A new approach to model regime switching[☆]

Yoosoon Chang [a], Yongok Choi [b], Joon Y. Park [a,c,*]

[a] *Department of Economics, Indiana University, United States*
[b] *Department of Financial Policy, Korea Development Institute, Republic of Korea*
[c] *Department of Economics, Sungkyunkwan University, Republic of Korea*

A B S T R A C T

This paper introduces a new approach to model regime switching using an autoregressive latent factor, which determines regimes depending upon whether it takes a value above or below some threshold level. In our approach, the latent factor is allowed to be correlated with the innovation to the observed time series. If the latent factor becomes exogenous, our approach reduces to the conventional Markov switching. We develop a modified Markov switching filter to estimate the mean and volatility models with Markov switching that are frequently analyzed, and find that the presence of endogeneity in regime switching is indeed strong and ubiquitous.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Regime switching models have been used extensively in econometric time series analysis. In most of these models, two regimes are introduced with a state process determining one of the regimes to take place in each period. The bivalued state process is typically modeled as a Markov chain. The autoregressive model with this type of Markov switching in the mean was first considered by Hamilton (1989), and was further analyzed in Kim (1994). Subsequently, Markov switching has been introduced in a more general class of models such as regression models and volatility models by numerous authors. Moreover, various statistical properties of the model have been studied by Hansen (1992); Hamilton (1996); Garcia (1998); Timmermann (2000), and Cho and White (2007), among others. For an introduction and overview of the related literature, the reader is referred to the monograph by Kim and Nelson (1999). Markov-switching models with endogenous explanatory variables have also been considered recently by Kim (2004, 2009).

Though Markov switching models have been used and proven to be useful in a wide range of contexts, they have some drawbacks. Most importantly, with a very few exceptions including Diebold et al. (1994) and Kim et al. (2008),[1] they all assume that the Markov

---

---

[1] Diebold et al. (1994) considers a Markov-switching driven by a set of observed variables, while our approach is based on a latent factor. The approach in Kim et al. (2008) is more closely related to our approach. See also Kalliovirta et al. (2015) for a related approach.

chain determining regimes is completely independent from all other parts of the model, which is extremely unrealistic in many cases.[2] Note that exogenous regime switching in particular implies that future transitions between states are completely determined by the current state, and does not rely on the realizations of underlying time series. This is highly unlikely in many practical applications. Instead, we normally expect that future transitions depend critically on the realizations of underlying time series as well as the current and possibly past states. Furthermore, the Markov chain determining the state of regime in virtually all of the existing switching models is assumed to be strictly stationary, and cannot accommodate nonstationarity in the transition probability. This can be restrictive if the transition is strongly persistent.

In this paper, we propose a novel approach to modeling regime switching. In our approach, the mean or volatility process is switching between two regimes, depending upon whether the underlying autoregressive latent factor takes values above or below some threshold level. The innovation of the latent factor, on the other hand, is assumed to be correlated with the previous innovation in the model. A current shock to the observed time series therefore affects the regime switching in the next period. Moreover, we allow the autoregressive latent factor to have a unit root and accommodate a strongly persistent regime change. Consequently, our approach provides remedies to both of the aforementioned shortcomings in conventional Markov switching models, and yields a broad class of models with endogenous and possibly nonstationary regime changes. Moreover, the latent factor involved in our approach can be estimated and used to investigate the dynamic interactions of the mean or volatility process of a given time series with the levels of other observed time series. Our model can be estimated by a modified Markov switching filter that we develop in the paper.

If the autoregressive latent factor is exogenous and stationary, the regime switching based on our approach reduces to the conventional Markov switching. In this case, the conventional two state Markov switching specified by two transition probabilities has the exact one-to-one correspondence with our regime switching specified by the autoregressive coefficient of the latent factor and the threshold level. Therefore, we may always find our regime switching model with an exogenous autoregressive latent factor corresponding to a conventional two state Markov switching model. They are observationally equivalent and have exactly the same likelihood. Consequently, our model may be regarded as a natural extension of the conventional Markov switching model, with the extension made to relax some of its important restrictive features. In the presence of endogeneity, however, our model diverges sharply from the conventional Markov switching model. In particular, we show that the state process in our model is given by a Markov process jointly with the underlying time series, and the transition of state systematically interacts with the realizations of underlying time series.

To evaluate the performance of our model and estimation procedure, we conduct an extensive set of simulations. Our simulation results can be summarized as follows. First, the endogeneity of regime switching, if ignored, has a significantly deleterious effect on the estimates of model parameters and transition probabilities. This is more so for the mean model than for the volatility model, and for the models with stationary latent factors relative to the models with nonstationary latent

factors. Second, the presence of endogeneity, if taken into account properly, improves the efficiency of parameter estimates and the precision of estimated transition probabilities. This is because the presence of endogeneity helps to extract more information from the data on the latent regimes and their transitions. The efficiency gain and the precision enhancement are substantial in some cases, particularly when the latent factor is stationary and endogeneity is strong. Finally, the likelihood ratio tests for endogeneity work reasonably well in all cases we consider. Though they tend to overreject the null hypothesis of no endogeneity for relatively small samples, they overall appear to be very powerful. In fact, their power increases sharply up to unity as the degree of endogeneity increases.

As an empirical illustration of our approach, we analyze the US GDP growth rates and the NYSE/AMEX index returns, respectively, for our mean and volatility models. For both models, the evidence of endogeneity is unambiguously strong. The estimated correlations between the current shock to the observed time series and the latent factor determining the state in the next period are all significantly different from zero. In our volatility model, the correlation is estimated to be strongly negative with the values -0.970 and -0.999 for the two sample periods we consider. Such a large negative correlation implies that a negative shock to stock returns in the current period is very likely to entail an increase in volatility in the next period, and provides an evidence for the presence of strong leverage effects in stock returns. On the other hand, the correlation in our mean model is estimated to be strongly negative with the value -0.923 for the earlier sample period considered in Kim and Nelson (1999) and has nearly perfect positive correlation for the recent subsample. The negative correlation in our stationary mean model implies that the mean reversion of the observed time series occurs at two different levels. Not only does the observed time series revert to its state dependent mean at the first level, but also the state dependent mean itself moves to offset the effect of a shock to the observed time series at the second level. In contrast, the positive correlation entails an unstabilizing movement of the state dependent mean at the second level. For both mean and volatility models, the inferred probabilities appear to be more accurately predicting the true regimes if we allow for endogeneity in regime switching.

The rest of the paper is organized as follows. In Section 2, we introduce our model and compare it with the conventional Markov switching model. In particular, we show that our model becomes observationally equivalent to the conventional Markov switching model, if endogeneity is not present. Section 3 explains how to estimate our model using a modified Markov switching filter. The Markov property of the state process is also discussed in detail. Section 4 reports our simulation studies, which evaluate the performance of our model relative to the conventional Markov switching model. The empirical illustrations in Section 5 consist of the analysis of the US GDP growth rates and the NYSE/AMEX index returns, using our mean and volatility models respectively. Section 6 concludes the paper, and Appendix collects the proofs of theorems and additional figures.

A note on notation. We denote respectively by φ and Φ the density and distribution functions of the standard normal distribution. The equality in distribution is written as $=_d$. Moreover, we use $p(\cdot)$ or $p(\cdot|\cdot)$ as the generic notation for density or conditional density function. Finally, $\mathbb{N}(a, b)$ signifies the density of normal distribution, or normal distribution itself, with mean a and variance b.

2. Models with endogenous regime switching

In this section, we introduce an approach to model endogenous regime switching and compare it with the approach used in the conventional Markov switching model.

---

[2] This is true only for the Markov switching models analyzed by the frequentist approach. In the literature on Bayesian regime switching models, Chib (1996), Chib and Dueker (2004), and Bazzi et al. (2014), among others, allow for endogenous Markov chains. See also Kang (2014), which extends Kim et al. (2008) to a general state space model using a Bayesian approach.

2.1. A new regime switching model

In our model, we let a latent factor (w_t) be generated as an autoregressive process

$$w_t = \alpha w_{t-1} + v_t \qquad (1)$$

for $t = 1, 2, \ldots$, with parameter $\alpha \in (-1, 1]$ and i.i.d. standard normal innovations (v_t). We use (π_t) as a generic notation to denote a state dependent parameter taking values $\pi_t = \underline{\pi}$ or $\bar{\pi}$, $\underline{\pi} < \bar{\pi}$, depending upon whether we have $w_t < \tau$ or $w_t \geq \tau$ with τ being a threshold level, or more compactly,

$$\pi_t = \pi(w_t) = \underline{\pi} \mathbf{1}\{w_t < \tau\} + \bar{\pi} \mathbf{1}\{w_t \geq \tau\}, \qquad (2)$$

where τ and $(\underline{\pi}, \bar{\pi})$ are parameters, $\pi : \mathbb{R} \to \{\underline{\pi}, \bar{\pi}\}$, and $\mathbf{1}\{\cdot\}$ is the indicator function. In subsequent discussion of our models, we interpret the two events $\{w_t < \tau\}$ and $\{w_t \geq \tau\}$ as two regimes that are switching by the realized value of the latent factor (w_t) and the level τ of the threshold.

To compare our model with the conventional Markov switching model, we may set

$$s_t = \mathbf{1}\{w_t \geq \tau\}, \qquad (3)$$

so that we have

$$\pi_t = \pi(s_t) = \underline{\pi}(1 - s_t) + \bar{\pi} s_t$$

exactly as in the conventional Markov switching model.[3] The state process (s_t) represents the low or high state depending upon whether it takes value 0 or 1. The conventional Markov switching model simply assumes that (s_t) is a Markov chain taking value either 0 or 1, whereas our approach introduces an autoregressive latent factor (w_t) to define the state process (s_t). In the conventional Markov switching model, (s_t) is assumed to be completely independent of the observed time series. Contrastingly, in our approach, it will be allowed to be endogenous, which appears to be much more realistic in a wide range of models used in practical applications.

For the identification of parameters $\underline{\pi}$ and $\bar{\pi}$ in (2), we need to assume that $\underline{\pi} < \bar{\pi}$. To see this, note that (v_t) has the same distribution as $(-v_t)$, and that our level function is invariant with respect to the joint transformation $w \mapsto -w$, $\tau \mapsto -\tau$ and $(\underline{\pi}, \bar{\pi}) \mapsto (\bar{\pi}, \underline{\pi})$. Recall also that, to achieve identification of our level function, we must restrict the variance of the innovations (v_t) to be unity. This is because, for any constant $c > 0$, (cv_t) generates (cw_t) and our level function remains unchanged under the joint transformation $w \mapsto cw$ and $\tau \mapsto c\tau$ in scale. If $\alpha = 1$ and the latent factor (w_t) becomes a random walk, we have an additional issue of joint identification for the initial value w_0 of (w_t) and the threshold level τ. In this case, we have $w_t = w_0 + \sum_{i=1}^{t} v_i$ for all t and the transformation $w_0 \mapsto w_0 + c$ for any constant c yields $(w_t + c)$ in place of (w_t). However, our level function does not change under the joint transformation $w \mapsto w + c$ and $\tau \mapsto \tau + c$ in location. Therefore, we set $w_0 = 0$ in this case. On the other hand, the identification problem of the initial value w_0 of (w_t) does not arise if we assume $|\alpha| < 1$. Under this assumption, the latent factor (w_t) becomes asymptotically stationary, and we set

$$w_0 =_d \mathbb{N}\left(0, \frac{1}{1 - \alpha^2}\right)$$

to make (w_t) a strictly stationary process.

We specify our model as

$$
\begin{aligned}
y_t &= m(x_t, y_{t-1}, \ldots, y_{t-k}, w_t, \ldots, w_{t-k}) + \sigma(x_t, w_t, \ldots, w_{t-k})u_t \\
&= m(x_t, y_{t-1}, \ldots, y_{t-k}, s_t, \ldots, s_{t-k}) + \sigma(x_t, s_t, \ldots, s_{t-k})u_t
\end{aligned}
\qquad (4)
$$

with mean and volatility functions m and σ respectively, where (x_t) is exogenous and (u_t) and (v_t) in (1) are jointly i.i.d. and distributed as

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} =_d \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \qquad (5)$$

with unknown parameter ρ.[4] For the brevity of our subsequent exposition, we write

$$
\begin{aligned}
m_t &= m(x_t, y_{t-1}, \ldots, y_{t-k}, w_t, \ldots, w_{t-k}) \\
&= m(x_t, y_{t-1}, \ldots, y_{t-k}, s_t, \ldots, s_{t-k})
\end{aligned}
\qquad (6)
$$

$$\sigma_t = \sigma(x_t, w_t, \ldots, w_{t-k}) = \sigma(x_t, s_t, \ldots, s_{t-k}). \qquad (7)$$

Note that m_t and σ_t are the conditional mean and volatility of the state dependent variable (y_t) given present and past values of latent factors $w_t, \ldots, w_{t-k}$, as well as the current values of exogenous variables x_t and lagged endogenous variables $y_{t-1}, \ldots, y_{t-k}$.[5]

Our model (4) includes as special cases virtually all models considered in the literature. In our simulations and empirical illustrations, we mainly consider the model

$$\gamma(L)(y_t - \mu_t) = \sigma_t u_t, \qquad (8)$$

where $\gamma(z) = 1 - \gamma_1 z - \cdots - \gamma_k z^k$ is a kth order polynomial, $\mu_t = \mu(w_t) = \mu(s_t)$ and $\sigma_t = \sigma(w_t) = \sigma(s_t)$ are the state dependent mean and volatility of (y_t) respectively. We may easily see that the model introduced in (8) is a special case of our general model (4). The model describes an autoregressive process with conditional mean and volatility that are state dependent. It is exactly the same as the conventional Markov switching model considered by Hamilton (1989) and many others, except that the regimes in our model (8) are determined by an endogenous latent autoregressive factor (w_t) specified as in (1). In fact, it turns out that if we set $\rho = 0$, together with $|\alpha| < 1$, our model in (8) becomes observationally equivalent to the conventional Markov switching model. This is shown below.

