

Regime Switching Model with Endogenous Autoregressive Latent Factor*

Yoosoon Chang[†] Yongok Choi[‡] Joon Y. Park[§]

Abstract

This paper introduces a model with regime switching, which is driven by an autoregressive latent factor correlated with the innovation to the observed time series. In our model, the mean or volatility process is switched between two regimes, depending upon whether the underlying autoregressive latent factor takes values above or below some threshold level. If the latent factor becomes exogenous, our model reduces to the conventional markov switching model, and therefore, our model may be regarded as an extended markov switching model allowing for endogeneity in regime switching. Our model is estimated by the maximum likelihood method using a newly developed modified markov switching filter. For both mean and volatility models that are frequently analyzed in markov switching framework, we demonstrate that the presence of endogeneity in regime switching is indeed strong and ubiquitous.

This version: April 8, 2015

JEL Classification: C13, C32

Key words and phrases: regime switching model, latent factor, endogeneity, mean reversion, leverage effect, maximum likelihood estimation, markov chain

*We thank the co-editor, Oliver Linton, an associate editor and an anonymous referee for helpful comments. We are also grateful for many useful comments to James Hamilton, Chang-Jin Kim, Frank Schorfheide and the participants at 2013 Princeton-QUT-SMU Conference on Measuring Risk (Princeton), 2013 SETA (Seoul), Conference on Stochastic Dominance & Related Themes (Cambridge, UK), 2013 African Econometric Society Meeting (Accra), 2013 MEG Meeting (Bloomington), 2014 Hitotsubashi Conference on Econometrics for Macroeconomics and Finance, 2014 Workshop on Time Series Econometrics (Frankfurt), 2014 ESAM (Hobart), 2014 NSF-NBER Time Series Conference (St. Louis FED), and 2014 CEF (Oslo), and the seminar attendants at UC Riverside, University of Washington, VU Amsterdam-Tinbergen, Groningen, CORE, Humboldt, Chung-Ang, Keio, Bank of Japan, Bundesbank, ECB, Toulouse School of Economics, Bank of Portugal, Durham, Queen Mary University London, Yale, Atlanta FED, and Bank of Canada.

[†]Department of Economics, Indiana University.

[‡]Department of Financial Policy, Korea Development Institute.

[§]Department of Economics, Indiana University and Sungkyunkwan University.

1 Introduction

Regime switching models have been used extensively. In most of these models, two regimes, designated as high and low states of an economy, are introduced with a state process determining one of the regimes to take place in each period. The bi-valued 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), which was further analyzed in Kim (1994). Subsequently, the markov switching has been introduced in 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 a nice overview and some extensions 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 the markov switching models have been used and proven to be quite useful in a wide range of contexts, they have some drawbacks. Most importantly, with a very few exceptions including Diebold et al. (1994)

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 remedies both of the aforementioned shortcomings in the conventional markov switching model, and yields a broad class of models with endogenous and possibly non-stationary regime changes. Moreover, it also provides an extracted autoregressive latent factor, which can be 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, our model reduces to the conventional markov switching model. Indeed, we show in this case that the conventional two state markov switching model specified by two transition probabilities has the exact one-to-one correspondence with our model specified by the autoregressive coefficient of the latent factor and the threshold level. Therefore, we may always find our 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 in the paper 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.

Our paper is closely related to Kim et al. (2008), which considers a regime switching model driven by an endogenous i.i.d. latent factor with the threshold level determined by the previous state and possibly lagged values of the underlying time series.³ Our model has some important advantages over their model. First, they require the state transition to be dependent only on its immediate past, and this is in contrast with our approach which allows for a high order markov structure.⁴ Second, the innovation in our model is set to be correlated with the state variable in the next period, in contrast to their model where it is assumed to be contemporaneously correlated with the state variable in the current period. We believe that the endogeneity of regime switching is more appropriately structured in our approach. In fact, the presence of contemporaneous correlation between the state variable and the innovation of the error term makes their regression model seriously misspecified from

³In their model, as well as in our model, the threshold level is also allowed to be dependent upon other exogenous covariates.

⁴This is a serious restriction. For instance, their model cannot be used directly to fit a mean switching AR($\mathbf{p}$) model with $\mathbf{p} > 1$, including the original Hamilton's model.

the conventional point of view.⁵ Finally, we may easily allow for nonstationary transition by letting our autoregressive latent factor have a unit root, whereas their model strictly requires stationarity in transition.

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 significant deleterious effect on the estimates of model parameters and transition probabilities. This is more so for the mean model than 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 in the data on the latent states and their transitions. The efficiency gain and the precision enhancement are substantial in some cases, particularly when the latent factor is stationary and the endogeneity is strong. Finally, the likelihood ratio tests for endogeneity work reasonably well in all cases we consider. Though they tend to overreject

other hand, the movement of the state dependent mean at the second level would entail an unstabilizing effect on the observed time series. For both mean and volatility models, the inferred probabilities appear to be much more accurately predicting the true states of changing 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 respectively our mean and volatility models. Section 6 concludes the paper, and Appendix collects the proofs of theorems in the paper and additional figures.

A word on notation. $W(m)$ and $\Phi(p)$ respectively denote the distribution function of standard normal distribution. The equality is written as $=_d$. Moreover, we use $p(\cdot)$ or $p(\cdot|\cdot)$ as the generic notation.

level, or more compactly,

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

where τ and $(\underline{\pi}, \bar{\pi})$ are parameters, $\pi : \mathbb{R} \rightarrow \{\underline{\pi}, \bar{\pi}\}$, and $1\{\cdot\}$ is the indicator function. In subsequent discussions of our models, we interpret two events $\{w_t < \tau\}$ and $\{w_t \geq \tau\}$ as two regimes that are switched by the realized value of the latent factor (w_t) and the level τ of threshold, and call π the level function of state dependent parameter (π_t) .

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

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

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. The state process (s_t) represents 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. In contrast, it will be allowed in our approach to be endogenous, which appears to be much more realistic in a wide range of models used in practical applications.

For identification of the level function π 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 it a strictly stationary process.

We specify our model as

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

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. as

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

with unknown parameter ρ .⁶ For the brevity of notation, we write

$$m_t = m(x_t, y_{t-1}, \dots, y_{t-k}, w_t, \dots, w_{t-k}) = m(x_t, y_{t-1}, \dots, y_{t-k}, s_t, \dots, s_{t-k}) \quad (6)$$

$$\sigma_t = \sigma(x_t, w_t, \dots, w_{t-k}) = \sigma(x_t, s_t, \dots, s_{t-k}), \quad (7)$$

subsequently. Note that m_t and σ_t are conditional mean and volatility of the state dependent variable (y_t) given present and past values of latent factors $w_t, \dots, w_{t-k}$, as well as the current values of exogenous variables x_t and lagged endogenous variables $y_{t-1}, \dots, y_{t-k}$.⁷

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

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

where $\gamma(z) = 1 - \gamma_1 z - \dots - \gamma_k z^k$ is a k -th order polynomial, $\mu_t = \mu(w_t) = \mu(s_t)$ and $\sigma_t = \sigma(w_t) = \sigma(s_t)$ are respectively the state dependent mean and volatility of (y_t) . 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

⁶We may equivalently define the dynamics of $(\mathbf{w}_t)$ as $\mathbf{w}_t = \alpha \mathbf{w}_{t-1} + \rho \mathbf{u}_{t-1} + \sqrt{1 - \rho^2} \mathbf{v}_t$, where $(\mathbf{u}_t)$ and $(\mathbf{v}_t)$ are independent i.i.d. standard normals.

⁷The endogenous latent factor $(\mathbf{w}_t)$ may naturally be regarded as an economic fundamental determining the regimes of an economy.

considered by Hamilton (1989) and many others, except that the states in our model (8) are determined by an endogenous latent autoregressive factor (w_t) and the level function π specified as in (1) and (2), respectively. 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, and we may represent it as the standard regime switching model driven by an exogenous two state markov chain. 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 . We observe that ρ is significantly different from zero both in mean and volatility models for many economic and financial time series including the GDP growth rates and stock returns we analyze for our empirical illustrations in the paper.

