

N-State Endogenous Markov-Switching Models*

Shih-Tang Hwu[†] Chang-Jin Kim[‡] Jeremy Piger[§]

December 2015

Abstract: We develop an N -regime Markov-switching regression model in which the latent state variable driving the regime switching is endogenous. The model admits a wide variety of patterns of correlation between the state variable and the regression disturbance term, while still maintaining computational feasibility. We provide an iterative filter that generates objects of interest, including the model likelihood function and estimated regime probabilities. The parameterization of the model also allows for a simple test of the null hypothesis of exogenous switching. Using simulation experiments we demonstrate that the maximum likelihood estimator performs well in finite samples, and that a likelihood ratio test of exogenous switching has good size and power properties. We provide results from two applications of the endogenous switching model: a three state model of U.S business cycle dynamics and a three-state volatility model of U.S. equity returns. In both cases we find statistically significant evidence in favor of endogenous switching.

Keywords: Regime-switching, business cycle asymmetry, nonlinear models, volatility feedback, equity premium

JEL Classifications: C13, C22, E32, G12

*We thank seminar participants at the Federal Reserve Bank of St. Louis, the University of Oregon, and the University of Washington for helpful comments.

[†]Department of Economics, University of Washington, (hwus@uw.edu)

[‡]Department of Economics, University of Washington, (changjin@u.washington.edu)

[§]Department of Economics, University of Oregon, (jpiger@uoregon.edu)

1 Introduction

Regression models with time-varying parameters have become a staple of the applied econometrician’s toolkit. A particularly prevalent version of these models is the Markov-switching regression of Goldfeld and Quandt (1973), in which parameters switch between some finite number of regimes, and this switching is governed by an unobserved Markov process. Hamilton (1989) makes an important advance by extending the Markov-switching framework to an autoregressive process, and providing an iterative filter that produces both the model likelihood function and filtered regime probabilities. Hamilton’s paper initiated a large number of applications of Markov-switching models, and these models are now a standard approach to describe the dynamics of many macroeconomic and financial time series. For surveys of this literature see Hamilton (2008) and Piger (2009).

Hamilton’s Markov-switching regression model assumes that the Markov state variable governing the timing of regime switches is strictly exogenous, and thus independent of the regression disturbance at all leads and lags. Diebold et al. (1994) and Filardo (1994) extend the Hamilton model to allow the transition probabilities governing the Markov process to be partly determined by strictly exogenous or predetermined information, which could include lagged values of the dependent variable. However, this time-varying transition probability (TVTP) formulation maintains the assumption that the state variable is independent of the contemporaneous value of the regression disturbance. The large literature applying Markov-switching models has almost exclusively focused on either the Hamilton (1989) fixed transition probability model or the TVTP extension, which we will collectively refer to as “exogenous switching” models.

Despite the popularity of this exogenous switching framework, it seems natural in many applications to think the state process is better modeled as endogenous. For example, a very common application of the Markov-switching regression is to models where the dependent variable is an aggregate measure of some macroeconomic or financial variable, and the state variable is meant to capture the business cycle regime (e.g. expansion and recession). It

seems reasonable that shocks to these aggregate quantities, such as real GDP, would contribute simultaneously to changes in the business cycle phase. More generally, both the state variable and the disturbance term to the dependent variable may be influenced simultaneously by a number of unmodeled elements. For example, in the Hamilton (1989) regime-switching autoregressive model of real GDP growth, both the state variable capturing the business cycle phase and the shock to real GDP are likely influenced by other factors, such as monetary and fiscal policy.

Motivated by such arguments, Kim et al. (2008) develop an “endogenous switching” regression model, in which the state variable and the regression disturbance term are determined simultaneously. Kang (2014) incorporates the Kim et al. (2008) model of endogenous switching inside of a more general state-space model. However, a significant drawback of this existing endogenous switching literature is that it is largely limited to the case of two regimes. This limits the potential application of the model considerably, as there is evidence for more than two regimes in many empirical implementations of the Markov-switching model. In models of real activity, Boldin (1996) finds evidence for a three regime switching model of business cycle dynamics for real GDP, while Hamilton (2005) does the same for the unemployment rate. For asset prices, Garcia and Perron (1996) and Guidolin and Timmermann (2005) provide evidence for a three-regime switching mean and volatility model of U.S. interest rates and equity returns respectively. In a Markov-switching VAR, Sims and Zha (2006) find the best fit using nine regimes, primarily capturing changes in conditional volatility.

In this paper, we develop an N -regime endogenous Markov-switching regression model. In the two regime case, the model collapses to that in Kim et al. (2008). For more than two regimes, the model admits a wide variety of patterns of correlation between the state variable and regression disturbance term. Despite this flexibility, the model maintains computational feasibility, and can be estimated via maximum likelihood using extensions to the filter in Hamilton (1989). The parameterization of the model also allows for a simple test of the null hypothesis of exogenous switching. Using simulation experiments, we demonstrate that the

maximum likelihood estimator performs well in finite samples, and that a likelihood ratio test of the null hypothesis of exogenous switching has good size and power properties.

We consider two applications of our N -regime endogenous switching model. In the first, we test for endogenous switching in a three regime switching mean model of U.S. real GDP growth. In the second we consider endogenous switching inside of a three-regime version of the Turner et al. (1989) volatility feedback model of U.S. equity returns. We find statistically significant evidence of endogenous switching in both of these models, as well as quantitatively large differences in parameter estimates resulting from allowing for endogenous switching.