The model given in (8) may therefore be viewed as an extension of the conventional autoregressive Markov switching model, which allows in particular for endogeneity and nonstationarity in regime changes. The autoregressive parameter α of the latent factor (w_t) in (1) controls the persistency of regime changes. In particular, if $\alpha = 1$, the regime change driven by (w_t) becomes nonstationary, and such a specification may be useful in describing regime changes that are highly persistent. On the other hand, the parameter ρ in the joint distribution (5) of the current model innovation u_t and the next period shock v_{t+1} to the latent factor determines the endogeneity of regime changes. As ρ approaches to unity in modulus, the endogeneity of regime change driven by (w_t) becomes stronger, i.e., the determination of the regime in time $t + 1$ is more strongly influenced by the realization of innovation (u_t) at time t.

The interpretation of the endogeneity parameter ρ, especially its sign, is quite straightforward in our volatility model. If $\rho < 0$, the innovation u_t of y_t at time t becomes negatively correlated with the volatility σ_{t+1} of y_{t+1} at time $t + 1$. This implies that a negative

---

[3] Chib and Dueker (2004) consider the model given by (1) and (3) and interpret (w_t) as representing the strengths of regimes. They contrast this model with the model given by $w_t = \alpha s_{t-1} + v_t$ in which the transition is determined entirely by the previous regime itself, rather than its strength.

[4] We may equivalently define the dynamics of (w_t) as $w_t = \alpha w_{t-1} + \rho u_{t-1} + \sqrt{1 - \rho^2} v_t$, where (u_t) and (v_t) are independent i.i.d. standard normals.

[5] The endogenous latent factor (w_t) may naturally be regarded as an economic fundamental determining the regimes of an economy.

shock to (y_t) in the current period entails an increase in volatility in the next period. This is often referred to as the leverage effect if, in our model, (y_t) denotes returns from a financial asset. See, e.g., Yu (2005) for more discussions on how we should model the leverage effect in financial asset returns. Of course, $\rho > 0$ means that there is an anti-leverage effect in the model.

For the mean model, the sign of ρ has a more subtle effect on the sample path of the observed time series (y_t). If the lag polynomial $\gamma(z)$ satisfies the stationarity condition, (y_t) becomes stationary. In this case, (y_t) reverts to its state dependent mean (μ_t), as well as to its global mean $\mathbb{E}y_t$. This is true for both, the cases of $\rho < 0$ and of $\rho > 0$. The mean reverting behavior of (y_t), however, differs depending upon whether $\rho < 0$ or $\rho > 0$. If $\rho < 0$, a positive realization of u_t at time t increases the probability of having a low regime in the state dependent mean μ_{t+1} of y_{t+1} at time $t + 1$, and in this sense, the state dependent mean (μ_t) of the observed time series (y_t) is also reverting. Therefore, the mean reversion of (y_t) takes place at two distinct levels: the reversion of (y_t) to its state dependent mean (μ_t) at the first level, and the movement of (μ_t) to offset the effect of a shock to (y_t) at the second level. This would not be the case if $\rho > 0$. In this case, the movement of (μ_t) at the second level would entail an unstabilizing effect on (y_t). Furthermore, in the cases of $\rho > 0$ and of $\rho < 0$, a regime switching is more likely to happen at time $t + 1$ if y_t takes a value inside and outside of the two state dependent means at time t, respectively. Therefore, in the mean model, regime switching can be more clearly seen when $\rho < 0$.

Kim et al. (2008) consider a regression model similar to our model in (4), yet they set the state dependent regression coefficients (β_t) to be dependent only on the current state variable (s_t). Therefore, their model is more restrictive than our model in (8). Moreover, in their model, the state process is defined as $s_t = 1\{v_t \geq \pi_{t-1}\}$, where (v_t) is specified simply as a sequence of i.i.d. latent random variables that is contemporaneously correlated with innovation (u_t) in regression error $(\sigma_t u_t)$.[6] Though their state process (s_t) is endogenous, it is strictly restricted to be first order Markovian and stationary as in the conventional Markov switching model.[7] Furthermore, in their approach, (u_t) is jointly determined with (s_t) for each time t. The presence of contemporaneous correlation between (u_t) and (s_t) entails undesirable consequences for their model: State dependent coefficients of regressors (β_t) are contemporaneously correlated with regression errors $(\sigma_t u_t)$, as well as regression errors $(\sigma_t u_t)$ are serially correlated.[8] Their regression model is therefore seriously misspecified from a conventional point of view. Contrastingly, in our model, innovation (u_t) affects the transition of (s_t) only in the next period, and therefore, (s_t) becomes pre-determined in this sense. Modeling endogeneity as in our model yields a model that is correctly specified as a conventional regression model. This is critical to interpret (m_t) and (σ_t) respectively as the conditional mean and volatility of (y_t) in (4).

2.2. Relationship with conventional markov switching models

Our model reduces to the conventional Markov switching model when the underlying autoregressive latent factor is stationary and independent of the model innovation. This will be explored below. In what follows, we assume

$$\rho = 0$$

to make our models more directly comparable to the conventional Markov switching models, and obtain the transition probabilities of the Markovian state process (s_t) defined in (3). In our approach, they are given as functions of the autoregressive coefficient α of the latent factor and the level τ of threshold. Note that

$$\mathbb{P}\{s_t = 0 | w_{t-1}\} = \mathbb{P}\{w_t < \tau | w_{t-1}\} = \Phi(\tau - \alpha w_{t-1}) \qquad (9)$$

$$\mathbb{P}\{s_t = 1 | w_{t-1}\} = \mathbb{P}\{w_t \geq \tau | w_{t-1}\} = 1 - \Phi(\tau - \alpha w_{t-1}). \qquad (10)$$

Therefore, if we let $|\alpha| < 1$ and denote the transition probabilities of the state process (s_t) from the low state to the low state and from the high state to the high state by

$$a(\alpha, \tau) = \mathbb{P}\{s_t = 0 | s_{t-1} = 0\},$$
$$b(\alpha, \tau) = \mathbb{P}\{s_t = 1 | s_{t-1} = 1\}, \qquad (11)$$

then it follows that

Lemma 2.1. *For $|\alpha| < 1$, we have*

$$a(\alpha, \tau) = \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right) \varphi(x)dx}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}$$

$$b(\alpha, \tau) = 1 - \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right) \varphi(x)dx}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

where $a(\alpha, \tau)$ and $b(\alpha, \tau)$ are defined in (11).

In particular, the state process (s_t) defined in (3) is a Markov chain on $\{0, 1\}$ with transition density

$$p(s_t | s_{t-1}) = (1 - s_t)\omega(s_{t-1}) + s_t[1 - \omega(s_{t-1})], \qquad (12)$$

where $\omega(s_{t-1})$ is transition probability to the low state given by

$$\omega(s_{t-1})$$
$$= \frac{\left[(1 - s_{t-1})\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} + s_{t-1}\int_{\tau\sqrt{1-\alpha^2}}^{\infty}\right] \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right) \varphi(x)dx}{(1 - s_{t-1})\Phi\left(\tau\sqrt{1-\alpha^2}\right) + s_{t-1}\left[1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)\right]}$$

with respect to the counting measure on $\{0, 1\}$.

The contours of the transition probabilities $a(\alpha, \tau)$ and $b(\alpha, \tau)$ obtained in Lemma 2.1 are presented in Fig. 1 for various levels of $0 < a(\alpha, \tau) < 1$ and $0 < b(\alpha, \tau) < 1$. Fig. 1 provides the contours of $a(\alpha, \tau)$ and $b(\alpha, \tau)$ in the (α, τ)-plane with $-1 < \alpha < 1$ and $-\infty < \tau < \infty$ for the levels of $a(\alpha, \tau)$ and $b(\alpha, \tau)$ starting from 0.05 to 0.95 in increment of 0.05. It is quite clear from Fig. 1 that there exists a unique pair of α and τ values yielding any given levels of $a(\alpha, \tau)$ and $b(\alpha, \tau)$, since any contour of $a(\alpha, \tau)$ intersects with that of $b(\alpha, \tau)$ once and only once. For instance, the only pair of α and τ values that yields $a(\alpha, \tau) = b(\alpha, \tau) = 1/2$ is given by $\alpha = 0$ and $\tau = 0$, in which case we have entirely random switching from the high state to the low state and vice versa with equal probability.

To more clearly demonstrate the one-to-one correspondence between the pair (α, τ) of parameters and the pair $(a(\alpha, \tau), b(\alpha, \tau))$ of transition probabilities derived in Lemma 2.1, we show how we may find the corresponding values of α and τ when the values of $a(\alpha, \tau)$ and $b(\alpha, \tau)$ are given. In Fig. 2, we set $a(\alpha, \tau) = 0.796$ and $b(\alpha, \tau) = 0.901$, the transition probabilities we obtain from the Hamilton's model for US GDP growth rates estimated over the period 1952–1984, and plot their contours in the

---

[6] For an easier comparison, we present their model using our notation introduced earlier in this section. Their model also includes other predetermined variables, which we ignore here to more effectively contrast their approach with ours. Our model may also easily accommodate the presence of other covariates.

[7] A referee believes that the required extension is possible and straightforward.

[8] Note also that (σ_t) does not represent the conditional volatility of their error process $(\sigma_t u_t)$, since (σ_t) is contemporaneously correlated with (u_t).
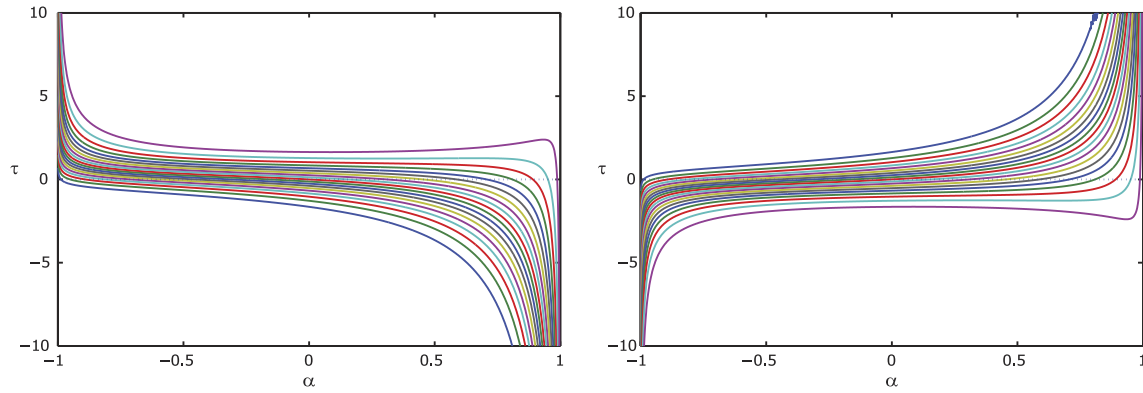
Fig. 1. Contours of transition probabilities in (α, τ)-plane. Notes: The contours of $a(\alpha, \tau)$ and $b(\alpha, \tau)$ are presented respectively in the left and right panels for the levels from 0.05 to 0.95 in increment of 0.05, upward for $a(\alpha, \tau)$ and downward for $b(\alpha, \tau)$. Hence the top line in the left panel is the contour of $a(\alpha, \tau) = 0.95$, and the bottom line on the right panel represents the contour of $b(\alpha, \tau) = 0.95$.
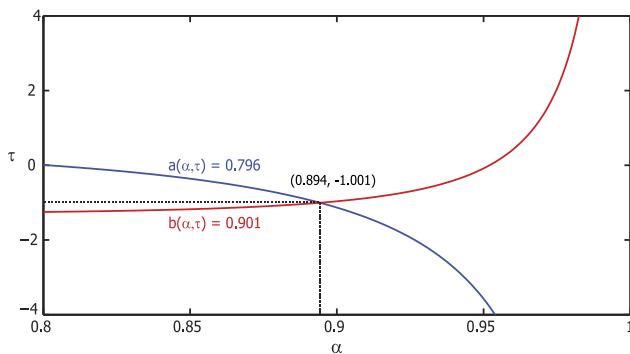
Fig. 2. Correspondence between (α, τ) and $(a(\alpha, \tau), b(\alpha, \tau))$. Notes: The increasing and decreasing curves are, respectively, the contours of $a(\alpha, \tau) = 0.796$ and $b(\alpha, \tau) = 0.901$ in the (α, τ)-plane. Their intersection at $(\alpha, \tau) = (0.894, -1.001)$ provides the (α, τ)-pair that yields the transition probabilities $a(\alpha, \tau) = 0.796$ and $b(\alpha, \tau) = 0.901$.

(α, τ)-plane. It is shown that the two contours intersect at one and only one point, which is given by $\alpha = 0.894$ and $\tau = -1.001$.

If we set $\rho = 0$ in our model (4), the transition probabilities of the state process (s_t) in (3) alone completely determine the regime switching without any interaction with other parts of the model. This implies that by setting $\rho = 0$ and obtaining the values of α and τ corresponding to the given values of $a(\alpha, \tau)$ and $b(\alpha, \tau)$, we may always find a regime switching model with an autoregressive latent factor that is observationally equivalent to any given conventional Markov switching model. Our approach, however, produces an important by-product that is not available from the conventional approach: an extracted time series of the autoregressive latent factor driving the regime switching.

Now we let $\alpha = 1$. In this case, the state process (s_t) defined in (3) becomes nonstationary and its transition evolves with time t. For $t \geq 1$, we subsequently define the transition probabilities explicitly as functions of time as

$$a_t(\tau) = \mathbb{P}\{s_t = 0 | s_{t-1} = 0\}, \quad b_t(\tau) = \mathbb{P}\{s_t = 1 | s_{t-1} = 1\}, \quad (13)$$

and show that

Corollary 2.2. *Let* $\alpha = 1$, *and let* $a_t(\tau)$ *and* $b_t(\tau)$ *be defined as in* (13). *For* $t = 1$, $a_1(\tau) = \Phi(\tau)$ *with* $\mathbb{P}\{s_0 = 0\} = 1$ *if* $\tau > 0$, *and,* $b_1(\tau) = 1 - \Phi(\tau)$ *with* $\mathbb{P}\{s_0 = 1\} = 1$ *if* $\tau \leq 0$. *Moreover, we have*

$$a_t(\tau) = \frac{\int_{-\infty}^{\tau/\sqrt{t-1}} \Phi\left(\tau - x\sqrt{t-1}\right) \varphi(x) dx}{\Phi\left(\tau/\sqrt{t-1}\right)}$$

$$b_t(\tau) = 1 - \frac{\int_{\tau/\sqrt{t-1}}^{\infty} \Phi\left(\tau - x\sqrt{t-1}\right) \varphi(x) dx}{1 - \Phi\left(\tau/\sqrt{t-1}\right)}$$

for $t \geq 2$.