The interpretation of endogeneity parameter ρ , especially its sign, is quite straightforward for our volatility model. If $\rho < 0$, the innovation u_t in the level 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 shock to (y_t) in the current period would entail an increase in volatility in the next period. This is often referred to as the leverage effect, if (y_t) is used to model returns from a financial asset. See, e.g., Yu (2005) for more discussions on the econometric modeling of leverage effect in volatility model for 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 cases of $\rho < 0$ and $\rho > 0$. 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 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 in two distinct levels: the reversion of (y_t) to its state dependent mean

(μ_t) , and the movement of (μ_t) to offset the effect of a shock to (y_t) . 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, the regime switching is more likely to happen for $\rho > 0$ if (y_t) is between the two state dependent means, whereas it is so for $\rho < 0$ if (y_t) is outside of the two state dependent means. Therefore, we may expect that regime switching becomes relatively more conspicuous if $\rho < 0$, compared with the case $\rho > 0$.

Kim et al. (2008) consider a regression model similar to ours in (4), yet they specify the state dependent regression coefficients (β_t) in their model as being dependent only on the current state variable (s_t) and their model is not directly applicable for model (8) with $k \geq 1$. 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 $(\sigma_t u_t)$.⁸ 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. 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 on their model: State dependent coefficients (β_t) of regressors are contemporaneously correlated with regression errors $(\sigma_t u_t)$, in addition to that regression errors $(\sigma_t u_t)$ are serially correlated.⁹ Their regression model is therefore seriously misspecified from the conventional point of view.

Our approach is different. In our model, the state process (s_t) is driven by an endogenous autoregressive latent factor, instead of an independent and identically distributed sequence of random variables. One important consequence of modeling the latent factor as an endogenous autoregressive process is that (s_t) alone is no longer markovian: It is markovian only jointly with the underlying time series (y_t) , and consequently, for our model in (8) the conditional distribution of s_t is determined by the past observations of the state dependent variable y_{t-i} 's as well as the past states s_{t-i} 's for $1 \leq i \leq k+1$ at any time t . This is shown more explicitly in the next section. Furthermore, 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 not only appears to be more realistic, but also yields a model that is correctly specified as a conventional regression model. Note that (m_t) and (σ_t) become respectively the mean and volatility of (y_t) in our model (4).

⁸For an easier comparison, we present their model using our notation. Their model also includes other predetermined variables, which we ignore here to more effectively contrast their approach with ours. As we explain in more detail later, our model may also easily accommodate the presence of other covariates.

⁹Note also that (σ_t) does not represent the conditional volatility of their error process $(\sigma_t \mathbf{u}_t)$, since (σ_t) is contemporaneously correlated with $(\mathbf{u}_t)$.

2.2 Relationship with Conventional Markov Switching Model

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

$$\begin{aligned} \mathbb{P}\{s_t = 0 \mid w_{t-1}\} &= \mathbb{P}\{w_t < \tau \mid w_{t-1}\} = \Phi(\tau - \alpha w_{t-1}) \\ \mathbb{P}\{ \end{aligned} \tag{9}$$

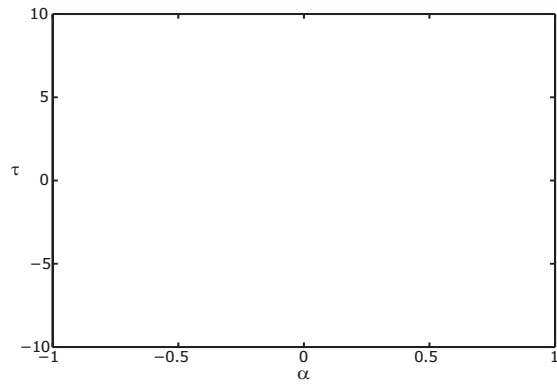
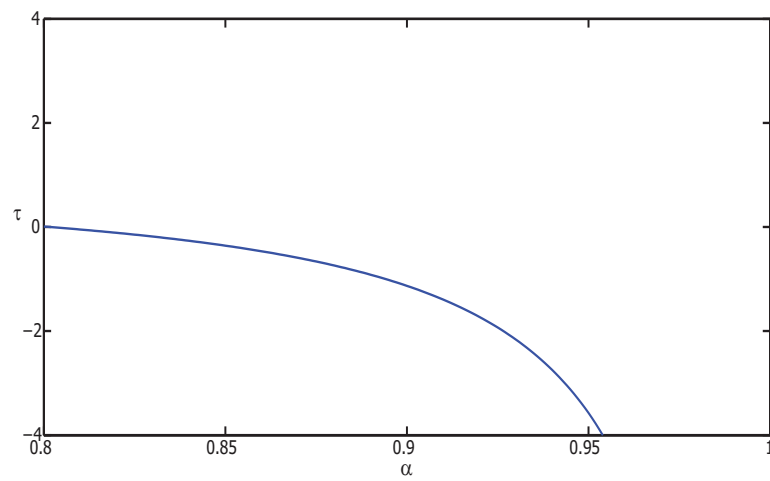
Figure 1: Contours of Transition Probabilities in (α, τ) -Plane

Figure 2: Correspondence Between (α, τ) and $(a(\alpha, \tau), b(\alpha, \tau))$



$\tau \leq 0$. *Moreover, we have*

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

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

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 low state $\omega(s_{t-1})$ given by

$$\omega(s_{t-1}) = \left[(1 - s_{t-1}) \int^{\tau/}$$

where

$$p(x, y) = \frac{1}{2\pi} \exp\left(-\frac{(1+c^2)x^2 + 2cxy + y^2}{2}\right) = \mathbb{N}\left(0, \begin{pmatrix} 1 & -c \\ -c & 1+c^2 \end{pmatrix}\right).$$

Therefore, the integrals can be solved, if bivariate normal distribution function is provided. Note that

$$\int_a^\infty \Phi(b+cx)\varphi(x)dx = M(-a, b, -c),$$

which can also be easily obtained, once we compute the integral in (14) for all $a, b, c \in \mathbb{R}$.

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

$$\ell(y_1, \dots, y_n) = \log p(y_1) + \sum_{t=2}^n \log p(y_t | \mathcal{F}_{t-1}) \quad (15)$$

where $\mathcal{F}_t = \sigma(x_t, (y_s)_{s \leq t})$, i.e., the information given by $x_t, y_1, \dots, y_t$ for each $t = 1, \dots, n$. Of course, the log-likelihood function includes a vector of unknown parameters $\theta \in \Theta$, say, which specifies $m_t = m(x_t, y_{t-1}, \dots, y_{t-k}, w_t, \dots, w_{t-k}) = m(x_t, y_{t-1}, \dots, y_{t-k}, s_t, \dots, s_{t-k})$ and $\sigma_t = \sigma(x_t, w_t, \dots, w_{t-k}) = \sigma(x_t, s_t, \dots, s_{t-k})$. It is, however, suppressed for the sake of notational brevity. The maximum likelihood estimator $\hat{\theta}$ of θ is given by

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

as usual. For the model we consider in (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 $(\gamma_1, \dots, \gamma_k)$.