2 An N-State Endogenous Markov-Switching Model

Consider the following Gaussian regime-switching model:

$$y_t = g(x_t, y_{t-1}, \dots, y_{t-p}, S_t, S_{t-1}, \dots, S_{t-p}) + \sigma_{S_t} \varepsilon_t, \quad (1)$$

$$\varepsilon_t \sim i.i.d. N(0, 1),$$

where $g(\cdot)$ is a conditional mean function, y_t is scalar, x_t is a $k \times 1$ vector of observed exogenous variables, and $S_t \in \{0, 1, \dots, N-1\}$ is an integer valued state variable indicating which of N different regimes is active at time t . Both y_t and x_t are assumed to be covariance-stationary variables. Examples of equation (1) include a regime-switching regression model:

$$y_t = x_t' \beta_{S_t} + \sigma_{S_t} \varepsilon_t, \quad (2)$$

as well as a regime-switching autoregression:

$$y_t = \mu_{S_t} + \phi_1 (y_{t-1} - \mu_{S_{t-1}}) + \phi_2 (y_{t-2} - \mu_{S_{t-2}}) + \dots + \phi_p (y_{t-p} - \mu_{S_{t-p}}) + \sigma_{S_t} \varepsilon_t. \quad (3)$$

For simplicity of exposition, we focus on the regime-switching regression model in equation

(2) throughout this paper. However, the algorithms presented below for estimation and filtering are easily extended to the more general case of equation (1).¹

In a Markov-switching model, the discrete regime indicator variable S_t follows an N -state Markov-process. Here we will allow the Markov-process to have time-varying transition probabilities as in Diebold et al. (1994) and Filardo (1994):

$$p_{ij,t} = \Pr(S_t = i | S_{t-1} = j, z_t) \quad (4)$$

In (4), the transition probability is influenced by the strictly exogenous or predetermined conditioning information z_t , and is thus time varying. Throughout, we assume that z_t is covariance stationary.

To model the dependence of the transition probability on the conditioning information z_t , it will be useful to alternatively describe S_t as the outcome of the values of $N - 1$ continuous latent variables, $S_{1,t}^*, S_{2,t}^*, \dots, S_{N-1,t}^*$, as follows:²

$$S_t = \left\{ \begin{array}{ll} 0, & S_{1,t}^* < 0 \\ 1, & S_{1,t}^* \geq 0, S_{2,t}^* < 0 \\ 2, & S_{1,t}^* \geq 0, S_{2,t}^* \geq 0, S_{3,t}^* < 0 \\ \vdots & \\ N-1, & S_{1,t}^* \geq 0, S_{2,t}^* \geq 0, S_{3,t}^* \geq 0, \dots, S_{N-1,t}^* \geq 0 \end{array} \right\} \quad (5)$$

¹The model likelihood for Markov-switching models will be invariant to an arbitrary relabeling of regimes. We assume throughout that the model has been appropriately normalized. Specific strategies for normalization will be discussed for the empirical analysis presented in Section 5.

²To describe S_t in terms of underlying continuous latent variables, we require a rule that divides the probability space among the N possible outcomes. There are many ways to accomplish this, with a leading example in the literature being the specification used for the multinomial probit model. As will be discussed in more detail in Section 3, we use the particular rule in (5) because it leads to a computationally efficient algorithm for computing multivariate CDFs.

Each of the $N - 1$ continuous latent variables follow symmetric processes:

$$\begin{aligned} S_{\tau,t}^* &= \gamma_{\tau,S_{t-1}} + z_t' \delta_{\tau} + \eta_{\tau,t} \\ \eta_{\tau,t} &\sim i.i.d.N(0, 1) \\ \tau &= 1, 2, \dots, N - 1 \end{aligned}$$

This provides enough structure to parameterize the transition probabilities for the Markov-process:

$$p_{ij,t} = \Pr(\eta_{1,t} \geq c_{1,j}, \eta_{2,t} \geq c_{2,j}, \dots, \eta_{i,t} \geq c_{i,j}, \eta_{i+1,t} < c_{i+1,j}), \quad (6)$$

where $c_{q,j} = -(\gamma_{q,j} + z_t' \delta_{q,j})$.

Typical applications of the Markov-switching model assume that the Markov-process driving S_t is either strictly exogenous, or in the case of time-varying transition probabilities, possibly dependent on lagged dependent variables included in z_t . In the formulation of the Markov-switching model given above, this is captured by assuming that the Gaussian model disturbance term ε_t is independent of each of the Gaussian disturbance terms for the latent variables $\eta_{\tau,t}$ at all leads and lags:

$$E(\varepsilon_t \eta_{\tau,t+\tau'}) = 0, \quad \forall t, \tau, \tau'$$

Here, we are alternatively interested in endogenous switching models, where S_t is potentially contemporaneously correlated with ε_t . To model this endogeneity, we assume that the joint probability density between ε_t and each $\eta_{\tau,t}$ is bivariate normal:

$$\begin{bmatrix} \eta_{\tau,t} \\ \varepsilon_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{\tau} \\ \rho_{\tau} & 1 \end{bmatrix} \right), \quad \tau = 1, 2, \dots, N - 1 \quad (7)$$

while maintaining the assumption that ε_t is independent of $\eta_{\tau,t}$ at leads and lags:

$$E(\varepsilon_t \eta_{\tau,t+\tau'}) = 0, \quad \forall t, \tau, \tau' \neq 0. \quad (8)$$

In exogenous switching models, the state variable S_t arrives exogenously, and then influences y_t by causing a switch in the data generating process for y_t . By contrast, in the endogenous switching model above, S_t may be determined endogenously with y_t . We do not take a stand on the direction of causality, and the model could be consistent with causality running from ε_t to S_t , from S_t to ε_t , or bi-directional causality. It is also worth noting that the exogenous switching model is obtained through the simple parameter restriction, $\rho_1 = \rho_2 = \dots \rho_{N-1} = 0$. This will be useful in developing tests for endogenous switching.

In order to calculate the transition probabilities in (6), we require an assumption about the joint distribution of the $\eta_{\tau,t}$, $\tau = 1, 2, \dots, N-1$. This assumption is needed to map the rule in (5) into transition probabilities, and is without loss of generality. Here we assume that conditional on ε_t , the $\eta_{\tau,t}$ are independent:

$$\text{cov}(\eta_{\tau,t}, \eta_{\tau',t} | \varepsilon_t) = 0, \quad \forall \tau \neq \tau'. \quad (9)$$

As will be described in Section (3), this conditional independence assumption will aid with efficient calculation of the multivariate CDFs in (6).