The state process (s_t) is a Markov chain with transition density $p(s_t | s_{t-1})$ in (12) which is defined now with the transition probability to the low state $\omega(s_{t-1})$ given by

$$\omega(s_{t-1})$$
$$= \frac{\left[(1 - s_{t-1})\int_{-\infty}^{\tau/\sqrt{t-1}} + s_{t-1}\int_{\tau/\sqrt{t-1}}^{\infty}\right] \Phi\left(\tau - x\sqrt{t-1}\right) \varphi(x) dx}{(1 - s_{t-1})\Phi\left(\tau/\sqrt{t-1}\right) + s_{t-1}\left[1 - \Phi\left(\tau/\sqrt{t-1}\right)\right]}$$

with respect to the counting measure on $\{0, 1\}$. We may easily see that

$$a_t(\tau), b_t(\tau) \approx 1 - \frac{1}{\pi\sqrt{t-1}}$$

for large t, where π is a mathematical constant given by $\pi = 3.14159\ldots$, and therefore, the transition becomes more persistent in this case as t increases. Moreover, the threshold parameter τ is unidentified asymptotically. For the asymptotic identifiability of the threshold parameter when $\alpha = 1$, we must set $\tau = \bar{\tau}\sqrt{n}$ for some fixed $\bar{\tau}$. This is obvious because in this case the latent factor (w_t) increases stochastically at the rate $\sqrt{n}$.

3. Estimation

Our endogenous regime switching model (4) can be estimated by the maximum likelihood method. For the maximum likelihood estimation of our model, we write the log-likelihood function as

$$\ell(y_1, \ldots, y_n) = \log p(y_1) + \sum_{t=2}^{n} \log p(y_t | \mathcal{F}_{t-1}) \quad (14)$$

where $\mathcal{F}_t = \sigma\left(x_t, (y_s)_{s \leq t}\right)$, i.e., the information given by $x_t, y_1, \ldots, y_t$ for each $t = 1, \ldots, n$. Of course, the log-likelihood function includes a vector of unknown parameters $\theta \in \Theta$, say, which specifies the conditional mean and volatility processes (m_t) and (σ_t) of our model. It is, however, suppressed for the sake of notational brevity. The maximum likelihood estimator $\hat{\theta}$ of θ is given by

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, \ell(y_1, \ldots, y_n)$$

as usual. For the model (8), θ consists of state dependent mean and volatility parameters, $(\underline{\mu}, \bar{\mu})$ and $(\underline{\sigma}, \bar{\sigma})$, as well as the threshold τ level, the autoregressive coefficient α of the latent factor,

the correlation coefficient ρ, and the autoregressive coefficients $(\gamma_1, \ldots, \gamma_k)$.

To estimate our general switching model (4) by the maximum likelihood method, we develop a modified Markov switching filter. The conventional Markov switching filter is no longer applicable, since the state process (s_t) defined in (3) for our model is not a Markov chain unless $\rho = 0$. To develop the modified Markov switching filter that can be used to estimate our model, we let

$$\Phi_\rho(x) = \Phi\left(x/\sqrt{1-\rho^2}\right) \tag{15}$$

for $|\rho| < 1$ and

$$u_t = \frac{y_t - m_t}{\sigma_t}.$$

We have

Theorem 3.1. Let $|\rho| < 1$. The bivariate process (s_t, y_t) on $\{0, 1\} \times \mathbb{R}$ is a $(k+1)$-st order Markov process, whose transition density with respect to the product of the counting and Lebesgue measure is given by

$$p(s_t, y_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= p(y_t | s_t, \ldots, s_{t-k}, y_{t-1}, \ldots, y_{t-k})$$
$$\times p(s_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}),$$

where

$$p(y_t | s_t, \ldots, s_{t-k}, y_{t-1}, \ldots, y_{t-k}) = \mathbb{N}\left(m_t, \sigma_t^2\right) \tag{16}$$

and

$$p(s_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= (1-s_t)\omega_\rho + s_t(1-\omega_\rho) \tag{17}$$

with the transition probability $\omega_\rho = \omega_\rho(s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$ of (s_t) to the low state conditional on the previous states and the past values of observed time series. If $|\alpha| < 1$,

$$\omega_\rho$$
$$= \frac{\left[(1-s_{t-1})\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} + s_{t-1}\int_{\tau\sqrt{1-\alpha^2}}^{\infty}\right]\Phi_\rho\left(\tau - \rho u_{t-1} - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\varphi(x)dx}{(1-s_{t-1})\Phi(\tau\sqrt{1-\alpha^2}) + s_{t-1}\left[1 - \Phi(\tau\sqrt{1-\alpha^2})\right]}$$

and, if $\alpha = 1$, for $t = 1$, $\omega_\rho(s_0) = \Phi(\tau)$ with $\mathbb{P}\{s_0 = 0\} = 1$ and $\mathbb{P}\{s_0 = 1\} = 1$ respectively when $\tau > 0$ and $\tau \leq 0$ and, for $t \geq 2$,

$$\omega_\rho$$
$$= \frac{\left[(1-s_{t-1})\int_{-\infty}^{\tau/\sqrt{t-1}} + s_{t-1}\int_{\tau/\sqrt{t-1}}^{\infty}\right]\Phi_\rho\left(\tau - \rho u_{t-1} - x\sqrt{t-1}\right)\varphi(x)dx}{(1-s_{t-1})\Phi(\tau/\sqrt{t-1}) + s_{t-1}\left[1 - \Phi(\tau/\sqrt{t-1})\right]}.$$

Theorem 3.1 fully specifies the joint transition of (s_t) and (y_t) in case of $|\rho| < 1$.

If $|\rho| = 1$, we have perfect endogeneity and Φ_ρ in (15) is not defined. In this case, the current shock to model innovation u_t fully dictates the realization of latent factor w_{t+1} determining the state in the next period. Consequently the transition of the state process (s_t), which is derived above for $|\rho| < 1$ in Theorem 3.1, is no longer applicable. When $|\rho| = 1$, the transition probability ω_ρ to the low state conditional on the previous states and the past values of observed time series behaves differently, which in turn implies that transition density of the state process needs to be modified accordingly. In this case, ω_ρ is given explicitly below for various values of AR coefficient α of the latent factor (w_t).

Corollary 3.2. If $|\rho| = 1$, the transition probability $\omega_\rho = \omega_\rho(s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$ of (s_t) to the low state conditional on the previous states and the past values of observed time series is given as follows: (a) If $\alpha = 0$,

$$\omega_\rho = 1\{\rho u_{t-1} < \tau\}.$$

(b) If $0 < \alpha < 1$,

$$\omega_\rho = (1-s_{t-1}) \min\left(1, \frac{\Phi\left((\tau - \rho u_{t-1})\frac{\sqrt{1-\alpha^2}}{\alpha}\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}\right)$$
$$+ s_{t-1} \max\left(0, \frac{\Phi\left((\tau - \rho u_{t-1})\frac{\sqrt{1-\alpha^2}}{\alpha}\right) - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}\right).$$

(c) If $-1 < \alpha < 0$,

$$\omega_\rho = s_{t-1} \min\left(1, \frac{1 - \Phi\left((\tau - \rho u_{t-1})\frac{\sqrt{1-\alpha^2}}{\alpha}\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}\right)$$
$$+ (1-s_{t-1}) \max\left(0, \frac{\Phi\left(\tau\sqrt{1-\alpha^2}\right) - \Phi\left((\tau - \rho u_{t-1})\frac{\sqrt{1-\alpha^2}}{\alpha}\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}\right).$$

(d) If $\alpha = 1$, for $t = 1$, $\omega_\rho(s_0, y_0) = \Phi(\tau - \rho u_0)$ with $\mathbb{P}\{s_0 = 0\} = 1$ and $\mathbb{P}\{s_0 = 1\} = 1$ respectively when $\tau > 0$ and $\tau \leq 0$ and, for $t \geq 2$,

$$\omega_\rho$$
$$= \begin{cases} 1 - s_{t-1}, & \text{if } \rho u_{t-1} > 0 \\ \frac{\Phi\left((\tau - \rho u_{t-1})\frac{1}{\sqrt{t-1}}\right) - s_{t-1}\Phi\left(\tau/\sqrt{t-1}\right)}{(1-s_{t-1})\Phi\left(\tau/\sqrt{t-1}\right) + s_{t-1}\left[1 - \Phi\left(\tau/\sqrt{t-1}\right)\right]}, & \text{otherwise} \end{cases}.$$

As shown in Theorem 3.1 and Corollary 3.2, the transition density of the state process (s_t) at time t from time $t-1$ depends upon $y_{t-1}, \ldots, y_{t-k-1}$ as well as $s_{t-1}, \ldots, s_{t-k-1}$. The state process (s_t) alone is therefore not Markovian, although the state process augmented with the observed time series (s_t, y_t) becomes a $(k+1)$-st order Markov process. If $\rho = 0$, we have $\omega_\rho = \omega(s_{t-1})$. In this case, the state process (s_t) reduces to a first order Markov process independent of (y_t) as in the conventional Markov switching model, with the transition probabilities obtained in Lemma 2.1.

Our modified Markov switching filter consists of the prediction and updating steps, which are entirely analogous to those in the usual Kalman filter. To develop the modified Markov switching filter, we write

$$p(y_t | \mathcal{F}_{t-1})$$
$$= \sum_{s_t} \cdots \sum_{s_{t-k}} p(y_t | s_t, \ldots, s_{t-k}, \mathcal{F}_{t-1})p(s_t, \ldots, s_{t-k} | \mathcal{F}_{t-1}). \tag{18}$$

Since $p(y_t | s_t, \ldots, s_{t-k}, \mathcal{F}_{t-1}) = p(y_t | s_t, \ldots, s_{t-k}, y_{t-1}, \ldots, y_{t-k})$ is given by (16), it suffices to have $p(s_t, \ldots, s_{t-k} | \mathcal{F}_{t-1})$ to compute the log-likelihood function in (14), which we obtain in the prediction step. For the prediction step, we note that

$$p(s_t, \ldots, s_{t-k} | \mathcal{F}_{t-1})$$
$$= \sum_{s_{t-k-1}} p(s_t | s_{t-1}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})p(s_{t-1}, \ldots, s_{t-k-1} | \mathcal{F}_{t-1}), \tag{19}$$

and that $p(s_t | s_{t-1}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) = p(s_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$, which is given in (17). Consequently, $p(s_t, \ldots, s_{t-k} | \mathcal{F}_{t-1})$ can be readily computed from (19), once we obtain

$p(s_{t-1}, \ldots, s_{t-k-1}|\mathcal{F}_{t-1})$ from the previous updating step. Finally, for the updating step, we have

$$
\begin{aligned}
p(s_t, \ldots, s_{t-k}|\mathcal{F}_t) &= p(s_t, \ldots, s_{t-k}|y_t, \mathcal{F}_{t-1}) \\
&= \frac{p(y_t|s_t, \ldots, s_{t-k}, \mathcal{F}_{t-1})p(s_t, \ldots, s_{t-k}|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})},
\end{aligned}
\tag{20}
$$

where $p(y_t|s_t, \ldots, s_{t-k}, \mathcal{F}_{t-1})$ is given by (16), and we may readily obtain $p(s_t, \ldots, s_{t-k}|\mathcal{F}_t)$ from $p(s_t, \ldots, s_{t-k}|\mathcal{F}_{t-1})$ and $p(y_t|\mathcal{F}_{t-1})$ computed in the prediction step above.

Using our modified Markov switching filter based on the state process (s_t), we can also easily extract the latent autoregressive factor (w_t). This can be done through the prediction and updating steps described above in (19) and (20). In the prediction step, we note that

$$
\begin{aligned}
&p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_{t-1}) \\
&= \sum_{s_{t-k-1}} p(w_t|s_{t-1}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})p(s_{t-1}, \ldots, s_{t-k-1}|\mathcal{F}_{t-1}).
\end{aligned}
\tag{21}
$$

Since $p(s_{t-1}, \ldots, s_{t-k-1}|\mathcal{F}_{t-1})$ is obtained from the previous updating step, we may readily compute $p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_{t-1})$ from (21) once we find $p(w_t|s_{t-1}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})$, the conditional density of latent factor (w_t) on previous states and past information on the observed time series, which is derived below for various values of AR coefficient α of latent factor and endogeneity parameter ρ.