To estimate our general switching model in (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

$$\Phi_\rho(x) = \Phi(x/\sqrt{1-\rho^2}) \quad (16)$$

for $|\rho| < 1$. 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*

$$\begin{aligned} & p(s_t, y_t | s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= p(y_t | s_t, \dots, s_{t-k}, y_{t-1}, \dots, y_{t-k}) p(s_t | s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}), \end{aligned}$$

where

$$p(y_t | s_t, \dots, s_{t-k}, y_{t-1}, \dots, y_{t-k}) = \mathbb{N}(m_t, \sigma_t^2) \quad (17)$$

and

$$\begin{aligned} & p(s_t | s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= (1 - s_t) \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &+ s_t [1 - \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1})] \end{aligned} \quad (18)$$

with the transition probability ω_ρ of the endogenous state process (s_t) to low state. If $|\alpha| < 1$, ω_ρ is given by

$$\begin{aligned} & \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= \frac{\left[(1 - s_{t-1}) \int_{-\infty}^{\tau \sqrt{1-\alpha^2}} + s_{t-1} \int_{\tau \sqrt{1-\alpha^2}}^{\infty} \right] \Phi_\rho \left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} - \frac{\alpha x}{\sqrt{1-\alpha^2}} \right) \varphi(x) dx}{(1 - s_{t-1}) \Phi(\tau \sqrt{1-\alpha^2}) + s_{t-1} [1 - \Phi(\tau \sqrt{1-\alpha^2})]}, \end{aligned}$$

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 $t \geq 2$,

$$\begin{aligned} & \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= \frac{\left[(1 - s_{t-1}) \int_{-\infty}^{\tau/\sqrt{t-1}} + s_{t-1} \int_{\tau/\sqrt{t-1}}^{\infty} \right] \Phi_\rho \left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} - x \sqrt{t-1} \right) \varphi(x) dx}{(1 - s_{t-1}) \Phi(\tau/\sqrt{t-1}) + s_{t-1} [1 - \Phi(\tau/\sqrt{t-1})]}. \end{aligned}$$

Theorem 3.1 fully specifies the joint transition of (s_t) and (y_t) in case of $|\rho| < 1$.¹⁰

If $|\rho| = 1$, we have perfect endogeneity and Φ_ρ in (16) 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

¹⁰It is clearly seen from Theorem 3.1 that the parameters α , τ and ρ in our model are all identified.

(s_t) given by the density $p(s_t|s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1})$, which is derived above for $|\rho| < 1$ in Theorem 3.1, is no longer applicable. When $|\rho| = 1$, the transition probability to low state $\omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1})$ behaves differently, which in turn implies that transition density of the state process needs to be modified accordingly. The transition probability to low state ω_ρ in this case 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 of the endogenous state process (s_t) to low state $\omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1})$ on the previous states and past observed time series, introduced in Theorem 3.1, is given as follows:*

(a) If $\alpha = 0$

$$\omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) = \begin{cases} 1, & \text{if } \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} < \tau \\ 0, & \text{otherwise} \end{cases}$$

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

$$\begin{aligned} & \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= (1 - s_{t-1}) \min \left(1, \frac{\Phi \left(\left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \right) \frac{\sqrt{1-\alpha^2}}{\alpha} \right)}{\Phi \left(\tau \sqrt{1-\alpha^2} \right)} \right) \\ &+ s_{t-1} \max \left(0, \frac{\Phi \left(\left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \right) \frac{\sqrt{1-\alpha^2}}{\alpha} \right) - \Phi \left(\tau \sqrt{1-\alpha^2} \right)}{1 - \Phi \left(\tau \sqrt{1-\alpha^2} \right)} \right) \end{aligned}$$

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

$$\begin{aligned} & \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= s_{t-1} \min \left(1, \frac{1 - \Phi \left(\left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \right) \frac{\sqrt{1-\alpha^2}}{\alpha} \right)}{1 - \Phi \left(\tau \sqrt{1-\alpha^2} \right)} \right) \\ &+ (1 - s_{t-1}) \max \left(0, \frac{\Phi \left(\tau \sqrt{1-\alpha^2} \right) - \Phi \left(\left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \right) \frac{\sqrt{1-\alpha^2}}{\alpha} \right)}{\Phi \left(\tau \sqrt{1-\alpha^2} \right)} \right) \end{aligned}$$

(d) If $\alpha = 1$, for $t = 1$, $\omega_\rho(s_0, y_0) = \Phi(\tau - \rho(y_0 - m_0)/\sigma_0)$ with $\mathbb{P}\{s_0 = 0\} = 1$ and

$\mathbb{P}\{s_0 = 1\} = 1$ *respectively when $\tau > 0$ and $\tau \leq 0$ and, for $t \geq 2$,*

$$\begin{aligned} & \omega_\rho(s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\ &= \begin{cases} 1 - s_{t-1}, & \text{if } \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} > 0 \\ \frac{\Phi\left(\left(\tau - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}}\right) \frac{1}{\sqrt{t-1}}\right) - s_{t-1} \Phi(\tau/\sqrt{t-1})}{(1 - s_{t-1}) \Phi(\tau/\sqrt{t-1})} & \text{otherwise} \end{cases} \end{aligned}$$

where $p(y_t|s_t, \dots, s_{t-k}, \mathcal{F}_{t-1})$ is given by (17), and we may readily obtain $p(s_t, \dots, s_{t-k}|\mathcal{F}_t)$ from $p(s_t, \dots, s_{t-k}|\mathcal{F}_{t-1})$ and $p(y_t|\mathcal{F}_{t-1})$.

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 (20) and (21). In the prediction step, we note that

$$p(w_t, s_{t-1}, \dots, s_{t-k}|\mathcal{F}_{t-1}) = \sum_{s_{t-k-1}} p(w_t|s_{t-1}, \dots, s_{t-k-1}, \mathcal{F}_{t-1})p(s_{t-1}, \dots, s_{t-k-1}|\mathcal{F}_{t-1}). \quad (22)$$

Since $p(s_{t-1}, \dots, s_{t-k-1}|\mathcal{F}_{t-1})$ is obtained from the previous updating step, we may readily compute $p(w_t, s_{t-1}, \dots, s_{t-k}|\mathcal{F}_{t-1})$ from (22) once we find $p(w_t|s_{t-1}, \dots, s_{t-k-1}, \mathcal{F}_{t-1})$, 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 latent factor (w_t) on previous states and past observed time series is given as follows:*

(a) When $|\alpha| < 1$ and $|\rho| < 1$,

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

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

$$\begin{aligned}
 & p(w_t | s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
 &= \begin{cases} \frac{\frac{\sqrt{1-\alpha^2}}{\alpha} \phi\left(\frac{w_t - \rho \frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}}{\alpha} \sqrt{1-\alpha^2}\right)}{1-\Phi(\tau\sqrt{1-\alpha^2})}, & \text{if } w_t \geq \alpha\tau + \rho \frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}} \\ 0, & \text{otherwise,} \end{cases} \\
 & p(w_t | s_{t-1} = 0, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
 &= \begin{cases} \frac{\frac{\sqrt{1-\alpha^2}}{\alpha} \phi\left(\frac{w_t - \rho \frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}}{\alpha} \sqrt{1-\alpha^2}\right)}{\Phi(\tau\sqrt{1-\alpha^2})}, & \text{if } w_t \leq \alpha\tau + \rho \frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}} \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}$$

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

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

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

$$\begin{aligned}
& p(w_t | s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \begin{cases} \frac{\frac{1}{\sqrt{t-1}} \phi\left(\frac{w_t - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}}}{t-1}\right)}{1 - \Phi(\tau/\sqrt{t-1})}, & \text{if } w_t \geq \tau + \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \\ 0, & \text{otherwise,} \end{cases} \\
& p(w_t | s_{t-1} = 0, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= \begin{cases} \frac{\frac{1}{\sqrt{t-1}} \phi\left(\frac{w_t - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}}}{t-1}\right)}{1 - \Phi(\tau/\sqrt{t-1})}, & \text{if } w_t \leq \alpha\tau + \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1}} \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

We may then obtain

$$p(w_t, s_{t-1}, \dots, s_{t-k} | \mathcal{F}_t) = \frac{p(y_t | w_t, s_{t-1}, \dots, s_{t-k}, \mathcal{F}_{t-1}) p(w_t, s_{t-1}, \dots, s_{t-k} | \mathcal{F}_{t-1})}{p(y_t | \mathcal{F}_{t-1})}, \quad (23)$$

in the updating step. By marginalizing $p(w_t, s_{t-1}, \dots, s_{t-k} | \mathcal{F}_t)$ in (23), we get

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

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 $t = 1, 2, \dots$. 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 level of threshold for regime switching is therefore given as a linear function of some predetermined and observable covariates. All of our previous results extend to this more general model only with some trivial modifications. Since the required modifications are quite clear, we do not explain them in detail here. This model is more directly comparable to the one considered in Kim et al. (2008).