To understand how the exogenous and endogenous Markov-switching models differ, it is useful to consider the probability of transitioning between states, conditional on ε_t :

$$\tilde{p}_{ij,t} = \Pr(S_t = i | S_{t-1} = j, z_t, \varepsilon_t) \quad (10)$$

For the exogenous switching model, this conditional transition probability is equal to the unconditional transition probability, so that $\tilde{p}_{ij,t} = p_{ij,t}$. For the endogenous switching model the conditional and unconditional transition probabilities will not be equal, and the

realization of ε_t can signal markedly different probabilities of transitioning between regimes. As an example of this, Figure 1 plots unconditional and conditional transition probabilities against alternative realizations of $\varepsilon_t \in (-3, 3)$ for a particular parameterization of a three state ($N = 3$) endogenous switching model. In this example, the dependence on z_t has been eliminated, and the correlation parameters have been set to $\rho_1 = -0.5$ and $\rho_2 = 0.9$. The figure shows that the conditional probability of transitioning regimes can vary in extreme directions depending on the outcome of ε_t . For example, focusing on the diagonal entries, the probability of remaining in the $S_t = 0$ regime, $\tilde{p}_{00,t}$ increases from around 0.4 to 1.0 as ε_t moves from a large negative value (-2) to a large positive value (2), while $\tilde{p}_{11,t}$ moves in the opposite direction by an even larger amount. These responses do not have to be monotonic, as is shown by the probability $\tilde{p}_{22,t}$, which moves from near 0 when $\varepsilon_t = -2$ to near 0.8 when $\varepsilon_t = 0$, but then falls to near 0.4 when $\varepsilon_t = 2$. Alternative parameterizations for ρ_1 and ρ_2 give alternative patterns of $\tilde{p}_{ij,t}$, as is seen in Figure 2 which depicts the transition probabilities when $\rho_1 = 0.9$ and $\rho_2 = 0.9$. These figures also demonstrate that the conditional transition probability can differ markedly from the unconditional transition probability, which is depicted by the horizontal dashed lines in each figure. As will be shown in detail in the next section, the ratio of these two probabilities is an important quantity in distinguishing the likelihood function for the endogenous switching model from that for the exogenous switching model.

3 Likelihood Calculation, State Filtering and Tests for Endogenous Switching

In this section we describe how both the likelihood function and filtered and smoothed probabilities of the states can be calculated for the endogenous switching model. We will also describe how these calculations differ from those for the exogenous switching model. Finally, we discuss how tests of the null hypothesis of exogenous switching vs. the alternative

hypothesis of endogenous switching can be conducted.

Collect the model parameters into the vector θ , and let $\mathbf{Z}_t = \{z_t, z_{t-1}, \dots\}$ and $\Psi_t = \{y_t, y_{t-1}, \dots\}$ indicate the history of observed z_t and y_t through date t . As in Filardo (1994), the conditional likelihood value for y_t , $f(y_t|\Psi_{t-1}, \mathbf{Z}_t, \theta)$, $t = 1, \dots, T$, can be constructed recursively using an extension of the iterative formulas in Hamilton (1989) to the case of time-varying transition probabilities:³

$$f(y_t|\Psi_{t-1}, \mathbf{Z}_t, \theta) = \sum_{S_t=0}^{N-1} \sum_{S_{t-1}=0}^{N-1} f(y_t|S_t, S_{t-1}, \Psi_{t-1}, \mathbf{Z}_t, \theta) p_{ij,t} \Pr(S_{t-1}|\Psi_{t-1}, \mathbf{Z}_{t-1}, \theta) \quad (11)$$

$$\Pr(S_t = i|\Psi_t, \mathbf{Z}_t, \theta) \propto \sum_{S_{t-1}=0}^{N-1} f(y_t|S_t, S_{t-1}, \Psi_{t-1}, \mathbf{Z}_t, \theta) p_{ij,t} \Pr(S_{t-1}|\Psi_{t-1}, \mathbf{Z}_{t-1}, \theta) \quad (12)$$

These equations can be iterated recursively to obtain the log likelihood function $L(\theta) = \sum_{t=1}^T \log[f(y_t|\Psi_{t-1}, \mathbf{Z}_t, \theta)]$ and the filtered state estimates $\Pr(S_t = i|\Psi_t, \mathbf{Z}_t, \theta)$, $t = 1, \dots, T$. To initialize the recursion we require an initial filtered state probability, $\Pr(S_0 = i|\Psi_0, \mathbf{Z}_0, \theta)$, $i = 0, \dots, N-1$, calculation of which can be quite involved. Here we follow the usual practice, suggested by Hamilton (1989), of approximating this initial probability with an unconditional probability. In the case of time-varying transition probabilities, we use the unconditional state probability computed assuming z_t is always at its sample mean. Denote this probability as $\Pr(S_t = i|\bar{z})$, $i = 0, \dots, N-1$, where \bar{z} is the sample mean of z_t . Next, define $\bar{p}_{ij} = \Pr(S_t = i|S_{t-1} = j, \bar{z})$, and collect these in a matrix of transition probabilities as:

$$\bar{P} = \begin{bmatrix} \bar{p}_{00} & \bar{p}_{01} & \dots & \bar{p}_{0\ N-1} \\ \bar{p}_{10} & \bar{p}_{11} & \dots & \bar{p}_{1\ N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{p}_{N-1\ 0} & \bar{p}_{N-1\ 1} & \dots & \bar{p}_{N-1\ N-1} \end{bmatrix} \quad (13)$$

³For notational convenience, we suppress the dependence of probability density functions on the regressors, x_t , throughout this section. Equations (11) and (12) make use of the assumption, implicit in equation (2), that conditional on x_t and the state indicator S_t , the probability density function of y_t does not depend on z_t . This is without loss of generality, since x_t may include elements of z_t .