Corollary 3.3. *The transition density of (w_t) conditional on the previous states and the past values of observed time series is given as follows:* (a) *When $|\alpha| < 1$ and $|\rho| < 1$,*

$$
\begin{aligned}
&p(w_t|s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\left(1 - \Phi\left(\sqrt{\frac{1-\rho^2+\alpha^2\rho^2}{1-\rho^2}}\left(\tau - \frac{\alpha(w_t - \rho u_{t-1})}{1-\rho^2+\alpha^2\rho^2}\right)\right)\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)} \\
&\quad \times \mathbb{N}\left(\rho u_{t-1}, \frac{1-\rho^2+\alpha^2\rho^2}{1-\alpha^2}\right),
\end{aligned}
$$

$$
\begin{aligned}
&p(w_t|s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\Phi\left(\sqrt{\frac{1-\rho^2+\alpha^2\rho^2}{1-\rho^2}}\left(\tau - \frac{\alpha(w_t - \rho u_{t-1})}{1-\rho^2+\alpha^2\rho^2}\right)\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)} \\
&\quad \times \mathbb{N}\left(\rho u_{t-1}, \frac{1-\rho^2+\alpha^2\rho^2}{1-\alpha^2}\right).
\end{aligned}
$$

(b) *When $|\alpha| < 1$ and $|\rho| = 1$,*

$$
\begin{aligned}
&p(w_t|s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\frac{\sqrt{1-\alpha^2}}{\alpha}\phi\left(\frac{w_t - \rho u_{t-1}}{\alpha}\sqrt{1-\alpha^2}\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}1\{w_t \geq \alpha\tau + \rho u_{t-1}\},
\end{aligned}
$$

$$
\begin{aligned}
&p(w_t|s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\frac{\sqrt{1-\alpha^2}}{\alpha}\phi\left(\frac{w_t - \rho u_{t-1}}{\alpha}\sqrt{1-\alpha^2}\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}1\{w_t \leq \alpha\tau + \rho u_{t-1}\}.
\end{aligned}
$$

(c) *When $\alpha = 1$ and $|\rho| < 1$,*

$$
p(w_t|s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})
$$

$$
\begin{aligned}
&= \frac{\left(1 - \Phi\left(\sqrt{\frac{t-t\rho^2+\rho^2}{1-\rho^2}}\left(\tau - \frac{w_t - \rho u_{t-1}}{t-t\rho^2+\rho^2}\right)\right)\right)}{1 - \Phi\left(\tau/\sqrt{t-1}\right)} \\
&\quad \times \mathbb{N}\left(\rho u_{t-1}, \frac{t-t\rho^2+\rho^2}{t-1}\right),
\end{aligned}
$$

$$
\begin{aligned}
&p(w_t|s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\Phi\left(\sqrt{\frac{t-t\rho^2+\rho^2}{1-\rho^2}}\left(\tau - \frac{w_t - \rho u_{t-1}}{t-t\rho^2+\rho^2}\right)\right)}{\Phi\left(\tau/\sqrt{t-1}\right)}\mathbb{N}\left(\rho u_{t-1}, \frac{t-t\rho^2+\rho^2}{t-1}\right).
\end{aligned}
$$

(d) *When $\alpha = 1$ and $|\rho| = 1$,*

$$
\begin{aligned}
&p(w_t|s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\frac{1}{\sqrt{t-1}}\phi\left(\frac{w_t - \rho u_{t-1}}{t-1}\right)}{1 - \Phi\left(\tau/\sqrt{t-1}\right)}1\{w_t \geq \tau + \rho u_{t-1}\},
\end{aligned}
$$

$$
\begin{aligned}
&p(w_t|s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \frac{\frac{1}{\sqrt{t-1}}\phi\left(\frac{w_t - \rho u_{t-1}}{t-1}\right)}{1 - \Phi\left(\tau/\sqrt{t-1}\right)}1\{w_t \leq \tau + \rho u_{t-1}\}.
\end{aligned}
$$

We may then obtain

$$
\begin{aligned}
&p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_t) \\
&= \frac{p(y_t|w_t, s_{t-1}, \ldots, s_{t-k}, \mathcal{F}_{t-1})p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})},
\end{aligned}
\tag{22}
$$

in the updating step. By marginalizing $p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_t)$ in (22), we get

$$
p(w_t|\mathcal{F}_t) = \sum_{s_{t-1}} \cdots \sum_{s_{t-k}} p(w_t, s_{t-1}, \ldots, s_{t-k}|\mathcal{F}_t),
$$

which yields the inferred factor,

$$
\mathbb{E}(w_t|\mathcal{F}_t) = \int w_t p(w_t|\mathcal{F}_t)\,dw_t
$$

for all $t = 1, 2, \ldots$. Therefore, we may easily extract the inferred factor, once the maximum likelihood estimates of $p(w_t|\mathcal{F}_t)$, $1 \leq t \leq n$, are available.

We may generalize our model and allow for other covariates to affect the regime switching process. For instance, we may specify the state dependent parameter π_t as

$$
\pi_t = \pi(w_t, x_t),
$$

where (x_t) is a time series of covariates that are predetermined and observable, and accordingly the level function π as

$$
\pi(w, x) = \underline{\pi}1\{w < \tau_1 + \tau_2'x\} + \bar{\pi}1\{w \geq \tau_1 + \tau_2'x\}
$$

with parameters $(\underline{\pi}, \bar{\pi})$ and (τ_1, τ_2), in place of (2). The threshold for regime switching is therefore given as a linear function of some predetermined and observable covariates. This model is more directly comparable to the one considered in Kim et al. (2008). All of our previous theories and results can be easily extended to this model.

We may also easily extend our model to allow for a more general level function $\pi(w)$ than the one introduced in (2). One obvious possibility is to use the level function that allows for multiple regimes, more than two. In fact, it is not uncommon to find applications in which three or more regimes are detected. See, e.g., Garcia and Perron (1996) and Kim et al. (1998) among many others. The extended models with a more general level function allowing for multiple regimes can also be estimated using our modified Markov switching filter similar to that with the

simple two-regime level function that we discussed in detail in the previous sections. We may further extend our model to allow for a continuum of regimes. In this case, however, our modified Markov switching filter is no longer applicable.

4. Simulations

To evaluate the performance of our model and estimation procedure, we conduct an extensive set of simulations. In the sequel, we will present our simulation models and results.

4.1. Simulation models

In our simulations, we consider both mean and volatility switching models. Our volatility model is specified as

$$y_t = \sigma(s_t) u_t, \quad \sigma(s_t) = \underline{\sigma}(1 - s_t) + \overline{\sigma} s_t. \quad (23)$$

The parameters $\underline{\sigma}$ and $\overline{\sigma}$ are set at $\underline{\sigma} = 0.04$ and $\overline{\sigma} = 0.12$, which are roughly the same as our estimates for the regime switching volatilities for the stock returns we analyze in the next section. On the other hand, our simulations for the mean model rely on

$$y_t = \mu(s_t) + \gamma(y_{t-1} - \mu(s_{t-1})) + \sigma u_t,$$
$$\mu(s_t) = \underline{\mu}(1 - s_t) + \overline{\mu} s_t. \quad (24)$$

We set the parameter values at $\sigma = 0.8$, $\gamma = 0.5$, $\underline{\mu} = 0.6$ and $\overline{\mu} = 3$. They are approximately the same as the estimates that we obtain using the US real GDP growth rates analyzed in the next section.

For both mean and volatility models, (s_t) and (u_t) are generated as specified in (1), (3) and (5) for the samples of size 500, and iterated 1000 times. The correlation coefficient ρ between the current model innovation u_t and the next period innovation v_{t+1} of the latent autoregressive factor is set to be negative for both mean and volatility models, as in most of our empirical results reported in the next section. To more thoroughly study the impact of endogeneity on the estimation of our model parameters, we allow ρ to vary from 0 to -1 in increment of 0.1. On the other hand, we consider three pairs of the autoregressive coefficient α of the latent factor and the threshold τ given by $(\alpha, \tau) = (0.4, 0.5), (0.8, 0.7), (1, 9.63)$. The first two pairs with $|\alpha| < 1$ yield a stationary latent factor, while the last pair with $\alpha = 1$ makes the latent factor a random walk.

As discussed earlier, if $\rho = 0$, there exists a one-to-one correspondence between the (α, τ) pair and the pair (a, b) of transition probabilities of state process, where a and b denote respectively the transition probabilities from the low state to the low state and from the high state to the high state. The first pair $(\alpha, \tau) = (0.4, 0.5)$ corresponds to $(a, b) = (0.75, 0.5)$, and the second pair $(\alpha, \tau) = (0.8, 0.7)$ to $(a, b) = (0.86, 0.72)$. The transitions of these two pairs have the same equilibrium distribution given by $(a^*, b^*) = (2/3, 1/3)$, which also becomes the common invariant distribution.[9] This, in particular, implies that the unconditional probabilities of the state being in the low and high regimes are 2/3 and 1/3 respectively in every period. For the third pair with $\alpha = 1$, the state process is nonstationary and its transition varies over time without having an invariant distribution. Our choice of $\tau = 9.63$ in the third pair yields the unconditional probabilities $(2/3, 1/3)$ for the low and high regimes at the terminal period of our simulation, which makes it comparable to the first two pairs.[10]

4.2. Simulation results

In our simulations, we first examine the endogeneity bias. The estimators of parameters in our models are expected to be biased if the presence of endogeneity in regime switching is ignored. To see the magnitude of bias resulting from the neglected endogeneity in regime switching, we let $\rho = 0$ for the exogenous regime switching models. Our simulation results are summarized in Fig. 3. On the left panel of Fig. 3, the bias in the maximum likelihood estimates $\underline{\hat{\sigma}}$ and $\hat{\overline{\sigma}}$ of $\underline{\sigma}$ and $\overline{\sigma}$ in the volatility model are presented in the upper and lower parts of the panel for three different levels of $\alpha = 0.4, 0.8, 1$ measuring the persistence of the latent factor in each of the three columns in the panel. Each graph plots the bias of the estimates from the endogenous (red solid line) and exogenous (blue dashed line) models across different levels of endogeneity ρ on the horizontal axis. Similarly, the right panel of Fig. 3 presents the bias in the maximum likelihood estimates $\underline{\hat{\mu}}, \hat{\overline{\mu}}$ and $\hat{\gamma}$ of $\underline{\mu}, \overline{\mu}$ and γ in the mean model for three persistence levels $\alpha = 0.4, 0.8, 1$ of the latent factor.

The endogeneity in regime switching, if ignored, may yield substantial bias in the estimates of model parameters. This turns out to be true for both mean and volatility models, though the deleterious effect of the neglected endogeneity is relatively larger in the mean model. The magnitude of the bias tends to be larger when α is away from unity and the latent autoregressive factor is less persistent. For example, when $\alpha = 0.4$ and $\rho = -0.7$, the bias of the estimates $\underline{\hat{\mu}}, \hat{\overline{\mu}}$, and $\hat{\gamma}$ in the mean model are respectively 38.7%, -10.8%, and -86.7%. If, however, α is close to unity, the neglected endogeneity does not appear to yield any substantial bias. In fact, when $\alpha = 1$ and the latent factor becomes a random walk, the effect of endogeneity on the parameter estimates in both mean and volatility models becomes insignificant. In all cases, however, the magnitude of the bias becomes larger as $|\rho|$ gets bigger and the degree of endogeneity increases. Though we do not report any details to save space, our simulations show that the inferred probabilities of the latent regimes are also seriously affected if the endogeneity in regime switching is not properly taken care of.

Not only can the presence of endogeneity create a pitfall leading to a substantial bias in parameter estimates, but also an opportunity to improve the precision of parameter estimates in Markov switching models. In fact, in the endogenous regime switching model, additional information on the state process (s_t) is provided by the observed time series (y_t). Note that the transition of the state process (s_t) in our models is determined by lags of (y_t) as well as lags of (s_t), and therefore we have an additional channel for the information in (y_t) to be reflected in the likelihood function. This is not the case if we let $\rho = 0$ as in the conventional Markov switching model that does not allow for the presence of endogeneity. The simulation results in Fig. 4 show that the presence of endogeneity in regime switching indeed improves the efficiency of parameter estimates, if accounted for properly as in our endogenous models. The standard errors of ML estimates of the parameters in our volatility and mean models are presented respectively in the left and right panels of Fig. 4 in exactly the same manner as in Fig. 3.

As shown in Fig. 4, the efficiency gain from incorporating endogeneity explicitly in the analysis of regime switching can be substantial. This is equally true for mean and volatility models. For instance, if we set $\alpha = 0.8$, the standard deviations of the estimators $\hat{\overline{\mu}}$ and $\underline{\hat{\sigma}}$ from our endogenous mean and volatility regime switching models with $\rho = -0.9$ decrease by approximately 24% and 22%, respectively, if compared with the exogenous regime switching models with $\rho = 0$. Of course, the presence of endogeneity yields efficiency gains only when

---

[9] Recall that the invariant distribution of the two-state Markov transition given by a 2×2 transition matrix P is defined by $\pi^* = (a^*, b^*)$ such that $\pi^* = \pi^* P$.

[10] Note that $w_{500} =_d \mathbb{N}(0, 500)$ when $\alpha = 1$ and $\rho = 0$.
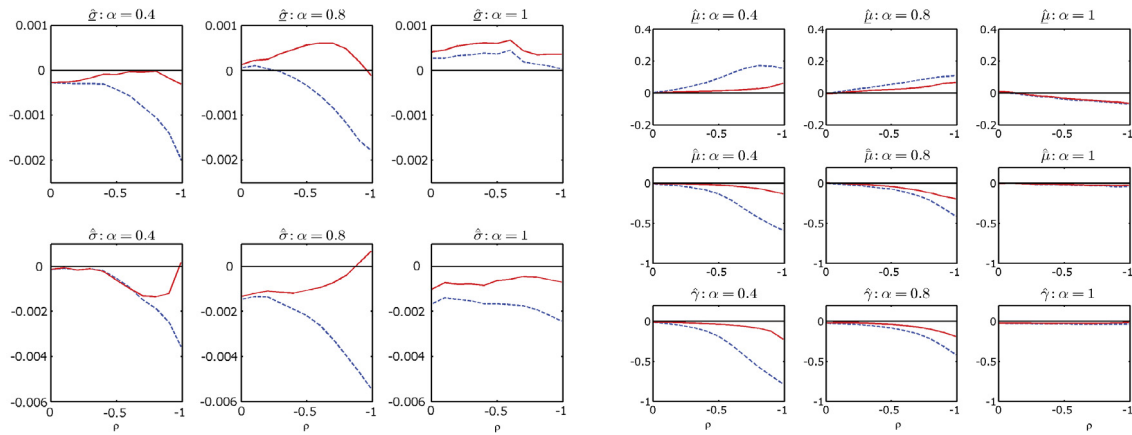
Fig. 3. Endogeneity bias. Notes: On the left panel, the bias in ML estimates $\hat{\underline{\sigma}}$ and $\hat{\overline{\sigma}}$ of low and high volatility levels $\underline{\sigma}$ and $\overline{\sigma}$ from the volatility model are presented respectively in the upper and lower parts, for three persistency levels of latent factor $\alpha = 0.4, 0.8, 1$, in each of its three columns. Each of the six individual graphs plots the bias from the endogenous (red solid line) and exogenous (blue dashed line) regime switching models across different levels of endogeneity parameter ρ on the horizontal axis. Presented in the same manner on the right panel are the bias in the ML estimates $\hat{\underline{\mu}}$, $\hat{\overline{\mu}}$ and $\hat{\gamma}$ of low and high mean levels and AR coefficient of observed time series, $\underline{\mu}$, $\overline{\mu}$ and γ, estimated from the mean model. There are 9 individual graphs covering the bias in three estimates for three persistency levels of the latent factor.
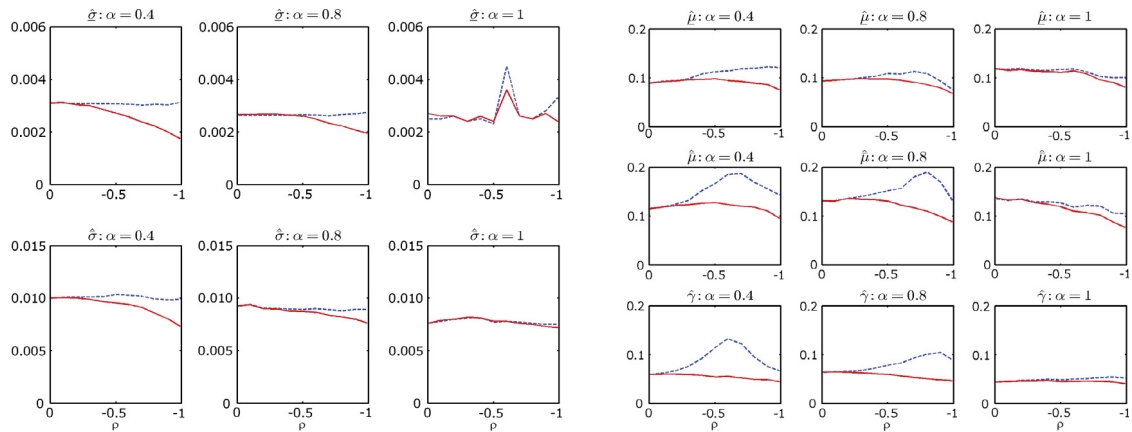
Fig. 4. Efficiency gain from endogeneity. Notes: Respectively presented in the left and right panels of Fig. 4 are the standard errors of the ML estimates of the parameters in our endogenous volatility and mean switching models. The 6 graphs on the left and 9 graphs on the right panels present the standard errors of ML estimates from the volatility and mean models in the exactly the same manner as in Fig. 3.