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. 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, and we need to use a density based filter to estimate the parameters.

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. For the volatility model, we consider

$$y_t = \sigma(s_t)u_t, \quad \sigma(s_t) = \underline{\sigma}(1 - s_t) + \bar{\sigma}s_t. \quad (24)$$

The parameters $\underline{\sigma}$ and $\bar{\sigma}$

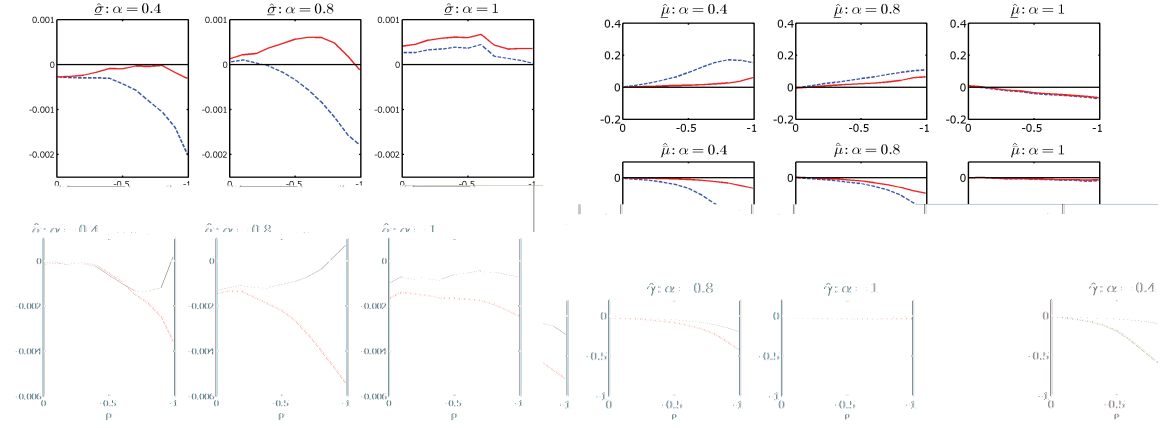
ρ between the current model innovation u_t and the next period innovation v_{t+1} of the latent autoregressive factor is set to be negative for both mean and volatility models, as in most of our empirical results reported in the next section. To more thoroughly study the impact of endogeneity on the estimation of our model parameters, we allow ρ to vary from 0 to -1 in the increment of 0.1. On the other hand, we consider three pairs of the autoregressive coefficient α of the latent factor and the threshold level τ given by $(\alpha, \tau) = (0.4, 0.5), (0.8, 0.7), (1, 9.63)$. The first two pairs with $|\alpha| < 1$ yield stationary latent factors, 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 low to low state and from high to 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.¹¹ This, in particular, implies that the unconditional probabilities of the state being in 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 with no existing invariant distribution. Our choice of $\tau = 9.63$ in the third pair yields the unconditional probabilities $(2/3, 1/3)$ of low and high regimes at the terminal period of our simulation, which makes it comparable to the first two pairs.¹²

4.2 Simulation Results

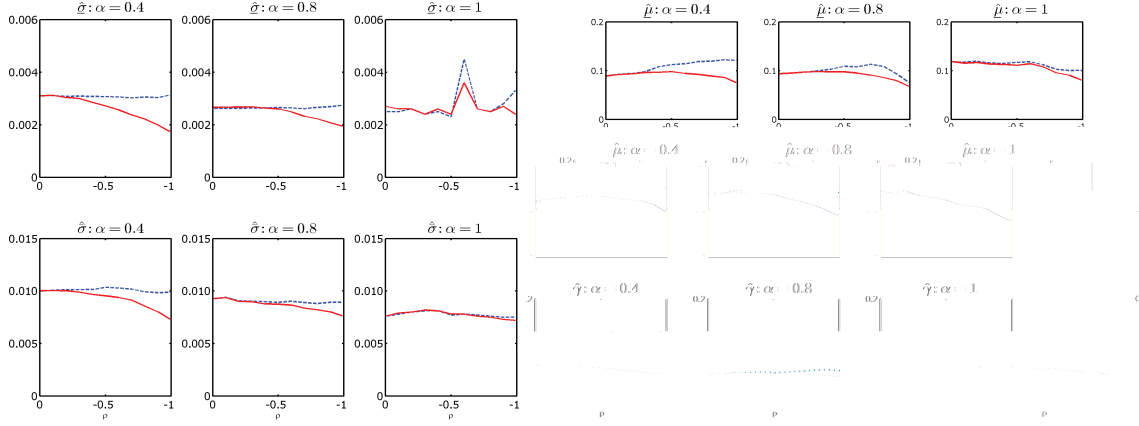
In our simulations, we first examine the endogeneity bias. The estimators of parameters in our models are expected to be biased if the presence of endogeneity in regime switching is ignored. To see the magnitude of bias resulting from the neglected endogeneity in regime switching, we let $\rho = 0$ for the exogenous regime switching models. Our simulation results are summarized in Figure 3. On the left panel of Figure 3, the bias in the maximum likelihood estimates $\hat{\underline{\sigma}}$ and $\hat{\bar{\sigma}}$ of low and high volatility levels $\underline{\sigma}$ and $\bar{\sigma}$ in the volatility

Figure 3: Endogeneity Bias



Notes: On the left panel, the bias in ML estimates $\hat{\underline{\sigma}}$ and $\hat{\bar{\sigma}}$ of low and high volatility levels $\underline{\sigma}$ and $\bar{\sigma}$ 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 $\hat{\underline{\mu}}, \hat{\bar{\mu}}$ and $\hat{\gamma}$ of low and high mean levels and AR coefficient of observed time series, $\underline{\mu}, \bar{\mu}$ and

Figure 4: Efficiency Gain from Endogeneity



Notes: Respectively presented in the left and right panels of Figure 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 Figure 3.

to save space, our simulations show that the inferred probabilities of latent states are also affected seriously if the endogeneity in regime switching is not properly taken care of.

Not only can the presence of endogeneity give us 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 accumulated 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 Figure 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 Figure 4 in exactly the same manner as in Figure 3.

As shown in Figure 4, the efficiency gain from the presence of endogeneity in regime switching can be quite substantial. This is equally true for both mean and volatility models. For instance, if we set $\alpha = 0.8$, the standard deviations of the estimators $\hat{\mu}$ and $\hat{\sigma}$ from our

mean and volatility models having endogenous regime switching with $\rho = -0.9$ decrease by approximately 24% and 22%, respectively, if compared with the models having exogenous regime switching with $\rho = 0$. Of course, the presence of endogeneity yields efficiency gain, only when it is properly taken into account. If the conventional markov switching model is used, the presence of endogeneity in most cases has a negative 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 we have endogeneity 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.¹³

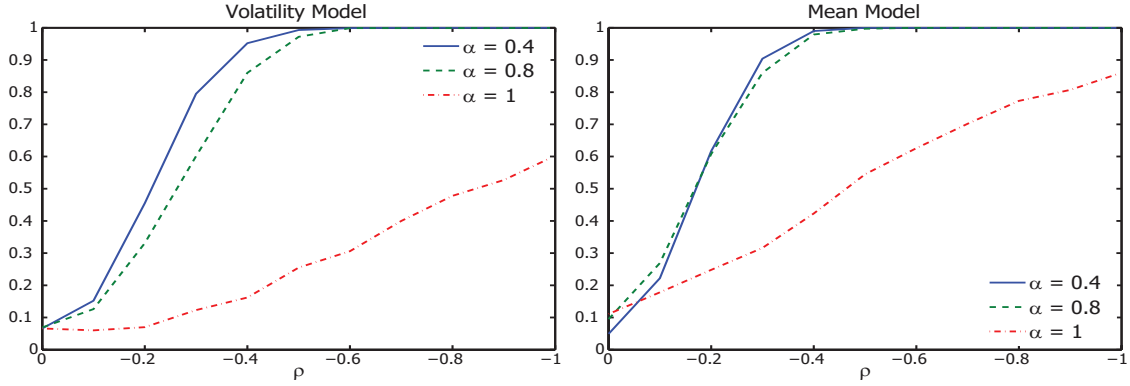
Finally, we consider testing for the presence of endogeneity in regime switching models based on the likelihood ratio test given by

$$2(\ell(\hat{\theta}) - \ell(\tilde{\theta})), \quad (26)$$

where ℓ signifies the log-likelihood function and the parameter θ with tilde and hat denote their maximum likelihood estimates with and without the no endogeneity restriction, $\rho = 0$. The likelihood ratio test has a chi-square limit distribution with one degrees of freedom. Presented in the left and right panels of Figure 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 actual power increasing rapidly as the value of endogeneity parameter $|\rho|$ gets large. Under the null hypothesis of no endogeneity, the test has good size properties overall 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. Though we do not report the details, the size distortion disappears as sample size increases. In contrast, the test does not work well when $\alpha = 1$ and the latent factor becomes nonstationary. In this

¹³On the average, the regime change occurs 160, 100, and 15 times out of 500, respectively, for 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.