Finally, define:

$$A = \begin{bmatrix} I_N - \bar{P} \\ \iota'_N \end{bmatrix}$$

where I_N is the $N \times N$ identity matrix and ι_N is an $N \times 1$ vector of ones. The vector holding $Pr(S_t = i | \bar{z})$, $i = 0, \dots, N-1$ is then computed as the last column of the matrix $(A'A)^{-1} A'$.

The key element required to compute each step of the the recursion in (11) and (12) is $f(y_t | S_t, S_{t-1}, \Psi_{t-1}, \mathbf{Z}_t, \theta)$, and it is here that we see the distinction in the likelihood function between the exogenous and endogenous switching models. In the exogenous switching model, the state indicators $S_t = i$ and $S_{t-1} = j$ simply defines the mean and variance of a Gaussian distribution for y_t , such that:

$$f(y_t | S_t = i, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta) = \frac{1}{\sigma_i} \phi\left(\frac{y_t - x'_t \beta_i}{\sigma_i}\right)$$

where $\phi()$ indicates the standard normal probability density function. By contrast, when there is endogenous switching, the state variables $S_t = i$ and $S_{t-1} = j$ indicate not just the parameters of the relevant data generating process, but additionally provide information about which values of the random disturbance, ε_t , are most likely. In the case of endogenous switching:

$$f(y_t | S_t = i, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta) = \frac{\tilde{p}_{ij,t}}{p_{ij,t}} \left[\frac{1}{\sigma_i} \phi\left(\frac{y_t - x'_t \beta_i}{\sigma_i}\right) \right] \quad (14)$$

This equation, which is derived in the appendix, can be interpreted as follows. The term in brackets is the regime-dependent conditional density for y_t for the exogenous switching model. This density is then weighted by a ratio of probabilities of transitioning from regime j to regime i , where the probability in the numerator is conditional on the regime-specific value of ε_t and the probability in the denominator is not. The unconditional transition prob-

ability $p_{ij,t}$ can be interpreted as the average value of $\tilde{p}_{ij,t}$ with respect to the unconditional distribution of ε_t . In other words, $p_{ij,t}$ gives the average probability of transitioning from state j to state i with respect to ε_t . Thus, equation (14) says that if the value of ε_t signals an above average probability of transitioning from state j to state i , then the likelihood value for y_t conditional on $S_t = i$ and $S_{t-1} = j$ will be higher than would be calculated under the exogenous switching model. Returning to Figures 1 and 2, the ratio $\tilde{p}_{ij,t}/p_{ij,t}$ can be far from unity, meaning the likelihood function for the exogenous switching model may be substantially misspecified in the presence of endogenous switching. In general, estimation assuming exogenous switching will lead to biased parameter estimates as well as biased filtered state probabilities when endogenous switching is present.

Calculation of the transition probabilities $\tilde{p}_{ij,t}$ and $p_{ij,t}$ require calculation of multivariate Gaussian CDFs, for which analytical formulas do not exist. A number of accurate approximations are available for the general case, but these are computationally intensive, which will significantly slow down estimation procedures requiring repeated likelihood calculations. Rather than rely on these general procedures, we instead exploit the specific structure of the endogenous switching model to speed calculations. Consider first the conditional transition probability:

$$\tilde{p}_{ij,t} = \Pr(\eta_{1,t} \geq c_{1,j}, \eta_{2,t} \geq c_{2,j}, \dots, \eta_{i,t} \geq c_{i,j}, \eta_{i+1,t} < c_{i+1,j} | \varepsilon_t)$$

The conditional independence assumption in equation (9) transforms this multivariate CDF into a product of univariate Gaussian CDFs, each of which can be approximated to a high degree of accuracy at little computational expense. Specifically:

$$\begin{aligned} \tilde{p}_{ij,t} &= \Pr(\eta_{1,t} \geq c_{1,j} | \varepsilon_t) \Pr(\eta_{2,t} \geq c_{2,j} | \varepsilon_t) \dots \Pr(\eta_{i,t} \geq c_{i,j} | \varepsilon_t) \Pr(\eta_{i+1,t} < c_{i+1,j} | \varepsilon_t) \quad (15) \\ &= \prod_{q=1}^i \left[1 - \Phi \left(\frac{c_{q,j} - \rho_q \varepsilon_t}{\sqrt{1 - \rho_q^2}} \right) \right] \Phi \left(\frac{c_{i+1,j} - \rho_{i+1,t} \varepsilon_t}{\sqrt{1 - \rho_{i+1}^2}} \right) \end{aligned}$$

where $\Phi()$ indicates the univariate Gaussian CDF.⁴

Next, consider the unconditional transition probability:

$$p_{ij,t} = \Pr[\eta_{1,t} \geq c_{1,j}, \eta_{2,t} \geq c_{2,j}, \dots, \eta_{i,t} \geq c_{i,j}, \eta_{i+1,t} < c_{i+1,j}]$$

The dependence of $\eta_{i,t}$ on ε_t in the endogenous switching model induces non-zero covariance terms in the unconditional joint distribution of the $\eta_{i,t}$, meaning that $p_{ij,t}$ does not collapse to a product of univariate CDFs. However, an efficient Monte Carlo integration technique is available, which bypasses the need to compute multivariate CDFs. Specifically, consider G realization of ε_t from its $N(0, 1)$ unconditional distribution, each denoted $\varepsilon_t^{[g]}$, $g = 1, \dots, G$. Further, denote the conditional transition probability computed for one of these draws as $\tilde{p}_{ij,t}^{[g]}$. We then have:

$$p_{ij,t} = \lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G \tilde{p}_{ij,t}^{[g]} \quad (16)$$

We can then construct an accurate estimate of $p_{ij,t}$ using equation (16) with G set to some suitably large value. This Monte Carlo estimate is very fast to compute, even for very large values of G . The key to the efficiency is again the conditional independence assumption in equation (9), which allows us to compute $\tilde{p}_{ij,t}^{[g]}$, $g = 1, \dots, G$, very quickly as a product of univariate CDFs. These operations can also be easily vectorized, making looping unnecessary.⁵

The recursion provided by equations (11) and (12) can be used to construct the value of the likelihood function for any value of θ , which can then be numerically maximized with respect to θ to obtain the maximum likelihood estimates, $\hat{\theta}$. Given these estimates, the recursion can be run again to provide the filtered state probability evaluated at the

⁴To approximate the univariate Gaussian CDF we use the approach of Vazquez-Leal et al. (2012), which we found to yield speed improvements of 75% over Matlab's built-in "normcdf" command.