it is properly taken into account. If the conventional Markov switching model is used, the presence of endogeneity in most cases has a deleterious effect on the standard deviations of parameter estimators.

In general, the standard deviations of parameter estimators are greatly reduced in both mean and volatility models if endogeneity exists in regime switching, as long as $|\alpha| < 1$ and the latent factor is stationary. Naturally, the efficiency gain increases as $|\rho|$ gets large and the degree of endogeneity increases. On the other hand, when the latent factor is nonstationary with $\alpha = 1$, the standard errors of parameter estimators from the endogenous model remain more or less constant across ρ, showing little or no sign of efficiency gain. This may be due to the fact that switching occurs rarely when the latent factor is highly persistent, reducing the opportunity for additional information contained in the observed time series on the switching to play a positive role.[11]

Finally, we consider testing for the presence of endogeneity in regime switching models on the basis of the likelihood ratio test

given by

$$2(\ell(\hat{\theta}) - \ell(\tilde{\theta})), \qquad (25)$$

where ℓ stands for the log-likelihood function and the parameter θ with tilde and hat denote their maximum likelihood estimates with and without the no endogeneity restriction, $\rho = 0$, respectively. The likelihood ratio test has a chi-square limit distribution with one degree of freedom. Presented in the left and right panels of Fig. 5 are the power functions of the likelihood ratio test computed from the simulated volatility and mean switching models for three different levels of persistency in the latent factor measured by its AR coefficient $\alpha = 0.4, 0.8, 1$. For the stationary regime switching model with $\alpha = 0.4$ or 0.8, the test is very powerful with the power increasing rapidly as the value of the endogeneity parameter $|\rho|$ gets large. Under the null hypothesis of no endogeneity, the test has good size properties in the volatility model, but it tends to over-reject in the mean model when the sample size is only moderately large as the latent factor becomes more persistent. Although we do not report the details, the size distortion disappears as the sample size increases. In contrast, the test does not work well when $\alpha = 1$ and the latent factor becomes nonstationary. In this case, the power function increases very slowly as $|\rho|$ gets large, and tends to over-reject in the mean model. The overall performance of the test in the nonstationary case is much worse than in the stationary cases.

---

[11] On the average, the regime change occurs 160, 100, and 15 times out of 500, respectively, for the three pairs of $(\alpha, \tau) = (0.4, 0.5)$, $(0.8, 0.7)$, and $(1, 9.63)$ we consider in our simulations. This clearly shows a rapid decline of regime change frequency as the value of AR coefficient α gets closer to 1 and the latent factor becomes a random walk.
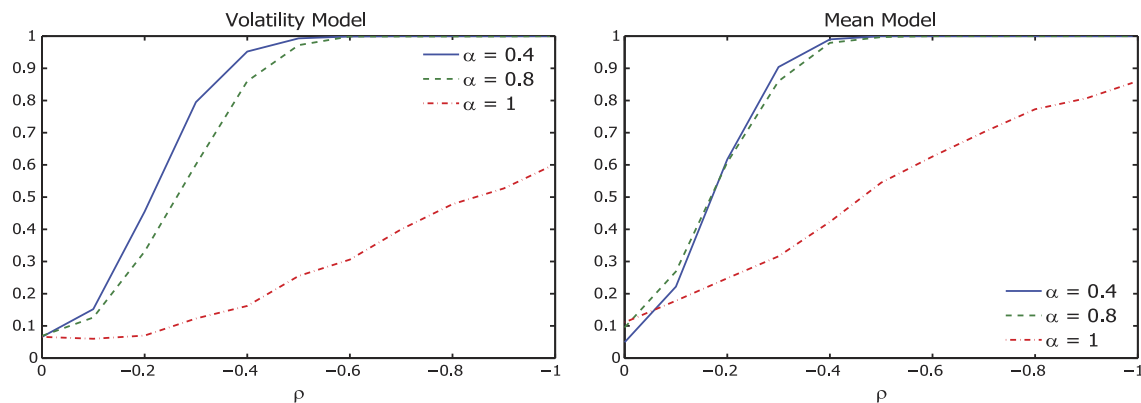
Fig. 5. Power function of LR test for endogeneity. Notes: The left and right hand side graphs of Fig. 5 present the power functions of the likelihood ratio test computed respectively from the volatility and mean switching models for three different levels of persistency in the latent factor (w_t) measured by its AR coefficient $\alpha = 0.4, 0.8, 1$.

5. Empirical illustrations

To empirically illustrate our approach, we analyze the US GDP growth rates and US excess stock market returns using the mean and volatility models with regime switching, respectively.

5.1. Regime switching in stock return volatility

As an illustration, our volatility model in (23) is estimated using the demeaned US excess stock market returns at monthly frequency. We define stock market returns as the monthly observations of value-weighted stock returns including dividends on the NYSE/AMEX index, and impute monthly risk free rates from the daily observations of three months T-bill rates.[12] Both NYSE/AMEX index returns and T-bill rates are obtained from the Center for Research in Security Prices (CRSP) for the period January 1926–December 2012. The monthly excess stock market returns are obtained by subtracting the monthly risk free rates from the stock market returns. We analyze two sample periods: the full sample period (1926–2012) and a recent subsample period (1990–2012). We choose this subsample to relate the extracted latent volatility factor obtained from our endogenous switching model with VIX, one of the most commonly used volatility indices, which is available only for this subsample period.

To estimate the volatility switching model by the ML method, we use our modified Markov switching filter implemented with the numerical optimization method including the commonly used BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm. Our estimates are reported in Table 1.[13] The estimates for the endogeneity parameter ρ are quite substantial in both samples, -0.970 for the full sample and -0.999 for the subsample,[14]

Table 1
Estimation results for volatility model.

| Sample periods	1926–2012		1990–2012	
Endogeneity	Ignored	Allowed	Ignored	Allowed
$\underline{\sigma}$	0.039	0.038	0.022	0.025
	(0.001)	(0.001)	(0.002)	(0.004)
$\overline{\sigma}$	0.115	0.115	0.050	0.055
	(0.009)	(0.009)	(0.003)	(0.008)
ρ		−0.970		−0.999
		(0.086)		(0.010)
log-likelihood	1742.536	1748.180	507.700	511.273
p-value (LR test for $\rho = 0$)		0.001		0.008

providing ample evidence of the presence of endogeneity in regime switching in market volatility. The presence of endogeneity in regime switching is formally tested using the usual likelihood ratio test given in (25). In both sample periods, we reject the null of no endogeneity at a 1% significance level, as reported in the bottom line of Table 1. For the full sample period 1926–2012, the estimates of α and τ in the exogenous model are 0.994 (0.004) and 11.375 (4.472), from which we obtain 0.991 and 0.928 as the estimates of the low-to-low and high-to-high transition probabilities. In the endogenous model, we obtain 0.986 (0.009) and 7.385 (2.567) for α and τ. On the other hand, for the recent subsample period 1990–2012, we find 0.997 (0.003) and −3.212 (7.055) as the estimates of α and τ in the exogenous model, which yield 0.973 and 0.981 for the low-to-low and high-to-high transition probabilities. For the endogenous model, we have 0.979 (0.044) and 0.685 (2.114) for α and τ.

What is most clearly seen from Fig. 6 is the striking difference in the time series plots of the transition probabilities estimated from the exogenous and endogenous volatility regime switching models. The transition probability estimated by the exogenous model is constant over the entire sample period, while the corresponding transition probabilities estimated by the endogenous model vary over time, and depend upon the lagged value of the excess market return y_{t-1} as well as the realized value of the previous state s_{t-1}. This point is clearly demonstrated in the left hand side graph of Fig. 6 which presents the transition probability from the low volatility regime at $t-1$ to the high volatility regime at t estimated by the exogenous and endogenous switching models. This low to high transition probability is estimated to be 2.7% throughout the entire sample period by the exogenous model, while in contrast the corresponding transition probabilities estimated by the endogenous model vary over time with the realized values of the lagged excess market return. It shows in particular that the transition probabilities have been changing drastically, and reach a value as high as 87.1%. The right hand side graph of Fig. 6 similarly illustrates the same point with the transition probability from high

---

[12] Monthly risk free rates are obtained by continuously compounding daily risk free rates, which we impute using the monthly series of annualized yield to maturity (TMYTM) constructed from nominal price of three month T-bill. The annualized yield to maturity is converted to the daily yield to maturity (TMYLD) by the conversion formula provided by CRSP, i.e., $TMYLD_t = (1/365)(1/100)TMYTM_t$. The monthly yield is then obtained by continuously compounding this daily yield as $\exp(TMYLD_{t-1} \times N_t) - 1$, where N_t is the number of days between the quote dates for the current and the previous month, $MCALDT_{t-1}$. The number of days between quote dates ranges from 28 to 33.

[13] The standard errors are presented in parenthesis.

[14] Here our estimate of ρ for the subsample is virtually identical to −1, in which case the current shock to the stock returns would be directly transmitted to the latent factor. This, however, does not mean that the regime becomes perfectly predictable, since (w_t) is never observed even if (s_t) is sometimes known. A referee suspects that it may also be spuriously obtained by fitting a model with regime switching in volatility for a process generated without any changing regimes. We found by simulation that this possibility is less than 30% at most.
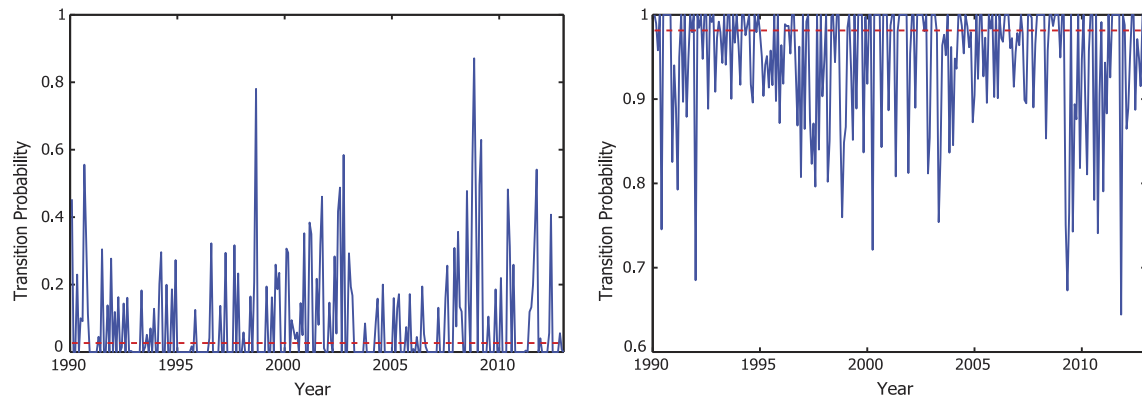
Fig. 6. Estimated transition probability from volatility model. Notes: Fig. 6 presents the transition probabilities from the volatility model. The left hand side graph shows the transition probability from low to high volatility state: the blue solid line refers to $\mathbb{P}(s_t = 1 \mid s_{t-1} = 0, y_{t-1})$ in our endogenous regime switching model, while the red dashed line corresponds to $\mathbb{P}(s_t = 1 \mid s_{t-1} = 0)$ in the exogenous regime switching model. Similarly, the right hand side graph shows the transition probabilities of staying at high volatility state.
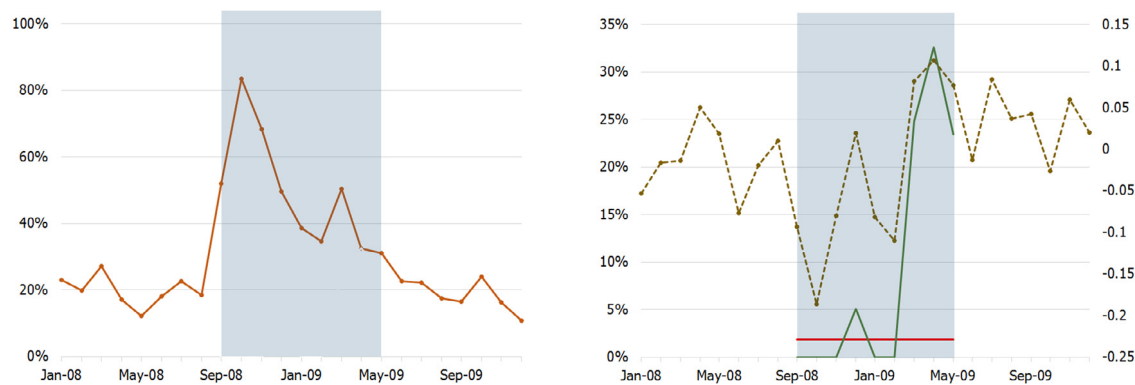
Fig. 7. Transition probabilities for recent financial crisis period. Notes: The high to low transition probabilities during the high volatility regime from September 2008 to May 2009 are presented on the right hand side graph of Fig. 7, where the green line signifies the time varying transition probability $\mathbb{P}(s_t = 0 \mid s_{t-1} = 1, y_{t-1})$ estimated from our endogenous regime switching model, while the red line corresponds to the constant transition probability $\mathbb{P}(s_t = 0 \mid s_{t-1} = 1)$ obtained from the exogenous switching model. The solid line and the dashed line on the left and right hand side graphs respectively present the time series of annualized monthly volatility and the monthly NYSE/AMEX index returns. The shaded areas on both graphs of Fig. 7 indicate the high volatility regime.

volatility state at $t-1$ to high volatility state at t by the exogenous and endogenous switching models.