Figure 5: Power Function of LR Test for Endogeneity



Notes: The left and right hand side graphs of Figure 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 $\alpha = 0.4, 0.8, 1$.

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 relatively much worse than the stationary cases.

5 Empirical Illustrations

To illustrate our approach empirically, we analyze the excess market returns using the volatility model with regime switching (24) studied in our earlier simulations, and the US GDP growth rates by a mean model with regime switching similar to (25) which is also examined in our simulations.

5.1 Stock Market Return Volatilities

For market returns, we consider the returns on NYSE/AMEX index from the Center for Research in Security Prices (CRSP). Specifically, we use the monthly series of value-weighted stock returns including dividend for the period from January 1926 to December 2012, along with the information on their quote date. For our analysis, we use the demeaned excess market returns (y_t) to fit the volatility model in (24).¹⁴ Table 1 reports the estimation results for the excess market return volatility model with regime switching for two sample

¹⁴To compute monthly excess return, we first obtain monthly risk free rate of return by continuously compounding daily risk free rate between the quotation dates. The number of days between quotation dates ranges from 28 to 33. CRSP provides monthly series of annualized yield to maturity (TMYTM), which is constructed from nominal price of three month treasury bill. Thus we obtain the yield to maturity at monthly frequency by first converting the annual yield to maturity to daily (TMYLD) by the conversion

periods: the full sample period (1926-2012) and the recent subsample period (1990-2012). We choose this subsample period to relate the extracted latent volatility factor obtained from our endogenous switching model with one of the most widely used volatility index VIX which is available only from 1990.

For the ML estimation of parameters and transition probabilities in the volatility switching model, we use our modified markov switching filter together with the numerical optimization method including the commonly used BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm. When likelihood function is highly nonlinear as in our switching models, it is well known that numerical optimization often fails to find global maximum and ends up with local maximum depending on the choice of initial values. However, we notice that in our switching models we can effectively profile out all the parameters but α and ρ , and once we fix α and ρ , our numerical maximization procedure successfully finds global maximum regardless of the choices on initial values.

For the exogenous model, we analyze profile likelihood in terms of α only, which is obtained by maximizing the likelihood for each value of α over all the other parameter values, and then maximize the profile likelihood over all α values considered. The left hand side graph of Figure 6 shows the profile likelihood function of α obtained from the exogenous switching model when we fix α at a value from 0.1 to 0.9 with 0.1 increment.¹⁵ For the endogenous switching model, we consider profile likelihood in terms of both α and ρ , and maximize it over all pairs of α and ρ considered. The graph on the right hand side of Figure 6 presents the profile likelihood surface plot obtained from our endogenous volatility switching model when we fix ρ as well as α . To obtain the surface plot, we consider 171 combinations of α and ρ from setting α at a value from 0.1 to 0.9 and ρ at a value from -0.9 to 0.9 with 0.1 increment.¹⁶ We may see clearly from Figure 6 that the profile likelihood

formula CRSP provides as

$$\mathbf{TMYLD}_t = \frac{1}{365} \frac{1}{100} \mathbf{TMYTM}_t,$$

and then continuously compounding this daily yield to obtain the monthly yield as

$$\mathbf{e}^{\mathbf{TMYLD}_{t-1} \times \mathbf{N}_t} - 1,$$

where $\mathbf{N}_t$ is the number of days between the quote date for the current month and the quote date $\mathbf{MCALDT}_{t-1}$ for the previous month. Finally, we obtain monthly excess market return data by subtracting the above monthly risk free rate from the market return.

¹⁵For each fixed value of α , we try 12 to 60 different initial value combinations for all the other parameters to find the maximum value for the profile likelihood function. However, regardless of the choices on initial values, the numerical maximization finds unique maximum once we fix α and the global maximum would be found around $\alpha = 0.9$.

¹⁶For each of these combinations of α and ρ , we try 7 to 35 different sets of initial values of all the other parameters to find the maximum. Our numerical method finds local maxima for some initial value

Figure 6: Profile Likelihood for Volatility Switching Model

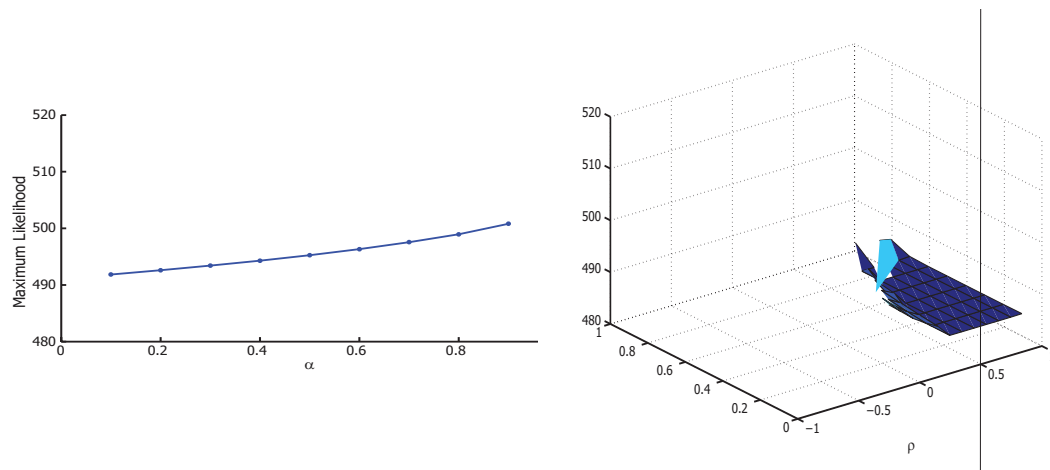


Table 1: Estimation Results for Volatility Model

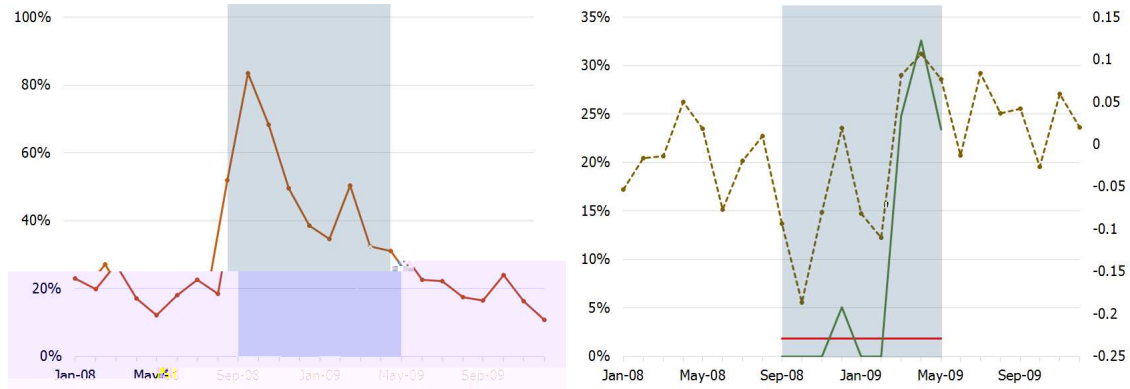
Sample Periods	1926-2012		1990-2012	
Endogeneity	Ignored	Allowed	Ignored	Allowed
$\underline{\sigma}$	0.039 (0.001)	0.038 (0.001)	0.022 (0.002)	0.025 (0.004)
$\overline{\sigma}$	0.115 (0.009)	0.115 (0.009)	0.050 (0.003)	0.055 (0.008)
ρ		-0.970 (0.086)		-0.999 (0.010)
log-likelihood	1742.536	1748.180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

τ in the exogenous model are 0.994 (0.004) and 11.375 (4.472), from which we have 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 Figure 7 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 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 Figure 7 which presents the transition probability from low volatility regime at $t - 1$ to 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 lagged excess market return. It shows in particular that the transition probabilities have been changing drastically, and reach as high as 87.1% at a time, which is 32 times bigger than its counterpart from the exogenous regime switching model. The right hand side graph of Figure 7 similarly illustrates the same point with the transition probability from high volatility state at $t - 1$ to high volatility state at t by the exogenous and endogenous switching models.