⁵The Monte Carlo estimate yields significant improvements in computation time over alternative, more generally applicable, numerical integration techniques. For example, for a specific test case with $N = 5$, we found that the estimate produced by (16) with $G = 100,000$ provides a high degree of accuracy and an 80% improvement in computation time over Matlab's "mvncdf" function.

maximum likelihood estimates, $\Pr(S_t = i | \Psi_t, \mathbf{Z}_t, \hat{\theta})$. In many applications we also require the so-called “smoothed” state probability $\Pr(S_t = i | \Psi_T, \mathbf{Z}_T, \theta)$, which provides inference on S_t conditional on all available sample information. To compute the smoothed probabilities, we can apply the recursive filter provided in Kim and Nelson (1999b), which remains valid for the N-state endogenous Markov-switching model described in Section 2. Beginning with the final filtered probability, $\Pr(S_T = j | \Psi_T, \mathbf{Z}_T, \theta)$, $j = 0, \dots, N - 1$, the following equation can be applied recursively, for $t = T - 1, \dots, 1$:

$$\Pr(S_t = i | \Psi_T, \mathbf{Z}_T, \theta) = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \Pr(S_{t-1} = j, S_t = i, S_{t+1} = k | \Psi_T, \mathbf{Z}_T, \theta) \quad (17)$$

where:

$$\begin{aligned} & \Pr(S_{t-1} = j, S_t = i, S_{t+1} = k | \Psi_T, \mathbf{Z}_T, \theta) \\ &= \frac{\Pr(S_t = i, S_{t+1} = k | \Psi_T, \mathbf{Z}_T, \theta) p_{ki,t} \Pr(S_t = i, S_{t-1} = j | \Psi_t, \mathbf{Z}_t, \theta)}{\Pr(S_t = j, S_{t+1} = k | \Psi_t, \mathbf{Z}_t, \theta)} \end{aligned} \quad (18)$$

For additional details of the derivation of equation (18), see Kim (1994) and Kim and Nelson (1999b).

To conclude this section, we describe how statistical hypothesis tests of the null hypothesis of exogenous switching can be conducted. Our N-state endogenous switching model collapses to a standard exogenous Markov-switching model in the case where:

$$\rho_1 = \rho_2 = \dots = \rho_{N-1} = 0, \quad (19)$$

Thus, the null hypothesis of exogenous switching can be tested by any suitable joint test of the $N - 1$ zero restrictions in 19. In the simulation studies presented in Section 4, we will consider the finite sample performance of both Wald and likelihood ratio tests of these restrictions.

4 Monte Carlo Evidence

In this section we describe results from a Monte Carlo simulation study designed to evaluate the finite sample performance of the maximum likelihood estimator (MLE) applied to data generated from an endogenous switching model. We also evaluate the size and power performance of hypothesis tests for endogenous switching. To focus on the results most germane to the addition of endogenous switching, we consider a simplified version of the general model presented in Section 2. In particular, we focus on the Gaussian Markov-switching mean and variance model:

$$\begin{aligned} y_t &= \mu_{S_t} + \sigma_{S_t} \varepsilon_t \\ \varepsilon_t &\sim i.i.d. N(0, 1) \end{aligned} \tag{20}$$

where $S_t \in \{0, 1, 2\}$ is a three-state Markov process that evolves with fixed transition probabilities $p_{ij} = \Pr(S_t = i | S_{t-1} = j)$.

We begin by studying the performance of the MLE applied to the incorrectly specified model that assumes exogenous switching. We can gain some initial insight into the bias that will result in the μ_i parameters by considering the state contingent expectation for the Markov-switching model:

$$E(y_t | S_t = i, S_{t-1} = j) = \mu_i + \sigma_i E(\varepsilon_t | S_t = i, S_{t-1} = j)$$

In the case of exogenous switching, the state indicator provides no information about ε_t and we have $E(y_t | S_t = i, S_{t-1} = j) = \mu_i$. However, with endogenous switching, this equality does not hold. Consider the case of $S_t = 0$:

$$\begin{aligned} E(y_t | S_t = 0, S_{t-1} = j) &= \mu_0 + \sigma_0 E(\varepsilon_t | \eta_{1,t} < -\gamma_{1,j}) \\ &= \mu_0 + \sigma_0 \left(-\rho_1 \frac{\phi(-\gamma_{1,j})}{\Phi(-\gamma_{1,jj})} \right) \end{aligned} \tag{21}$$

Here, μ_0 does not equal the state contingent expectation of y_t as would be the case with exogenous switching. The MLE assuming exogenous switching will then provide biased parameter estimates of μ_0 . The amount of this bias will depend on several factors, including the unconditional variance of ε_t , the extent of correlation between $\eta_{1,t}$ and ε_t , and the inverse Mills Ratio $\frac{\phi(-\gamma_{1,j})}{\Phi(-\gamma_{1,j})}$. This final term captures the extent to which the movement from $S_{t-1} = j$ to $S_t = 0$ is informative about $\eta_{1,t}$. For example, values of this ratio near zero correspond to $p_{0j} \approx 1$. In this case, a transition from $S_{t-1} = j$ to $S_t = 0$ provides little information about $\eta_{1,t}$, and thus little information about the value of ε_t .

