as $\boldsymbol{X}^{(i,j)}_{t|s}$. For example, conditional on the information up to time t, the mean and variance of the transition component $\boldsymbol{\Gamma}_t$ are, respectively,

$$\begin{aligned}
\boldsymbol{\Gamma}^{(i,j)}_{t|t} &= \mathbb{E}\big(\boldsymbol{\Gamma}_t \mid s^*_{t-1} = j, s^*_{t-2} = i, \mathcal{Z}^t\big),\\
\boldsymbol{\Upsilon}^{(i,j)}_{t|t} &= \mathbb{E}\left[(\boldsymbol{\Gamma}_t - \boldsymbol{\Gamma}^{(i,j)}_{t|t})(\boldsymbol{\Gamma}_t - \boldsymbol{\Gamma}^{(i,j)}_{t|t})' \mid s^*_{t-1} = j, s^*_{t-2} = i, \mathcal{Z}^t\right].
\end{aligned}$$

We also let $\boldsymbol{X}^{(j)}_{t|t}$ denote the expectation of $\boldsymbol{X}_t$, conditional on the information up to time t and the realized state value $s^*_{t-1} = j$.

The Kalman filter for (8) consists of the following updating equations:

$$\begin{aligned}
\boldsymbol{\Gamma}^{(i,j)}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\Gamma}^{(i)}_{t-1|t-1},\\
\boldsymbol{\Upsilon}^{(i,j)}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\Upsilon}^{(i)}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q}^*,\\
\boldsymbol{\eta}^{(i,j)}_{t|t-1} &= z_t - \mu_0 - \mu_1\boldsymbol{\xi}'_{t-1,j}\mathbf{1} - \boldsymbol{H}_j\boldsymbol{\Gamma}^{(i,j)}_{t|t-1},\\
\boldsymbol{f}^{(i,j)}_{t|t-1} &= \boldsymbol{H}_j\boldsymbol{\Upsilon}^{(i,j)}_{t|t-1}\boldsymbol{H}'_j,\\
\boldsymbol{\Gamma}^{(i,j)}_{t|t} &= \boldsymbol{\Gamma}^{(i,j)}_{t|t-1} + \boldsymbol{\Upsilon}^{(i,j)}_{t|t-1}\boldsymbol{H}'_j\big[\boldsymbol{f}^{(i,j)}_{t|t-1}\big]^{-1}\boldsymbol{\eta}^{(i,j)}_{t|t-1},\\
\boldsymbol{\Upsilon}^{(i,j)}_{t|t} &= \big(\boldsymbol{I} - \boldsymbol{\Upsilon}^{(i,j)}_{t|t-1}\boldsymbol{H}'_j\big[\boldsymbol{f}^{(i,j)}_{t|t-1}\big]^{-1}\boldsymbol{H}_j\big)\boldsymbol{\Upsilon}^{(i,j)}_{t|t-1},
\end{aligned}\tag{10}$$

where $\boldsymbol{Q}^*$ is the variance-covariance matrix of $\boldsymbol{\varepsilon}_t$ which involves the unknown parameter σ^2_ε. Note that this algorithm calculates $(2^{r+1} \times 2^{r+1})$ forecasts of $(\boldsymbol{\Gamma}^{(i,j)}_{t|t}, \boldsymbol{\Upsilon}^{(i,j)}_{t|t})$ for each date t, corresponding to every possible value for i and j.

The (2^{r+1}) elements of $(\boldsymbol{\Gamma}^{(j)}_{t|t}, \boldsymbol{\Upsilon}^{(j)}_{t|t})$ can be obtained by taking weighted averages over all state values at $t-2$:

$$\begin{aligned}
\boldsymbol{\Gamma}^{(j)}_{t|t} &\approx \sum_{i=1}^{2^{r+1}} \frac{\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^t)}{\mathbb{P}(s^*_{t-1} = j \mid \mathcal{Z}^t)}\boldsymbol{\Gamma}^{(i,j)}_{t|t},\\
\boldsymbol{\Upsilon}^{(j)}_{t|t} &\approx \sum_{i=1}^{2^{r+1}} \frac{\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^t)}{\mathbb{P}(s^*_{t-1} = j \mid \mathcal{Z}^t)}\big\{\boldsymbol{\Upsilon}^{(i,j)}_{t|t} + \big(\boldsymbol{\Gamma}^{(j)}_{t|t} - \boldsymbol{\Gamma}^{(i,j)}_{t|t}\big)\big(\boldsymbol{\Gamma}^{(j)}_{t|t} - \boldsymbol{\Gamma}^{(i,j)}_{t|t}\big)'\big\},
\end{aligned}\tag{11}$$

where by the Bayes theorem, $\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^t)$ can be calculated as

$$\begin{aligned}
&\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^t)\\
&\quad = \frac{\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^{t-1})f(z_t \mid s^*_{t-2} = i, s^*_{t-1} = j, \mathcal{Z}^{t-1})}{f(z_t \mid \mathcal{Z}^{t-1})}.
\end{aligned}\tag{12}$$