Figure 7: Estimated Transition Probability from Volatility Model

Figure 8: 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 Figure 8, where the green line signifies the time varying transition probability $\mathbb{P}(s_t = 0 | s_{t-1} = 1, y_{t-1})$ estimated from our endogenous regime switching model, while the red line corresponds to the constant transition probability $\mathbb{P}(s_t = 0 | s_{t-1} = 1)$ obtained from the exogenous switching model. The solid line on the left and the dashed line on the right graphs respectively present the time series of annualized monthly volatility and the monthly NYSE/AMEX index returns. The shaded areas on both graphs of Figure 8 indicate the high volatility regime.

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 early spring of 2009. The time varying transition probability produced from our endogenous switching model therefore represents the reality we observe much better, which in turn explains the higher precision and efficiency of the parameter estimates obtained from our endogenous regime switching model observed in our earlier simulation studies.

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 Figure 13 in Appendix, which presents the sample path of the extracted latent factor along with that of the CBOE (The Chicago Board Options Exchange) volatility index VIX for the period 1990-2012. VIX stayed relatively high during 1998-2004 and 2008 periods indicating that the volatility was high during those periods.

As shown in Figure 13, 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 the similar role played by VIX.

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 Figure 14 in Appendix. Overall the time series of the high volatility probabilities computed from both endogenous and exogenous switching models are similar. There are however some noticeable differences. In general, those obtained from our endogenous model tend to fluctuate more over time compared to those from its exogenous counterpart. Moreover, it is interesting to note that the probability of being in 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. Therefore we may also say that our endogenous switching model can be used as a warning system for detecting unusual instability in financial market.

5.2 GDP Growth Rates

In this section, we investigate 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:Q4.¹⁸ As in Hamilton (1989), we model the real GDP growth rates (y_t) as an AR(4) process similarly as in (25). Since there seems to be a structural break in the postwar U.S. 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. Kim and Nelson (1999) provides an empirical study of the regime switching model for US growth rates considered in Hamilton (1989) for the earlier sample period. We use the same data used in their study and compare our results with theirs.¹⁹

We estimate the mean switching model for the GDP growth rates using our new modified markov switching filter along with BFGS method. To ensure that we find the global maximum, we look at the profile likelihood as done in the volatility model for the market

¹⁸Source: Bureau of Economic Analysis, US Department of Commerce. The growth rate of real GDP is calculated as the first difference of log real GDP.

¹⁹We use the data provided in the website for Kim and Nelson (1999).

Figure 9: Profile Likelihood for Mean Switching Model

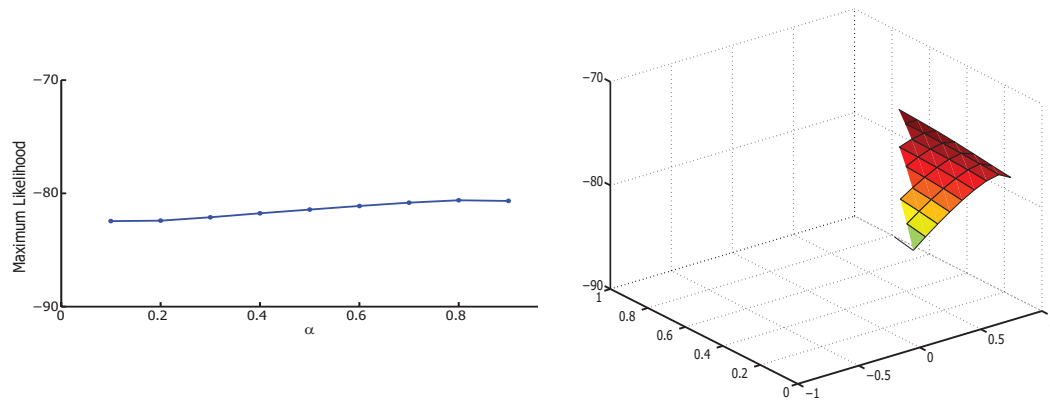
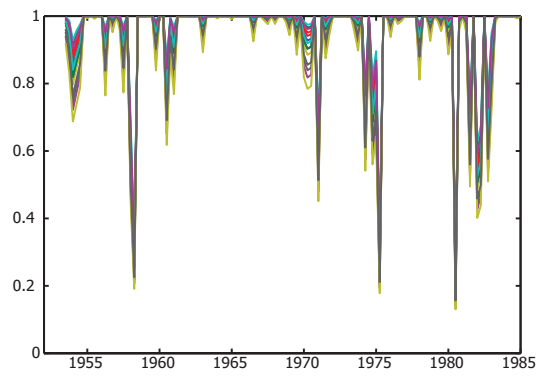


Table 2: Maximum Likelihood Estimates for Hamilton (1989) Model

Sample Periods		1952-1984		1984-2012	
Endogeneity		Ignored	Allowed	Ignored	Allowed
$\underline{\mu}$		-0.165 (0.219)	-0.083 (0.161)	-0.854 (0.298)	-0.758 (0.311)
$\overline{\mu}$		1.144	1.212	0.710	0.70T/R12061-1.2-15602.8Td() 3.0011(0)5

Figure 10: Estimated Transition Probabilities for Mean Model

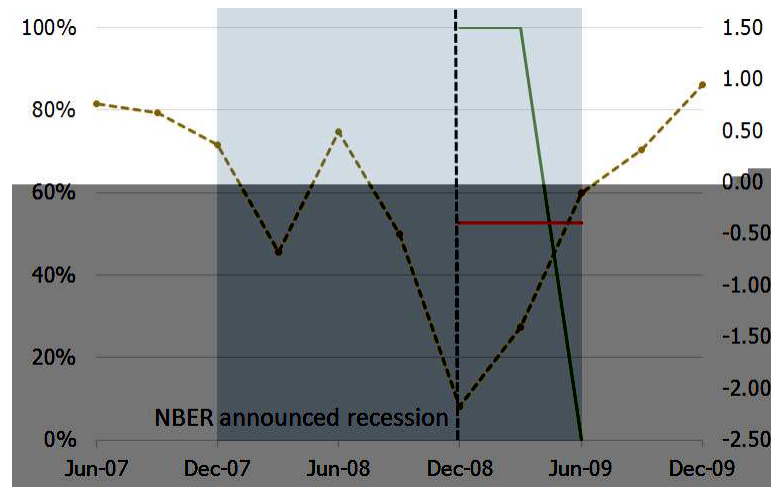


obtained from our endogenous regime switching model for each of the 16 possible realizations of the four lagged state variables, s_{t-2} , s_{t-3} , s_{t-4} , and s_{t-5} . Note that each of the four lagged state variables s_{t-2} , s_{t-3} , s_{t-4} , and s_{t-5} takes a value either 0 or 1, giving 16 possibilities for their joint realizations. We therefore calculate the transition probability from low state at $t-1$ to 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 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 Figure 10 is that the estimated transition probabilities estimated by our endogenous regime switching model are drastically different from the one obtained from its exogenous counterpart. The time varying transition probabilities estimated from our endogenous model are indeed much more sensible and realistic.