To provide a more comprehensive look at the performance of the MLE that incorrectly assumes exogenous switching, we present results from a simulation experiment in Table 1. In all Monte Carlo simulations we set $\mu_{S_t} \in \{-1, 0, 1\}$ and $\sigma_{S_t} \in \{0.33, 0.67, 1.00\}$. The Markov process evolves according to the endogenous switching model outlined in Section 2 with $z_t = 0, \forall t$. Across alternative Monte Carlo experiments we vary the persistence of the transition probabilities for remaining in a regime from a “high persistence” case ($p_{00} = p_{11} = p_{22} = 0.9$) to a “low persistence case” ($p_{00} = p_{11} = p_{22} = 0.7$). We also vary the size of the correlation parameters from $\rho_1 = \rho_2 = 0.9$ to $\rho_1 = \rho_2 = 0.5$. Finally, we consider two sample sizes, $T = 300$ and $T = 500$. Performance is measured using the mean and root mean squared error (RMSE) of the estimates of each parameter across 1000 Monte Carlo simulations. The RMSE, reported in parentheses, is computed relative to the true value for each parameter.

The results in Table 1 demonstrate that the bias in the MLE that ignores endogenous switching can be severe. The bias in the μ_i parameters increases as the state persistence falls, with the amount of bias reaching as high as 67% of the true parameter value in the case of μ_2 . Estimation bias is also visible in the estimates of the regime-switching conditional variance term, with the bias in some cases above 30% of the true parameter value. The estimation bias is not a small sample phenomenon, with similar bias observed for $T = 300$ as for $T = 500$. The bias decreases as the correlation parameters, ρ_1 and ρ_2 , fall from 0.9 to 0.5.

However, despite this substantially lessened importance of endogenous switching, the MLE that ignores endogenous switching still generates very biased parameter estimates, with bias reaching as high as 49% of the true parameter value for μ_2 .

Table 2 shows results for the same variety of data generating processes, but with the MLE now applied to the correctly specified model. These results demonstrate that the MLE of the correctly specified model performs very well, with mean parameter estimates that are close to the true value, and RMSE statistics that are small. The performance of the correctly specified estimator seems largely unaffected by the extent of state persistence or the value of the correlation parameters. The sample size also does not have large effects on the mean estimates although, not surprisingly, the RMSE is higher when the sample size is $T = 300$.

Finally, we show simulation results to assess the finite sample performance of both Wald and likelihood ratio (LR) tests of the null hypothesis of exogenous switching, which is parameterized as a test of the joint restriction $\rho_1 = \rho_2 = 0$. We again consider two sample sizes, as well as a high and low state persistence case. To evaluate the size of the Wald and LR tests, we first consider the case where the true data generating process has $\rho_1 = \rho_2 = 0$. To evaluate the power of these tests we consider two cases, one in which the extent of endogenous switching is high ($\rho_1 = \rho_2 = 0.9$) and a second where endogenous switching is more moderate ($\rho_1 = \rho_2 = 0.5$). The size results are based on rejection rates of 5%-level tests using asymptotic critical values. The power results are based on rejection rates using size-adjusted 5% critical values.

Beginning with the size of the tests, the Wald test is significantly oversized, with true size over 20% in the $T = 300$ case and over 10% in the $T = 500$ case. In contrast, the LR test has close to correct size for both the $T = 300$ and $T = 500$ cases, with rejection rates between 3.6% and 6.8%. Turning to the power results, the LR test displays high rejection rates ranging from between 86% and 100%. The Wald test is less consistent, with rejection rates ranging from 25.9% to 100%.

Overall, the Monte Carlo results suggest that ignoring endogenous switching can lead to

substantial bias in the MLE when endogenous switching is in fact present. This bias persists into large sample sizes, and for both high and moderate values of the parameters controlling the extent of endogenous switching. The MLE that accounts for endogenous switching performed very well, yielding accurate parameter estimates and low variability of these estimates. Finally, the LR test for exogenous switching was effective, with approximately correct size and good power.

5 Applications in Macroeconomics and Finance

In this section, we consider two applications of the N-state endogenous Markov-switching model. In Section 5.1, we consider endogenous switching in a three regime model of U.S. business cycle dynamics. In Section 5.2 we extend the two regime endogenous-switching volatility feedback model in Kim et al. (2008) to allow for three volatility regimes.

5.1 U.S. Business Cycle Fluctuations

One empirical characteristic of the U.S. business cycle highlighted by Burns and Mitchell (1946) is asymmetry in the behavior of real output across business cycle phases. In his seminal paper, Hamilton (1989) captures asymmetry in the business cycle using a two-state Markov-switching autoregressive model of U.S. real GNP growth. His model identifies one phase as relatively brief periods of steep declines in output, and the other as relatively long periods of gradual output increases. Using quarterly data from 1952:Q2 to 1984:Q4, Hamilton (1989) shows that the estimated shifts between the two phases accord well with the National Bureau of Economic Research (NBER) chronology of U.S. business cycle peaks and troughs.

While Hamilton’s original model captures the short and steep nature of recessions relative to expansions, it does not incorporate an important feature of the business cycle that was prevalent over the sample period he considered: recessions were typically followed by high-

growth recovery phases that pushed output back toward its pre-recession level. This “bounce back” effect is evident in the post-recession real GDP growth rates shown in Table 4. In order to capture this high growth recovery phase, Sichel (1994) and Boldin (1996) extend Hamilton’s original model to a three state Markov-switching model.

Here we also use a three state Markov-switching model to capture recessions, expansions, and a post-recession recovery phase. In particular, we assume that the U.S. real GDP growth rate is described by the following three state Markov-switching mean model:

$$\begin{aligned}\Delta y_t &= \mu_{S_t} + \sigma \varepsilon_t \\ \varepsilon_t &\sim i.i.d.N(0, 1)\end{aligned}\tag{22}$$

where y_t is U.S. log real GDP and $S_t \in \{0, 1, 2\}$ is a Markov-switching state variable that evolves with fixed transition probabilities p_{ij} . Note that in this model U.S. real GDP growth follows a white noise process inside of each regime. This intra-regime lack of dynamics is consistent with the results of Kim et al. (2005) and Camacho and Perez-Quiros (2007), who find that traditional linear autoregressive dynamics in U.S. real GDP growth are largely absent once mean growth is allowed to follow a three-regime Markov-switching process.