The first term of the numerator in (12) can be computed easily as

$$\mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^{t-1}) = p^*_{ij}\, \mathbb{P}(s^*_{t-2} = i \mid \mathcal{Z}^{t-1}); \tag{13}$$

the second term of the numerator is

$$f(z_t \mid s^*_{t-2} = i, s^*_{t-1} = j, \mathcal{Z}^{t-1}) = (2\pi)^{-\frac{T}{2}} \left| \boldsymbol{f}^{(i,j)}_{t|t-1} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\eta}^{(i,j)\,\prime}_{t|t-1} \boldsymbol{f}^{(i,j)\;-1}_{t|t-1} \boldsymbol{\eta}^{(i,j)}_{t|t-1} \right\},$$

the denominator can be expressed as

$$f(z_t \mid \mathcal{Z}^{t-1}) = \sum_{i=1}^{2^{r+1}} \sum_{j=1}^{2^{r+1}} f(z_t \mid s^*_{t-2} = i, s^*_{t-1} = j, \mathcal{Z}^{t-1})\, \mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^{t-1}).$$

Finally, summing over the values of s^*_{t-2}, the filter in Eq. (12) becomes

$$\mathbb{P}(s^*_{t-1} = j \mid \mathcal{Z}^{t}) = \sum_{i=1}^{2^{r+1}} \mathbb{P}(s^*_{t-2} = i, s^*_{t-1} = j \mid \mathcal{Z}^{t}). \tag{14}$$

Thus, with the initial values $(\boldsymbol{\Gamma}^{(i)}_{r|r}, \boldsymbol{\Upsilon}^{(i)}_{r|r})$ and $\mathbb{P}(s^*_{r-1} = i \mid \mathcal{Z}^{r})$, we can iterate the Eqs. (10)–(12) to obtain $\mathbb{P}(s^*_{t-1} = i, s^*_t = j \mid \mathcal{Z}^{t})$ from Eq. (13) for $t = r, r+1, \ldots T$. Then for each t, the desired filtering probability is

$$\mathbb{P}(s_t = 1 \mid \mathcal{Z}^{t}) = \sum_{j} \sum_{i=1}^{2^{r+1}} \mathbb{P}(s^*_{t-1} = i, s^*_t = j \mid \mathcal{Z}^{t}), \tag{15}$$

where the first summation is taken over all j that associated with $s_t = 1$. Clearly, $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^{t}) = 1 - \mathbb{P}(s_t = 1 \mid \mathcal{Z}^{t})$.

Following Kim and Nelson (1999) we set the initial values $(\boldsymbol{\Gamma}^{(i)}_{r|r}, \boldsymbol{\Upsilon}^{(i)}_{r|r})$ to the unconditional mean and variance: $\boldsymbol{\Gamma}^{(i)}_{r|r} = \boldsymbol{0}$ and

$$\mathrm{vec}\big(\Upsilon^{(i)}_{r|r}\big) = (\boldsymbol{I} - \boldsymbol{F} \otimes \boldsymbol{F})^{-1} \mathrm{vec}\big(\boldsymbol{Q}^*\big),$$

where the "vec" operator stacks the columns of a matrix into a vector, $\boldsymbol{I}$ is the identity matrix, and $\otimes$ denotes the Kronecker product. We also follow Hamilton (1989, 1994) and set the initial value $\mathbb{P}(s^*_{r-1} \mid \mathcal{Z}^{r})$ to its limiting unconditional counterpart: the $(2^{r+1}+1)$th column of the matrix $(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, where

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

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

We also follow 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 information in the sample. Observe that

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

In the current context,

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

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

$$\begin{aligned}
&\mathbb{P}(s_t^* = i \mid \mathcal{Z}^T) \\
&\qquad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \mathbb{P}(s_t^* = i \mid s_{t+1}^* = j, \mathcal{Z}^{t+1}) \\
&\qquad = \sum_{j=1}^{2^{r+1}} \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)\ \frac{\mathbb{P}(s_{t+1}^* = j \mid s_t^* = i, \mathcal{Z}^{t+1})\ \mathbb{P}(s_t^* = i \mid \mathcal{Z}^{t+1})}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})} \\
&\qquad = \mathbb{P}(s_t^* = i \mid \mathcal{Z}^{t+1}) \sum_{j=1}^{2^{r+1}} \frac{p_{ij}^*\ \mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^T)}{\mathbb{P}(s_{t+1}^* = j \mid \mathcal{Z}^{t+1})}.
\end{aligned} \tag{16}$$

Using the filtering probability $\mathbb{P}(s_T^* = i \mid \mathcal{Z}^T)$ as the initial value we can iterate the equations (14)–(16) backward for $t = T-1, \cdots, p+1$. Consequently, the desired smoothing probability for each t is

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

where the summation is taken over all i that are associated with $s_t = 1$. We also have $\mathbb{P}(s_t = 0 \mid \mathcal{Z}^T) = 1 - \mathbb{P}(s_t = 1 \mid \mathcal{Z}^T)$.

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. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, **50**, 987–1008.

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.

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.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297–303.

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.