The endogeneity in regime switching plays a more important role when the underlying regime is known. Fig. 7 illustrates that the time varying transition probabilities from the endogenous model can indeed produce a much more realistic assessment for the likelihood of moving into a low volatility regime from a known high volatility regime. The left hand side graph of Fig. 7 presents the time series of annualized monthly volatility. The time series of monthly stock returns is also presented in the graph on the right. The volatility increased dramatically in September 2008 when Lehman Brothers filed bankruptcy and stayed high until May 2009. We consider this period as a high volatility regime.[15] The shaded areas on both graphs of Fig. 7 indicate this high volatility regime.

The high to low transition probabilities during the high volatility regime are presented on the right hand side graph of Fig. 7, where the green line signifies the time varying transition probability $\mathbb{P}(s_t = 0 \mid s_{t-1} = 1, y_{t-1})$ estimated from our endogenous regime switching model, while the red line corresponds to the constant transition probability $\mathbb{P}(s_t = 0 \mid s_{t-1} = 1)$ obtained from

the exogenous switching model. Indeed the transition probability estimated by the exogenous model stays constant for the entire duration of the high volatility regime, which is in sharp contrast to the substantially time varying transition probabilities obtained from the endogenous model. Notice that the high to low transition probability from our endogenous model is smaller than that from the exogenous model at the beginning of the high volatility regime; however, it goes up drastically toward the end of the high volatility regime, which coincides, not surprisingly, with the rapid recovery of the stock market in the early spring of 2009.

To see how well our endogenous volatility switching model can explain the current state of market volatility, we compare the sample paths of the extracted latent factor with that of VIX, a popular measure for implied market volatility, over the subsample period 1990 to 2012, where VIX is available. See Fig. 10 in the Appendix, which presents the sample path of the extracted latent factor along with that of the CBOE (Chicago Board Options Exchange) volatility index VIX for the period 1990–2012. The VIX stayed relatively high during 1998–2004 and 2008 periods indicating that the volatility was high during those periods. As shown in Fig. 10, the extracted latent factor obtained from our endogenous volatility model also stays relatively high, moving closely with VIX during those high volatility periods. VIX has been used as a gauge for "fear factor" or an indicator for the overall risk level of market. Therefore the extracted latent factor from our volatility model may be considered as an alternative measure which can play a similar role played by VIX.

---

[15] The volatility level in each month during this high volatility regime from September 2008 to May 2009 is at least twice higher than the average volatility computed over the 32-month period ending at the start of this high volatility regime in September 2008.

Table 2
Maximum likelihood estimates for Hamilton (1989) model.

Sample periods	1952–1984		1984–2012	
Endogeneity	Ignored	Allowed	Ignored	Allowed
$\underline{\mu}$	−0.165	−0.083	−0.854	−0.758
	(0.219)	(0.161)	(0.298)	(0.311)
$\overline{\mu}$	1.144	1.212	0.710	0.705
	(0.113)	(0.095)	(0.092)	(0.085)
γ_1	0.068	0.147	0.154	0.169
	(0.123)	(0.104)	(0.105)	(0.105)
γ_2	−0.015	0.044	0.350	0.340
	(0.112)	(0.096)	(0.105)	(0.103)
γ_3	−0.175	−0.260	−0.036	−0.076
	(0.108)	(0.090)	(0.106)	(0.128)
γ_4	−0.097	−0.067	0.043	0.049
	(0.104)	(0.095)	(0.103)	(0.112)
σ	0.794	0.784	0.455	0.452
	(0.065)	(0.057)	(0.034)	(0.032)
ρ		−0.923		0.999
		(0.151)		(0.012)
Log-likelihood	−173.420	−169.824	−80.584	−76.447
p-value		0.007		0.004

We also compute for each period the inferred probability of being in the high volatility regime using our endogenous volatility switching model as well as the conventional exogenous switching model. They are presented in Fig. 11 in the Appendix. Overall the time series of the high volatility probabilities computed from both endogenous and exogenous switching models are comparable. However, those obtained from our endogenous model tend to fluctuate more over time compared to those from its exogenous counterpart. This is perhaps because the endogenous model more efficiently extracts the information on regime switching as shown in our simulation. It is also interesting to note that the probability of being in the high volatility regime computed from our endogenous model is exactly at 1 during the 2008 financial crisis, while that from the exogenous model does not quite reach 1, albeit close to 1, which demonstrates the potential of our model to more precisely estimate the probability of the financial market becoming unstable.

5.2. Regime switching in GDP growth rates

In this section, we investigate the regime switching behavior of the US real GDP growth rates constructed from the seasonally adjusted quarterly real GDP series for the period 1952:Q1-2012:Q4.[16] As in Hamilton (1989), we model the real GDP growth rates (y_t) as an AR(4) process with regime switching. Since there seems to be a structural break in the postwar US real GDP growth rates in 1984:Q1, as noted in Kim and Nelson (1999), we consider two sample periods: the earlier sample period covering 1952:Q1-1984:Q4, and the more recent sample period covering 1984:Q1-2012:Q4. We use the same data used in Kim and Nelson (1999) and compare our results with theirs.[17]

We estimate the mean switching model for the GDP growth rates using our new modified Markov switching filter along with BFGS method. Table 2 presents the estimation results for the two sample periods we consider.[18] The ML estimates obtained from the exogenous model with the constraint $\rho = 0$ imposed and those from the endogenous model with no constraint on ρ are largely comparable. However, the difference in the estimates from two sample periods is remarkable. In particular, the estimates

of $\underline{\mu}, \overline{\mu}$, and σ from two sample periods are quite different, which may be used as a supporting evidence for the presence of a structural break in the US GDP series. It is also interesting to note that the ML estimate of the correlation coefficient ρ measuring the degree of endogeneity for the earlier sample period is large and negative, −0.923, which is in contrast with the value, 0.999, estimated from the recent sample period.[19] For the earlier sample period 1952–1984, the estimates of α and τ in the exogenous model are 0.895 (0.077) and −1.009 (0.773), from which we obtain the estimates 0.796 and 0.901 for the low-to-low and high-to-high transition probabilities. In the endogenous model, we have 0.927 (0.041) and −0.758 (0.883) for α and τ. On the other hand, for the recent sample period 1984–2012, the estimates for α and ρ in the exogenous model are 0.842 (0.162) and −3.282 (1.489), which yield 0.526 and 0.981 for the low-to-low and high-to-high transition probabilities. In the endogenous model, we have 0.809 (0.145) and −2.782 (1.144) for α and τ. Moreover, the maximum value of the log-likelihood function from the unrestricted endogenous model is larger than that from the restricted exogenous model with $\rho = 0$ imposed, and consequently the null of no endogeneity is decisively rejected by the usual likelihood ratio test given in (25) at 1% significance level for both sample periods.

The estimated transition probabilities are presented in Fig. 8. Since (s_t, y_t) is jointly a fifth-order Markov process, the transition probabilities at time t depend on $s_{t-1}, \ldots, s_{t-5}$ as well as $y_{t-1}, \ldots, y_{t-5}$. The left hand side graph of Fig. 8 shows the transition probabilities from the low mean state at $t-1$ to the low mean state at t. There are 17 lines in the graph. The 16 solid lines represent the sample paths of the 16 transition probabilities obtained from our endogenous regime switching model for each of the 16 possible realizations of the four lagged state variables, $s_{t-2}, s_{t-3}, s_{t-4}$, and s_{t-5}. Note that each of the four lagged state variables $s_{t-2}, s_{t-3}, s_{t-4}$, and s_{t-5} takes a value either 0 or 1, giving 16 possibilities for their joint realizations. We therefore calculate the transition probability from the low state at $t-1$ to the low state at t, i.e., $\mathbb{P}(s_t = 0 | s_{t-1} = 0, s_{t-2} = i, s_{t-3} = j, s_{t-4} = k, s_{t-5} = \ell, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ for all 16 possible combinations of $i, j, k, \ell = 0, 1$. The one dashed line represents the corresponding probability of staying at low mean state obtained from the exogenous model. Similarly, the graph on the right hand side shows the sample paths of the 17 transition probabilities from the high to low mean state, 16 solid time varying lines from the endogenous model and one dashed straight line from the exogenous model. The most salient feature from the two figures presented in Fig. 8 is that the estimated transition probabilities estimated by our endogenous regime switching model are drastically different from the one obtained from its exogenous counterpart.

Fig. 9 presents the transition probabilities from the low mean regime to the low mean regime plotted along with the US GDP growth rates. NBER officially announced on December 1, 2008 that a recession began in December 2007 after the December 2007 peak which defined the turning point from expansion to recession, and it announced on October 21, 2010 that the recession ended in June 2009. Therefore, we knew that we were in recession on December 1, 2008. Using this information, we calculate the transition probability from the low to the low mean regime. The

---

[16] Source: Bureau of Economic Analysis, US Department of Commerce. The growth rate of real GDP is calculated as the first difference of log real GDP.

[17] We use the data set provided at the website for the monograph by Kim and Nelson (1999).

[18] Again, the standard errors are presented in parenthesis.

[19] Like one of the estimates for ρ in the volatility model considered earlier, here we also have an extreme case. The estimated ρ for the recent sample period is very close to 1, which suggests that the GDP growth rates evolve with the mean regimes determined almost entirely by the shocks to themselves. As discussed, a referee suspects that it may be spuriously obtained by fitting a model with regime switching in mean for a process generated without any changing regimes. However, we found by simulation that this possibility is low and only around 10%.
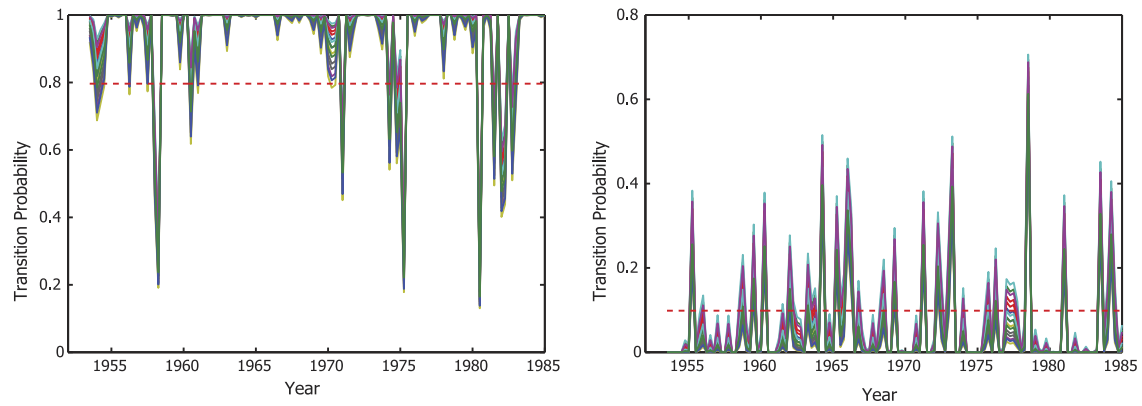
Fig. 8. Estimated transition probabilities for mean model. Notes: Fig. 8 presents the transition probabilities from the mean model for the US GDP growth rates. The left hand side graph shows the sample paths of the 17 transition probabilities of staying at the low mean state: the 16 solid time varying lines represent transition probabilities obtained from our endogenous switching model by computing $\mathbb{P}(s_t = 0 \mid s_{t-1} = 0, s_{t-2} = i, s_{t-3} = j, s_{t-4} = k, s_{t-5} = \ell, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ for all 16 possible combinations of $i, j, k, \ell = 0, 1$, while the one red dashed straight line represents the probability of staying at the low mean state $\mathbb{P}(s_t = 0 \mid s_{t-1} = 0)$ obtained from the exogenous model. In the same way, the right hand side graph shows the transition probabilities from the high to the low mean state.
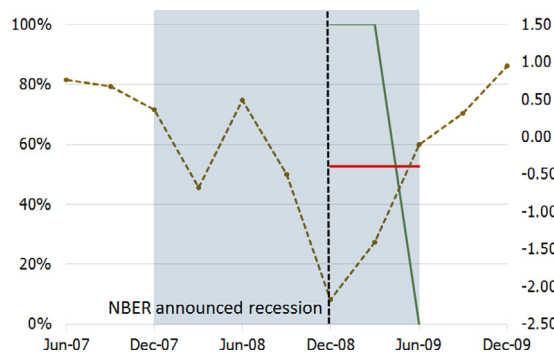
Fig. 9. Transition probabilities during recent recession period. Notes: Fig. 9 presents the transition probabilities from the mean switching model for the most recent US recession period, 2007–2009. The shaded area indicates the recession period which started on 2007:Q4 and ended on 2009:Q2, and the dashed vertical line marks December 1, 2008 when NBER announced the recession began on December 2007. The solid green (red) line signifies the low to low transition probability estimated by the endogenous (exogenous) switching model. The dashed blue line plotted on the right vertical axis represents the US real GDP growth rates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

red solid line is the corresponding transition probability from exogenous switching model. It is constant over time. However, the green solid line signifying the transition probability from our endogenous model drastically changes over time. Note that the transition probability in our endogenous model is determined not only by previous states but also by the lagged values of the GDP growth rates. The endogenous switching model exploits the information from the past values of the observed time series to update the transition probability. When the observed GDP growth rate is low, the transition probability from the low to the low regime is as high as 100%, but this transition probability sharply declines to virtually zero when we update our information with the high realized values of GDP growth rates.