Figure 11 presents the transition probabilities from low mean regime to low mean regime plotted along with the US GDP growth rates. NBER announced on December 1, 2008 that a recession began in December 2007 after the December 2007 peak which defines the turning point from expansion to recession, and it announced on October 21, 2010 that the recession ended in June 2009. By the time of the official announcement by NBER on December 1, 2008, everyone is presumed to know that we were in recession. Using this information, we calculate the transition probability from low to low mean regime. The 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. Therefore, when the observed GDP growth rate is low, the transition probability from low to 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 Figure 15 in 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 times of the

Figure 11: Transition Probabilities During Recent Recession Period



Notes: Figure 11 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.

movements of the growth rates. This is why we see much higher values of the recession probabilities computed from our endogenous model compared to those obtained from the exogenous model during the period prior to the 1981 recession. Our model could see the downward movements in the growth rates during this period, and accordingly increased the recession probabilities, even before the growth rates become negative. There is no way, however, for the exogenous model to update the recession probabilities upward even when it was evident that the growth rates were sinking, and it has to wait until after the levels of growth rates become negative to raise the recession probabilities.

6 Conclusions

In the paper, we propose a new regime switching model based on an autoregressive latent factor. As we demonstrate in the paper, our approach has several clear advantages over the conventional regime switching model. Most importantly, we may allow for endogeneity in regime switching. The endogeneity we introduce by using our approach is well structured. It models the effect of a shock to the observed time series in a very natural manner. In the mean model with regime switching, the presence of endogeneity implies that the mean reversion may occur in two different levels: one at the level of reversion of the observed time series to its state dependent mean, and the other at the level of movement of the state dependent mean to offset the effect of a shock. In the volatility model with regime switching, on the other hand, the presence of endogeneity means the leverage effect. Furthermore, our regime switching model becomes observationally equivalent to the conventional markov switching model, if the endogeneity of regime switching is not present. 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 very strong and unambiguous. This implies, in particular, that it is worthwhile to refit any of the previously fitted conventional markov switching models, allowing for endogeneity in regime switching. Our extensive simulations make it clear that neglecting endogeneity in regime switching not only incurs a substantial bias in the estimates of model parameters, but also does it lead to significant information loss. 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 much 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 channeled 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.

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. *Journal of 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, Oxford, UK, pp. 283–302.
- Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review* 39, 763–788.
- 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. *Journal of Econometrics* 70, 127–157.
- Hansen, B. E., 1992. The likelihood ratio test under non-standard conditions. *Journal of Applied Econometrics* 7, S61–82.
- Kalliovirta, L., Meitz, M., Saikkonen, P., 2015. A gaussian mixture autoregressive model. *Journal of Time Series Analysis* 36, 247–266.
- Kang, K. H., 2014. State-space models with endogenous Markov regime switching parameters. *Econometrics Journal* 17, 56–82.
- Kim, C.-J., 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60, 1–22.

- Kim, C.-J., 2004. Markov-switching models with endogenous explanatory variables. *Journal of Econometrics* 122, 127–136.
- Kim, C.-J., 2009. Markov-switching models with endogenous explanatory variables II: A two-step mle procedure. *Journal of Econometrics* 148, 46–55.
- Kim, C.-J., Nelson, C., 1999. *State-Space Models with Regime Switching*. MIT Press, Cambridge, MA.
- Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of Markov regime-switching regression models with endogeneous switching. *Journal of Econometrics* 143, 263–273.
- Timmermann, A., 2000. Moments of markov switching models. *Journal of Econometrics* 96, 75–111.
- Yu, J., 2005. On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–178.

Appendix

Appendix A: Mathematical Proofs

Proof of Lemma 2.1

From (9), we may deduce that

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

from which it follows that

$$\begin{aligned} \mathbb{P}\{s_t = 0 \mid w_{t-1} < \tau\} &= \mathbb{P}\left\{s_t = 0 \mid w_{t-1}\sqrt{1-\alpha^2} < \tau\sqrt{1-\alpha^2}\right\} \\ &= \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \mathbb{P}\left\{s_t = 0 \mid w_{t-1}\sqrt{1-\alpha^2} = x\right\} \varphi(x) dx}{\mathbb{P}\left\{w_{t-1}\sqrt{1-\alpha^2} < \tau\sqrt{1-\alpha^2}\right\}} \\ &= \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right) \varphi(x) dx}{\Phi\left(\tau\sqrt{1-\alpha^2}\right)}, \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}\{s_t = 1 \mid w_{t-1} \geq \tau\} &= \mathbb{P}\left\{s_t = 1 \mid w_{t-1}\sqrt{1-\alpha^2} \geq \tau\sqrt{1-\alpha^2}\right\} \\ &= \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \mathbb{P}\left\{s_t = 1 \mid w_{t-1}\sqrt{1-\alpha^2} = x\right\} \varphi(x) dx}{\mathbb{P}\left\{w_{t-1}\sqrt{1-\alpha^2} \geq \tau\sqrt{1-\alpha^2}\right\}} \\ &= \frac{\int_{\tau\sqrt{1-\alpha^2}}^{\infty} \left[1 - \Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\right] \varphi(x) dx}{1 - \Phi\left(\tau\sqrt{1-\alpha^2}\right)}, \end{aligned}$$

since

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

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

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. $\square$

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

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

we may easily deduce that

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

It follows that

$$\begin{aligned} & \mathbb{P} \left\{ w_t < \tau \mid w_{t-1}, w_{t-2}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \\ &= \mathbb{P} \left\{ \frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \mid w_{t-1}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \left\{ \tau - \rho^{y_{t-1} - m_{t-1}} \right\} \\ & \quad \left[\frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \mid w_{t-1}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right] \end{aligned}$$

$$\begin{aligned} & \left\{ \frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}} \mid w_{t-1}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \\ &= \mathbb{P} \left\{ s_t = 0 \mid t-1 \text{ is } 0, y_{t-1} - m_{t-1}, \dots, y_{t-k-1} \right\} \\ & \quad \frac{\Phi_\rho \left(\frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \right)}{\Phi \left(\frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \right)} \left\{ \mathbb{P} \left\{ w_t < \tau \mid w_{t-1} \geq \tau, w_{t-2}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \right. \\ & \quad \left. = \mathbb{P} \left\{ w_t < \tau \mid w_{t-1} \sqrt{1-\rho^2} \right\} \right\} \end{aligned}$$

$$\begin{aligned} & \left\{ \frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}} \mid w_{t-1}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \\ &= \mathbb{P} \left\{ s_t = 0 \mid t-1 \text{ is } 1, y_{t-1} - m_{t-1}, \dots, y_{t-k-1} \right\} \\ & \quad \frac{\Phi_\rho \left(\frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \right)}{1 - \Phi \left(\frac{\frac{w_{t-1} - \alpha w_{t-2}}{\sigma_{t-1}} - \rho \frac{y_{t-1} - m_{t-1}}{\sigma_{t-1} \sqrt{1-\rho^2}}}{\sqrt{1-\rho^2}} \right)} \left\{ \mathbb{P} \left\{ w_t < \tau \mid w_{t-1} \geq \tau, w_{t-2}, \dots, w_{t-k-1}, y_{t-1}, \dots, y_{t-k-1} \right\} \right. \\ & \quad \left. = \mathbb{P} \left\{ w_t < \tau \mid w_{t-1} \sqrt{1-\rho^2} \right\} \right\} \end{aligned}$$

$$= \int^{\tau \sqrt{1-\rho^2}}$$

It follows from (17) that

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

Moreover, we have

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

as we have shown above. Therefore, it follows that

$$\begin{aligned} & p(s_t, y_t | s_{t-1}, \dots, s_1, y_{t-1}, \dots, y_1) \\ &= p(y_t | s_t, \dots, s_{t-k}, y_{t-1}, \dots, y_{t-k}) p(s_t | s_{t-1}, \dots, s_{t-(k+1)}, y_{t-1}, \dots, y_{t-k-1}) \\ &= p(s_t, y_t | s_{t-1}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \end{aligned}$$

and (s_t, y_t) is a $(k+1)$ -st order markov process. $\square$

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 $|\alpha| < 1$.