We restrict the model in two ways. First, we restrict $\mu_0 > 0$, $\mu_1 < 0$, and $\mu_2 > 0$, which serves to identify $S_t = 1$ as the recession regime, and $S_t = 0$ and $S_t = 2$ as expansion regimes. Second, following Boldin (1996), we restrict the matrix of transition probabilities so that the states occur in the order $0 \rightarrow 1 \rightarrow 2$:

$$\bar{P} = \begin{bmatrix} p_{00} & 0 & 1 - p_{22} \\ 1 - p_{00} & p_{11} & 0 \\ 0 & 1 - p_{11} & p_{22} \end{bmatrix},$$

In combination with the restrictions on μ_{S_t} , this form of the transition matrix restricts the regimes to occur in the order: mature expansion \rightarrow recession \rightarrow post-recession expansion.

We will consider two versions of this model, one in which the Markov-switching is assumed to be exogenous, so that $\rho_1 = \rho_2 = 0$, and one that allows for endogenous switching.

We estimate this model using data on quarterly U.S. real GDP growth from 1952:Q1 - 2014:Q2. Over this sample period, there are two prominent types of structural change in the U.S. business cycle that are empirically relevant. The first is the well-documented reduction in real GDP growth volatility in the early 1980s known as the “Great Moderation” (Kim and Nelson (1999a), McConnell and Perez-Quiros (2000)). To capture this reduction in volatility, we include a one time change in the conditional volatility parameter, σ , in 1984:Q1, the date identified by Kim and Nelson (1999a) as the beginning of the Great Moderation. The second, as identified in Kim and Murray (2002) and Kim et al. (2005), is the lack of a high growth recovery phase following the three most recent NBER recessions. To capture this change in post-recession growth rates, we include a one-time break in μ_2 . Finally, to allow for the possibility that the nature of endogenous switching changed along with the nature of the post-recession regime, we allow for breaks in ρ_1 and ρ_2 . These breaks in μ_2 , ρ_1 and ρ_2 are also dated to 1984:Q1, although results are insensitive to alternative break dates between 1984:Q1 and the beginning of the 1990-1991 recession. All other model parameters are assumed to be constant over the entire sample period.

The second and third columns of Table 5 shows the maximum likelihood estimation results when we assume exogenous switching. The estimates show a prominent high growth recovery phase before 1984:Q1 ($\mu_{2,1} \gg \mu_0$). The estimates also show that this high growth recovery phase has disappeared in recent recessions, and indeed been replaced with a *low-growth* post-recession phase ($\mu_{2,2} < \mu_0$). The conditional volatility parameter, σ , falls by nearly 50% after 1984, consistent with the large literature on the Great Moderation.

The maximum likelihood estimates assuming endogenous switching are shown in the fourth and fifth columns of Table 5. A likelihood ratio test rejects the null hypothesis of exogenous switching at the 5% level (p-value = 0.049). Looking at the individual estimates and standard errors, this rejection is arising primarily from $\rho_{1,2}$, the value of ρ_1 after 1984.

The estimated value of $\rho_{1,2}$ is significantly positive, corresponding to a strong positive correlation between the regression disturbance term, ε_t , and η_{1t} , the disturbance to the latent state variable $S_{1,t}^*$. This implies that large values of $\eta_{1,t}$, which generate switches from expansion to recession, tend to be accompanied by large positive shocks to ε_t . This correlation creates a rounded path of real GDP around business cycle peaks, an empirical result that cannot be systematically captured by the exogenous switching version of the model.

There is also evidence of bias in the parameter estimates of the exogenous switching model. Each of the estimates of the regime-dependent growth rates are substantially different when accounting for endogenous switching. Also, the continuation probability for the post-recession phase, p_{22} , is substantially lower when accounting for endogenous switching, meaning the length of these phases are overstated by the results from exogenous switching models. Finally, results of the Ljung-Box test, shown in the bottom panel of the table, show that accounting for endogenous switching eliminates autocorrelation in the disturbance term that is present in the exogenous switching model.

Figure 3 displays the smoothed state probabilities for both the exogenous and endogenous switching models, and shows the distortion in estimated state probabilities that can occur from ignoring endogenous switching. From panel (a), we see that the smoothed probability that the economy is in a mature expansion ($S_t = 0$) is often lower for the exogenous switching model than the endogenous switching model, while panel (c) shows that the opposite is true for the smoothed probability of the post-recession recovery phase ($S_t = 2$). Put differently, the endogenous switching model suggests a quicker transition from the post-recession recovery phase to the mature expansion phase than does the exogenous switching model.

5.2 Volatility Regimes in U.S. Equity Returns

An empirical regularity of U.S. equity returns is that low returns are contemporaneously associated with high volatility. This is a counterintuitive result, as classical portfolio theory implies the equity risk premium should respond positively to the expectation of future

volatility. One explanation for this observation is that while investors do require an increase in expected return for expected future volatility, they are often surprised by news about realized volatility. This “volatility feedback” creates a reduction in prices in the period in which the increase in volatility is realized. The volatility feedback effect has been investigated extensively in the literature by French et al. (1987), Turner et al. (1989), Campbell and Hentschel (1992), Bekaert and Wu (2000) and Kim et al. (2004).

Turner et al. (1989) (TSN hereafter) model the volatility feedback effect with a two state Markov-switching model:

$$r_t = \theta_1 E(\sigma_{S_t}^2 | I_{t-1}) + \theta_2 [E(\sigma_{S_t}^2 | I_t^*) - E(\sigma_{S_t}^2 | I_{t-1})] + \sigma_{S_t} \varepsilon_t$$

$$\varepsilon_t \sim i.i.d.N(0, 1)$$

where r_t is a measure of excess equity returns, $I_t = \{r_t, r_{t-1}, \dots\}$, and I_t^* is an information set that includes I_{t-1} and the information investors observe during period t . $S_t \in \{0, 1\}$ is a discrete variable that follows a two state Markov process with fixed transition probabilities p_{ij} . To normalize the model, TSN restrict $\sigma_1 > \sigma_0$, so that state 1 is the higher volatility regime.