We also extract the latent factor determining the states from our endogenous mean switching model, and compare it with the recession periods identified by NBER. See Fig. 13 in the Appendix, which presents the sample path of the extracted latent mean switching factor and NBER recession periods during the two sample periods we consider, 1952–1984 and 1984–2012. In both sample periods, we can see clearly that the trough dates of the extracted latent factor coincide with NBER recession periods indicated by shaded areas in the graphs. It seems that we may use the extracted latent factor from our endogenous mean switching model as a potential indicator for business cycle.

Finally, we compute the inferred probability that we were in the recession regime, which are presented in Figs. 14 and 15 for both sample periods. Both endogenous and exogenous models produce high recession probabilities when the *levels* of growth rates become negative. However, our endogenous model has an additional channel via transition probabilities to reflect the *changes* in the observed growth rates, which its exogenous counterpart does not. The ability of our endogenous model to exploit the information on the changes in the growth rates can indeed result in strikingly different recession probabilities. Shortly before the 1981 recession formally announced by NBER, the GDP growth rates sharply went up and peaked in the second quarter of 1978, and then quickly declined and became negative in the second quarter of 1980. During the period prior to the 1981 recession, we see much higher values of the recession probabilities computed from our endogenous model compared to those obtained from the exogenous model. Our model could reflect the downward movements in the growth rates and accordingly increased the recession probabilities, even before the growth rates become negative.

6. Conclusions

In the paper, we propose a new approach to model regime switching based on an autoregressive latent factor. Our approach has several clear advantages over the conventional regime switching model. Most importantly, we may allow for endogeneity in regime switching, so that a shock to the observed time series affects the change in regime. In the mean model with regime switching, the presence of endogeneity implies that the mean reversion may occur at two different levels: the reversion of the observed time series to its state dependent mean, and the reversion of the state dependent mean to offset the effect of a shock. In the volatility model, on the other hand, the endogeneity in regime switching implies the presence of leverage effect. Furthermore, our regime switching model becomes observationally equivalent to the conventional Markov switching model, if there is no endogeneity in regime switching. Finally, our approach allows the transition of the state process to be nonstationary and strongly persistent.

The empirical evidence for the presence of endogeneity in regime switching appears to be strong and unambiguous. Our simulations make it clear that neglecting endogeneity in regime switching incurs not only a substantial bias, but also a significant information loss, in estimating model parameters. If endogeneity in the regime switching is ignored, the variability of parameter

estimates sharply increases and consequently the inferred probabilities of the latent states become less precise. This is because the endogeneity in regime switching creates an important additional link between the latent states and observed time series, and therefore, the information that can be reflected through this link cannot be exploited if the endogeneity is ignored. The additional information that we may extract from this new link is certainly more valuable in a Markov switching model, since the state process playing such a critical role in the model is latent and must be inferred from a single observable time series.

Appendix A. Mathematical proofs

Proof of Lemma 2.1. From (9), we may deduce that

$$\mathbb{P}\left\{s_t = 0 \Big| w_{t-1}\sqrt{1-\alpha^2} = x\right\} = \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right),$$

from which it follows that

$$\mathbb{P}\{s_t = 0 | w_{t-1} < \tau\} = \mathbb{P}\left\{s_t = 0 \Big| w_{t-1}\sqrt{1-\alpha^2} < \tau\sqrt{1-\alpha^2}\right\}$$

$$= \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \mathbb{P}\left\{s_t = 0 \Big| w_{t-1}\sqrt{1-\alpha^2} = x\right\}\varphi(x)dx}{\mathbb{P}\left\{w_{t-1}\sqrt{1-\alpha^2} < \tau\sqrt{1-\alpha^2}\right\}}$$

$$= \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\varphi(x)dx}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

upon noticing that $w_{t-1}\sqrt{1-\alpha^2} =_d \mathbb{N}(0,1)$. The stated result for $a(\alpha, \tau)$ can therefore be easily deduced from (11). Similarly, we have

$$\mathbb{P}\{s_t = 1 | w_{t-1} \geq \tau\} = \mathbb{P}\left\{s_t = 1 \Big| w_{t-1}\sqrt{1-\alpha^2} \geq \tau\sqrt{1-\alpha^2}\right\}$$

$$= \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \mathbb{P}\left\{s_t = 1 \Big| w_{t-1}\sqrt{1-\alpha^2} = x\right\}\varphi(x)dx}{\mathbb{P}\left\{w_{t-1}\sqrt{1-\alpha^2} \geq \tau\sqrt{1-\alpha^2}\right\}}$$

$$= \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \left[1 - \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\right]\varphi(x)dx}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

since

$$\mathbb{P}\left\{s_t = 1 \Big| w_{t-1}\sqrt{1-\alpha^2} = x\right\} = 1 - \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right),$$

due to (10), from which and (11) the stated result for $b(\alpha, \tau)$ follows readily as above. □

Proof of Corollary 2.2. The stated result for $t = 1$ is obvious, since $\mathbb{P}\{s_0 = 0\} = 1$ and $\mathbb{P}\{s_0 = 1\} = 1$ depending upon $\tau > 0$ and $\tau \leq 0$. Note that we set $w_0 = 0$ for identification when $\alpha = 1$. For $t \geq 2$, upon noticing that $w_{t-1}/\sqrt{t-1} =_d \mathbb{N}(0,1)$, the proof is entirely analogous to that of Lemma 2.1, and the details are omitted. □

Proof of Theorem 3.1. We only provide the proof for the case of $|\alpha| < 1$. The proof for the case of $\alpha = 1$ is virtually identical, except that we have $w_{t-1}/\sqrt{t-1} =_d \mathbb{N}(0,1)$ for $t \geq 2$ in this case, in place of $w_{t-1}\sqrt{1-\alpha^2} =_d \mathbb{N}(0,1)$ for the case of $|\alpha| < 1$. If we let

$$z_t = \frac{w_t - \alpha w_{t-1}}{\sqrt{1-\rho^2}} - \frac{\rho u_{t-1}}{\sqrt{1-\rho^2}},$$

we may easily deduce that

$$p(z_t | w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}) = \mathbb{N}(0,1).$$

It follows that

$$\mathbb{P}\left\{w_t < \tau \Big| w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \mathbb{P}\left\{z_t < \frac{\tau - \alpha w_{t-1}}{\sqrt{1-\rho^2}} - \frac{\rho u_{t-1}}{\sqrt{1-\rho^2}}\right| \\ \times w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \Phi_\rho\left(\tau - \rho u_{t-1} - \alpha w_{t-1}\right).$$

Note that

$$p(w_t | w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}) = p(w_t | w_{t-1}, u_{t-1}),$$

and that w_{t-1} is independent of u_{t-1}. Consequently, we have

$$\mathbb{P}\left\{w_t < \tau \Big| w_{t-1} < \tau, w_{t-2}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \mathbb{P}\left\{w_t < \tau \Big| w_{t-1}\sqrt{1-\alpha^2} \right. \\ \left. < \tau\sqrt{1-\alpha^2}, w_{t-2}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \mathbb{P}\left\{s_t = 0 \Big| s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi_\rho\left(\tau - \rho u_{t-1} - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\varphi(x)dx}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}$$

and

$$\mathbb{P}\left\{w_t < \tau \Big| w_{t-1} \geq \tau, w_{t-2}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \mathbb{P}\left\{w_t < \tau \Big| w_{t-1}\sqrt{1-\alpha^2} \right. \\ \left. \geq \tau\sqrt{1-\alpha^2}, w_{t-2}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \mathbb{P}\left\{s_t = 0 \Big| s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right\}$$

$$= \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \Phi_\rho\left(\tau - \rho u_{t-1} - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\varphi(x)dx}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

since in particular $w_{t-1}\sqrt{1-\alpha^2} =_d \mathbb{N}(0,1)$, from which the stated result for the transition density for (s_t, y_t) may be readily obtained.

Now we write

$$p(s_t, y_t | s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1)$$
$$= \quad p(y_t | s_t, s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1) \\ \times p(s_t | s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1).$$

It follows from (16) that

$$p(y_t | s_t, s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1)$$
$$= p(y_t | s_t, \ldots s_{t-k}, y_{t-1}, \ldots, y_{t-k}).$$

Moreover, we have

$$p(s_t | s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1)$$
$$= p(s_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}),$$

as we have shown above. Therefore, it follows that

$$p(s_t, y_t | s_{t-1}, \ldots, s_1, y_{t-1}, \ldots, y_1)$$

$$= p(y_t | s_t, \ldots, s_{t-k}, y_{t-1}, \ldots, y_{t-k})$$
$$\times \; p(s_t | s_{t-1}, \ldots, s_{t-(k+1)}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= p(s_t, y_t | s_{t-1}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$

and (s_t, y_t) is a $(k + 1)$-st order Markov process. □

Proof of Corollary 3.2. We only provide the proof for the case of $0 < \alpha < 1$. The proof for the case of $\alpha = 0$ is trivial and the proof for the case of $-1 < \alpha < 0$ can be easily done with a simple modification of the case of $0 < \alpha < 1$. The proof for the case of $\alpha = 1$ is virtually identical, except that we have $w_{t-1}/\sqrt{t-1} =_d \mathbb{N}(0, 1)$ for $t \geq 2$ in this case, in place of $w_{t-1}\sqrt{1-\alpha^2} =_d \mathbb{N}(0, 1)$ for the case of $|\alpha| < 1$.

It follows that

$$\mathbb{P}\left\{ w_t < \tau \,\middle|\, w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1} \right\}$$
$$= \mathbb{P}\left\{ \alpha w_{t-1} + v_t < \tau \,\middle|\, w_{t-1}, \ldots, w_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1} \right\}$$
$$= \mathbb{P}\left\{ \alpha w_{t-1} + \rho u_{t-1} < \tau \,\middle|\, w_{t-1}, u_{t-1} \right\}$$
$$= 1\{\alpha w_{t-1} + \rho u_{t-1} < \tau\}.$$

We note that when $0 < \alpha < 1$,

$$\omega_\rho(s_{t-1} = 0, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= \mathbb{P}\left\{ \alpha w_{t-1} + \rho u_{t-1} < \tau \,\middle|\, w_{t-1} < \tau, u_{t-1} \right\}$$
$$= \mathbb{P}\left\{ \sqrt{1-\alpha^2}\, w_{t-1} < \frac{\sqrt{1-\alpha^2}}{\alpha}(\tau - \rho u_{t-1}) \,\middle|\, \right.$$
$$\left. \times \; \sqrt{1-\alpha^2}\, w_{t-1} < \tau\sqrt{1-\alpha^2}, u_{t-1} \right\}$$
$$= \begin{cases} 1, & \text{if } \dfrac{1}{\alpha}(\tau - \rho u_{t-1}) < \tau, \\[2ex] \dfrac{\Phi\left((\tau - \rho u_{t-1})\dfrac{\sqrt{1-\alpha^2}}{\alpha}\right)}{\Phi(\tau\sqrt{1-\alpha^2})}, & \text{otherwise.} \end{cases}$$

Similarly, we have

$$\omega_\rho(s_{t-1} = 1, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= \mathbb{P}\left\{ \alpha w_{t-1} + \rho u_{t-1} < \tau \,\middle|\, w_{t-1} > \tau, u_{t-1} \right\}$$
$$= \mathbb{P}\left\{ \sqrt{1-\alpha^2}\, w_{t-1} < \frac{\sqrt{1-\alpha^2}}{\alpha}(\tau - \rho u_{t-1}) \,\middle|\, \right.$$
$$\left. \times \; \sqrt{1-\alpha^2}\, w_{t-1} > \tau\sqrt{1-\alpha^2}, u_{t-1} \right\}$$
$$= \frac{\Phi\left((\tau - \rho u_{t-1})\dfrac{\sqrt{1-\alpha^2}}{\alpha}\right) - \Phi(\tau\sqrt{1-\alpha^2})}{1 - \Phi(\tau\sqrt{1-\alpha^2})}$$
$$\times \; 1\left\{ \frac{\tau - \rho u_{t-1}}{\alpha} \geq \tau \right\}.$$