$\mathbb{I}_{t-1} \sqrt{}$

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

$$\begin{aligned}
\omega_\rho(s_{t-1} = 0, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) &= \mathbb{P}\{\alpha w_{t-1} + \rho u_{t-1} < \tau \mid w_{t-1} < \tau, u_{t-1}\} \\
&= \mathbb{P}\left\{\sqrt{1-\alpha^2}w_{t-1} < \frac{\sqrt{1-\alpha^2}}{\alpha}(\tau - \rho u_{t-1}) \mid \sqrt{1-\alpha^2}w_{t-1} < \tau\sqrt{1-\alpha^2}, u_{t-1}\right\} \\
&= \begin{cases} 1, & \text{if } \frac{1}{\alpha}\left(\tau - \rho\frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}\right) < \tau, \\ \frac{\Phi\left(\left(\tau - \rho\frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}\right)\frac{\sqrt{1-\alpha^2}}{\alpha}\right)}{\Phi(\tau\sqrt{1-\alpha^2})}, & \text{otherwise,} \end{cases}
\end{aligned}$$

where $u_{t-1} = (y_{t-1} - m_{t-1})/\sigma_{t-1}$. Similarly, we have

$$\begin{aligned}
\omega_\rho(s_{t-1} = 1, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) &= \mathbb{P}\{\alpha w_{t-1} + \rho u_{t-1} < \tau \mid w_{t-1} > \tau, u_{t-1}\} \\
&= \mathbb{P}\left\{\sqrt{1-\alpha^2}w_{t-1} < \frac{\sqrt{1-\alpha^2}}{\alpha}(\tau - \rho u_{t-1}) \mid \sqrt{1-\alpha^2}w_{t-1} > \tau\sqrt{1-\alpha^2}, u_{t-1}\right\} \\
&= \begin{cases} 0, & \text{if } \frac{1}{\alpha}\left(\tau - \rho\frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}\right) < \tau. \\ \frac{\Phi\left(\left(\tau - \rho\frac{y_{t-1}-m_{t-1}}{\sigma_{t-1}}\right)\frac{\sqrt{1-\alpha^2}}{\alpha}\right) - \Phi(\tau\sqrt{1-\alpha^2})}{1 - \Phi(\tau\sqrt{1-\alpha^2})}, & \text{otherwise.} \end{cases}
\end{aligned}$$

Proof of Corollary 3.3

We note that

$$\begin{aligned}
p(w_t \mid s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= p(w_t \mid s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\
&= p(w_t \mid w_{t-1} > \tau, u_{t-1}) \\
&= \frac{\int_\tau^\infty p(w_t, w_{t-1}, u_{t-1}) dw_{t-1}}{\int_\tau^\infty p(w_{t-1}, u_{t-1}) dw_{t-1}} \\
&= \frac{\int_\tau^\infty p(w_t \mid w_{t-1}, u_{t-1}) p(w_{t-1}) dw_{t-1}}{\int_\tau^\infty p(w_{t-1}) dw_{t-1}}. \tag{27}
\end{aligned}$$

The last equality comes from the independence between (w_t) and (u_t) . Likewise, we have

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

When $|\alpha| < 1$ and $|\rho| < 1$, we have $w_t|w_{t-1}, u_{t-1} =_d \mathbb{N}(\alpha w_{t-1} + \rho u_{t-1}, 1 - \rho^2)$. Since $w_t =_d w_{t-1} =_d \mathbb{N}\left(0, \frac{1}{1 - \alpha^2}\right)$, we can easily deduce

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

When $|\alpha| < 1$ and $|\rho| = 1$, we note that

$$\begin{aligned}
& p(w_t|s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, \mathcal{F}_{t-1}) \\
&= p(w_t|s_{t-1} = 1, s_{t-2}, \dots, s_{t-k-1}, y_{t-1}, \dots, y_{t-k-1}) \\
&= p(w_t|w_{t-1} > \tau, u_{t-1}) \\
&= p(\alpha w_{t-1} + v_t|w_{t-1} > \tau, v_t).
\end{aligned} \tag{28}$$

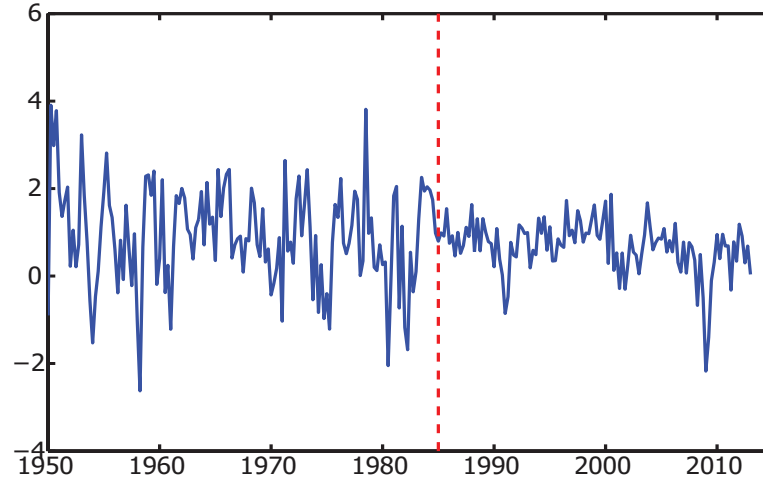
Since $p(w_{t-1}|w_{t-1} > \tau) = \frac{\sqrt{1 - \alpha^2} \phi(w_{t-1} \sqrt{1 - \alpha^2})}{1 - \Phi(\tau\sqrt{1 - \alpha^2})}$, due to (28), we can easily deduce

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

When $\alpha = 1$, the proofs are analogues to those of above cases except the fact that $w_t =_d \mathbb{N}(0, t)$ and therefore omitted.

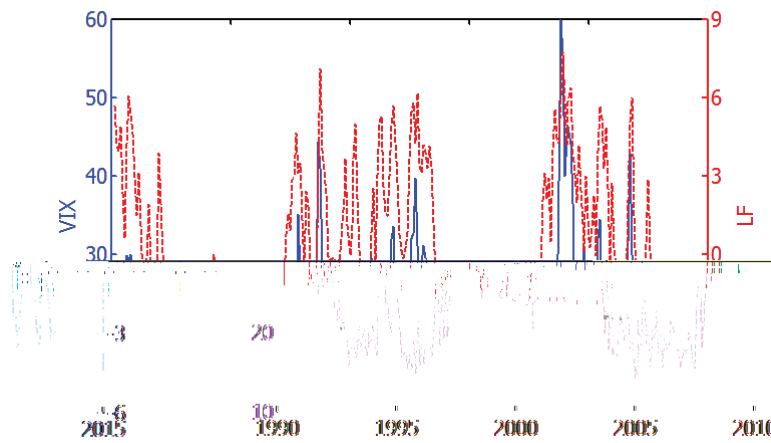
Appendix B: Additional Figures

Figure 12: US Real GDP Growth Rates



Notes: Figure 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.

Figure 13: Extracted Latent Factor and VIX



Notes: Figure 13 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.

Figure 14: High Volatility Probabilities

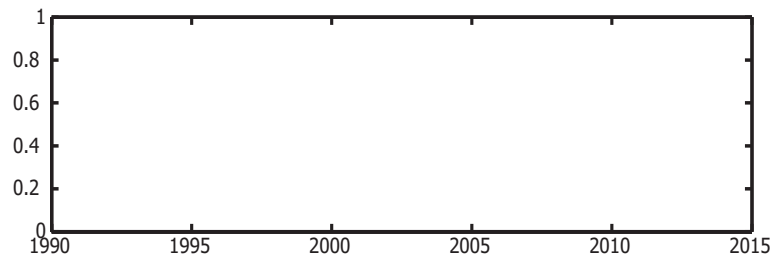


Figure 16: Recession Probabilities During 1952-1984