Proof of Corollary 3.3. We note that

$$p(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})$$
$$= p(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1})$$
$$= p(w_t | w_{t-1} > \tau, u_{t-1})$$
$$= \frac{\int_\tau^\infty p(w_t, w_{t-1}, u_{t-1})\, dw_{t-1}}{\int_\tau^\infty p(w_{t-1}, u_{t-1})\, dw_{t-1}}$$
$$= \frac{\int_\tau^\infty p(w_t | w_{t-1}, u_{t-1})\, p(w_{t-1})\, dw_{t-1}}{\int_\tau^\infty p(w_{t-1})\, dw_{t-1}},$$
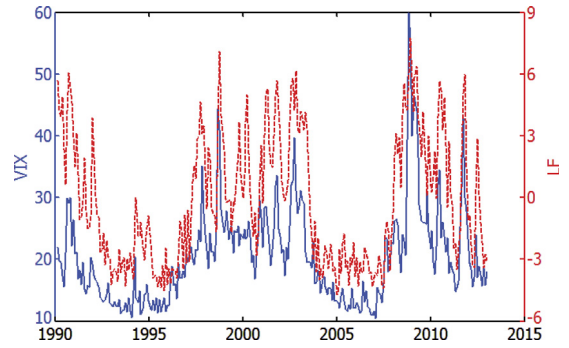
Fig. 10. Extracted Latent Factor and VIX. Notes: Fig. 10 presents the sample path of the latent factor extracted from the endogenous volatility switching model (dashed red line) along with that of the CBOE (The Chicago Board Options Exchange) volatility index VIX (solid blue line) for the period 1990–2012, respectively, on the left and right vertical axis.
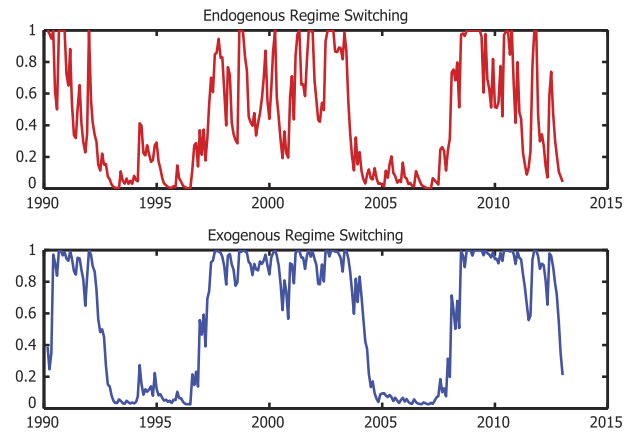
Fig. 11. High volatility probabilities. Notes: Fig. 11 presents the time series of the probabilities of being in the high volatility regime. Top panel plots the high volatility probability series obtained from the endogenous volatility switching model with red line, while the bottom panel plots those from its conventional exogenous counterpart with blue line.
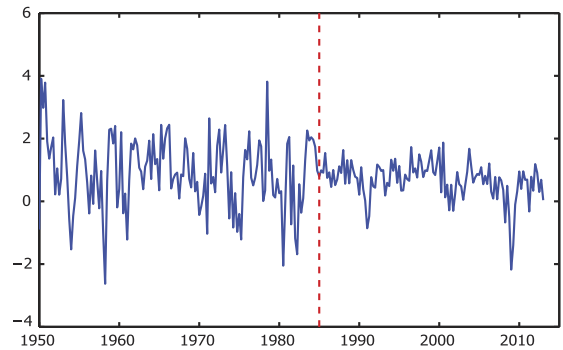

Fig. 12. US Real GDP Growth Rates. Notes: Fig. 12 presents the US real GDP growth rates which is calculated as 100 times the change in the log of real GDP. It is seasonally adjusted, annualized, and collected at the quarterly frequency from 1952 to 2012. The vertical dashed red line indicates 1983:Q4.

where the last equality follows from the independence between (w_t) and (u_t). We may similarly deduce that

$$p(w_t | s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1})$$
$$= \frac{\int_{-\infty}^\tau p(w_t | w_{t-1}, u_{t-1})\, p(w_{t-1})\, dw_{t-1}}{\int_{-\infty}^\tau p(w_{t-1})\, dw_{t-1}}.$$

In case $|\alpha| < 1$ and $|\rho| < 1$, we have $w_t | w_{t-1}, u_{t-1} =_d \mathbb{N}\left(\alpha w_{t-1} + \rho u_{t-1}, 1 - \rho^2\right)$. Since $w_t =_d w_{t-1} =_d \mathbb{N}\left(0, 1/(1 - \alpha^2)\right)$,
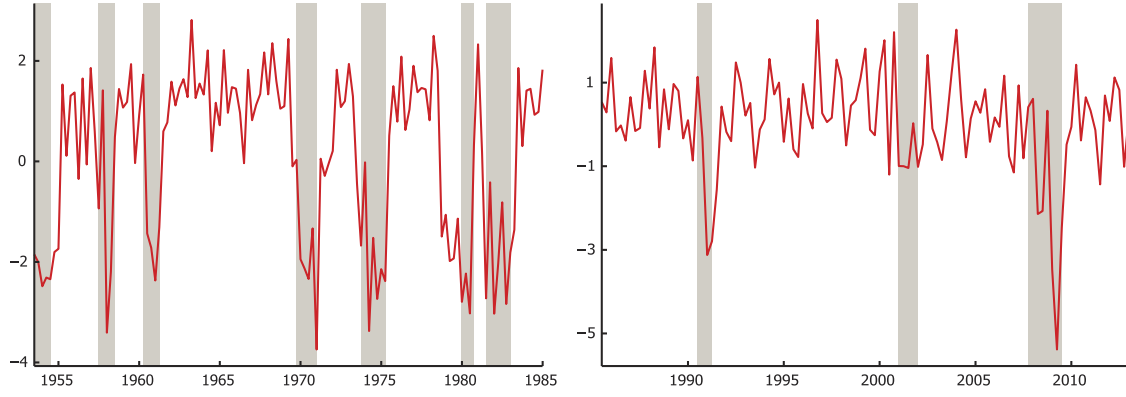
Fig. 13. NBER recession periods and extracted latent factor. Notes: Fig. 13 presents the latent factor determining the states extracted from the endogenous mean switching model, which is compared with the recession periods identified by NBER. The left hand side graph presents extracted latent factor plotted with solid red line and NBER recession periods displayed as shaded gray areas for the sample period 1952–1984, while the graph on the right presents those for the more recent sample period 1984–2012.
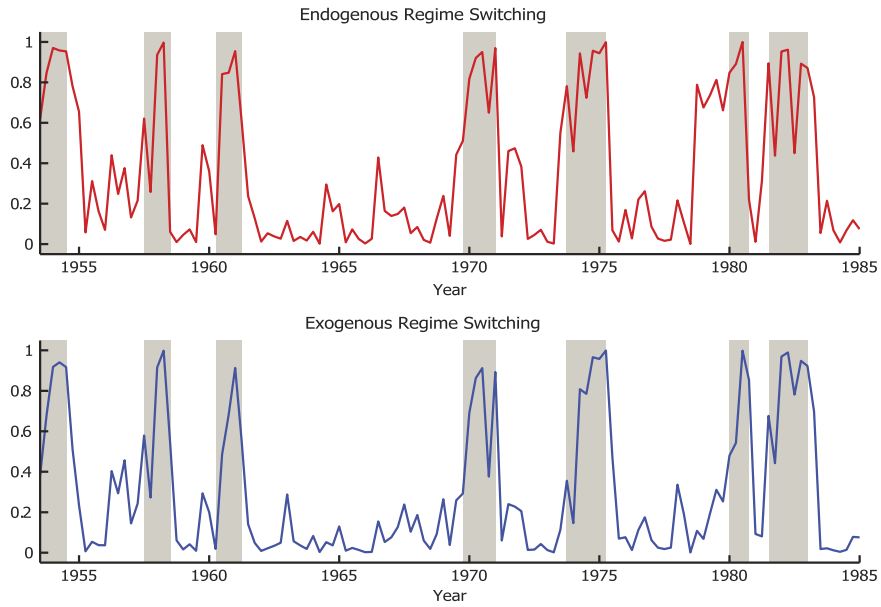
Fig. 14. Recession probabilities during 1952–1984. Notes: Fig. 14 presents the recession probabilities for the earlier sample period, 1952–1984. Top panel plots the recession probabilities from the endogenous mean switching model with red line, while the bottom panel plots those from its conventional exogenous counterpart with blue line. Both panels also show the recession periods identified by NBER.

it follows that

$$p\left(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1} \mathcal{F}_{t-1}\right)$$

$$= \frac{\left(1 - \Phi\left(\sqrt{\frac{1-\rho^2+\alpha^2\rho^2}{1-\rho^2}}\left(\tau - \frac{\alpha(w_t - \rho u_{t-1})}{1-\rho^2+\alpha^2\rho^2}\right)\right)\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}$$

$$\times \mathbb{N}\left(\rho u_{t-1}, \frac{1-\rho^2+\alpha^2\rho^2}{1-\alpha^2}\right)$$

$$p\left(w_t | s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1} \mathcal{F}_{t-1}\right)$$

$$= \frac{\Phi\left(\sqrt{\frac{1-\rho^2+\alpha^2\rho^2}{1-\rho^2}}\left(\tau - \frac{\alpha(w_t - \rho u_{t-1})}{1-\rho^2+\alpha^2\rho^2}\right)\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}$$

$$\times \mathbb{N}\left(\rho u_{t-1}, \frac{1-\rho^2+\alpha^2\rho^2}{1-\alpha^2}\right).$$

If $|\alpha| < 1$ and $|\rho| = 1$, we have

$$p\left(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}\right)$$

$$= p\left(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, y_{t-1}, \ldots, y_{t-k-1}\right)$$

$$= p\left(w_t | w_{t-1} > \tau, u_{t-1}\right)$$

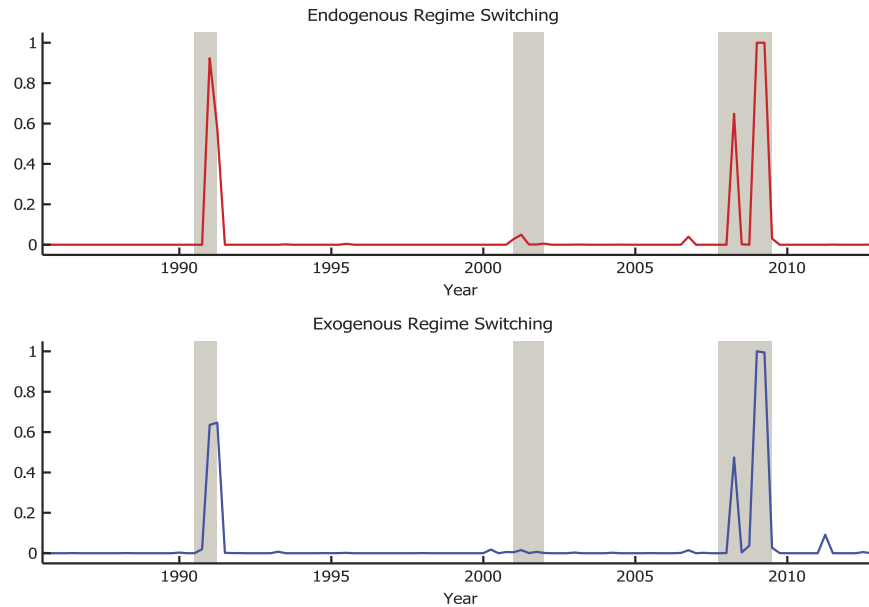$$= p\left(\alpha w_{t-1} + v_t | w_{t-1} > \tau, v_t\right), \tag{26}$$

and since

$$p\left(w_{t-1} | w_{t-1} > \tau\right) = \frac{\sqrt{1-\alpha^2}\phi\left(w_{t-1}\sqrt{1-\alpha^2}\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

it follows that

$$p\left(w_t | s_{t-1} = 1, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}\right)$$

$$= \frac{\frac{\sqrt{1-\alpha^2}}{\alpha}\phi\left(\frac{w_t - \rho u_{t-1}}{\alpha}\sqrt{1-\alpha^2}\right)}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)},$$

$$p\left(w_t | s_{t-1} = 0, s_{t-2}, \ldots, s_{t-k-1}, \mathcal{F}_{t-1}\right)$$

$$= \frac{\frac{\sqrt{1-\alpha^2}}{\alpha}\phi\left(\frac{w_t - \rho u_{t-1}}{\alpha}\sqrt{1-\alpha^2}\right)}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}.$$

Fig. 15. Recession probabilities during 1984–2012. Notes: Fig. 15 presents the recession probabilities for the recent sample period, 1984–2012. Top panel plots the recession probabilities from the endogenous mean switching model with red line, while the bottom panel plots those from its conventional exogenous counterpart with blue line. Both panels also show the recession periods identified by NBER.

The proof for the case of $\alpha = 1$ is entirely analogous except that $w_t =_d \mathbb{N}(0, t)$.

Appendix B. Additional Figures

See Figs. 10–15.

References

Bazzi, M., Blasques, F., Koopman, S.J., Lucas, A., 2014. Time Varying Transition Probabilities for Markov Regime Switching Models, Tinbergen Institute Discussion Paper TI 2014-072/III.

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econometrics 75, 79–97.

Chib, S., Dueker, M., 2004. Non-Markovian Regime Switching with Endogenous States and Time-Varying State Strengths, Working Paper 2008-31, Federal Reserve Bank of St. Louis.

Cho, J. S., White, H., 2007. Testing for regime switching. Econometrica 75, 1671–1720.

Diebold, F., Lee, J.-H., Weinbach, G., 1994. Regime switching with time-varying transition probabilities. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, pp. 283–302.

Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. Internat. Econom. Rev. 39, 763–788.

Garcia, R., Perron, P., 1996. An analysis of real interest rate under regime shift. Rev. Econom. Stat. 78, 111–125.

Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hamilton, J., 1996. Specification testing in Markov-switching time-series models. J. Econometrics 70, 127–157.

Hansen, B. E., 1992. The likelihood ratio test under non-standard conditions. J. Appl. Econometrics 7, S61–82.

Kalliovirta, L., Meitz, M., Saikkonen, P., 2015. A Gaussian mixture autoregressive model. J. Time Series Anal. 36, 247–266.

Kang, K. H., 2014. State-Space models with endogenous Markov regime switching parameters. Econom. J. 17, 56–82.

Kim, C.-J., 1994. Dynamic linear models with Markov-switching. J. Econometrics 60, 1–22.

Kim, C.-J., 2004. Markov-switching models with endogenous explanatory variables. J. Econometrics 122, 127–136.

Kim, C.-J., 2009. Markov-switching models with endogenous explanatory variables II: A Two-Step MLE procedure. J. Econometrics 148, 46–55.

Kim, C.-J., Nelson, C., 1999. State-Space models with regime switching. MIT Press.

Kim, C.-J., Nelson, C. R., Startz, R., 1998. Testing for mean reversion in heteroskedastic data based on gibbs-sampling-augmented randomization. J. Empir. Finance 5, 131–154.

Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of Markov regime-switching regression models with endogeneous switching. J. Econometrics 143, 263–273.

Timmermann, A., 2000. Moments of Markov switching models. J. Econometrics 96, 75–111.

Yu, J., 2005. On leverage in a stochastic volatility model. J. Econometrics 127, 165–178.