One estimation difficulty with the above model is that there exists a discrepancy between the investors’ and the econometrician’s information set. In particular, while I_{t-1} may be summarized by returns up to period $t - 1$, the information set I_t^* includes information that is not summarized in the econometrician’s data set on observed returns. To handle this estimation difficulty, TSN use actual volatility, $\sigma_{S_t}^2$ to approximate $E(\sigma_{S_t}^2 | I_t^*)$. That is, they estimate,

$$r_t = \theta_1 E(\sigma_{S_t}^2 | I_{t-1}) + \theta_2 [\sigma_{S_t}^2 - E(\sigma_{S_t}^2 | I_{t-1})] + \sigma_{S_t} u_t \tag{23}$$

$$u_t = \varepsilon_t + \theta_2 [E(\sigma_{S_t}^2 | I_t^*) - \sigma_{S_t}^2]$$

Kim, Piger, and Startz (2008) (KPS, hereafter) point out that this approximation in-

introduces classical measurement error into the state variable in the estimated equation, thus rendering it endogenous. KPS propose a two-state endogenous Markov-switching model to deal with this endogeneity problem. Again, this two-state model of endogenous switching is identical to the N -state endogenous switching model proposed in Section 2 when $N = 2$. However, there is substantial evidence for more than two volatility regimes in U.S. equity returns (Guidolin and Timmermann (2005)). Here, we extend the TSN and KPS exogenous and endogenous switching volatility feedback models to allow for three volatility regimes. Specifically, we extend the volatility feedback model in equation (23) to allow for three regimes, $S_t \in \{0, 1, 2\}$, with fixed transition probabilities p_{ij} . For normalization we assume $\sigma_2 > \sigma_1 > \sigma_0$, so that state 2 is the highest volatility regime.

To estimate the three state volatility feedback model, we measure excess equity returns using monthly returns for a value-weighted portfolio of all NYSE-listed stocks in excess of the one-month Treasury Bill rate. The sample period extends from January 1952 to December 2013. The second and third columns of Table 6 show the estimation results when we assume exogenous switching. The estimates are consistent with a positive relationship between the risk premium and expected future volatility ($\theta_1 > 0$) and a substantial volatility feedback effect ($\theta_2 \ll 0$). The estimates also suggest a dominant volatility feedback effect, as θ_1 is small in absolute value relative to θ_2 .

The fourth and fifth columns of Table 6 show the estimation results when we allow for endogenous switching. A likelihood ratio test rejects the null hypothesis of exogenous switching at the 10% level (p-value = 0.081). The results from the endogenous switching model show a smaller volatility feedback effect (smaller θ_2) and a much more persistent highest volatility regime (larger p_{22}) than the results from the exogenous switching model. The estimated correlation parameters have different signs, with $\rho_1 < 0$ and $\rho_2 > 0$. Given the expression for u_t in equation (23), the negative value of ρ_1 implies that investors set their expectations of volatility higher than the true volatility when the volatility regime is switching to state 1. In contrast, a positive ρ_2 implies that investors set their expectations

of volatility lower than the true volatility when the volatility regime is switching to state 2. Thus our model suggests that investors adjust their expectations of volatility asymmetrically with regard to the volatility regime.

Figure 4 shows the risk premium implied by three different volatility feedback models, the exogenous switching model with three states (red dashed line), the endogenous switching model with three states (blue solid line), and the endogenous switching model with two states (green dotted line). The three state endogenous switching model produces a risk premium that is more variable than the other models across volatility states. In particular, the risk premium from the three state endogenous switching model rises above the risk premium from the other models during the highest volatility state, which from Figure 5 can be seen to be highly correlated with NBER recessions. However, during the other volatility states, the risk premium from the three state endogenous switching model is generally below that from the other models. On average, our model suggests a 6% risk premium, which is below the 9% estimated by Kim et al. (2004) using the volatility feedback model assuming exogenous switching over the period 1952 to 1999. However, it is higher than Fama and French (2002), who estimate an average risk premium of 2.5% using the average dividend yield plus the average dividend growth rate for the S&P 500 index over the period 1951 to 2000.

6 Conclusion

We have proposed a novel N -state Markov-switching regression model in which the state indicator variable is correlated with the regression disturbance term. The model admits a wide variety of patterns for this correlation, while maintaining computational feasibility. Parameter estimates can be obtained via maximum likelihood using extensions to the filter in Hamilton (1989). The parameterization of the model also allows for a simple test of the null hypothesis of exogenous switching. In simulation experiments, the maximum likelihood estimator performed well, and a likelihood ratio test of the null hypothesis of exogenous

switching had good size and power properties. We considered two applications of the N -regime endogenous switching model, one to an empirical model of U.S. business cycles, and the other to a switching volatility model of U.S. equity returns. We find statistically significant evidence of endogenous switching in both of these models, as well as quantitatively large differences in parameter estimates resulting from allowing for endogenous switching.

Appendix: Derivation of $f(y_t|S_t, S_{t-1}, \Psi_{t-1}, \mathbf{Z}_t, \theta)$

The iterative filter presented in Section 3 requires calculation of the regime-dependent density $f(y_t|S_t, S_{t-1}, \Psi_{t-1}, \mathbf{Z}_t, \theta)$, where y_t represents the random variable described by the data generating process described in equation (2) along with the endogenous regime-switching process described in Section 2. We have again suppressed the conditioning of this density on the covariates x_t . This appendix derives this regime-dependent density.

Let y_t^* denote a realization of y_t for which we wish to compute $f(y_t^*|S_t = i, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta)$. Applying Bayes Rule yields:

$$f(y_t|S_t = i, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta) = \frac{f(y_t, S_t = i|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta)}{\Pr(S_t = i|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta)} \quad (\text{A.1})$$

The denominator of equation (A.1) is the time-varying transition probability, $p_{ij,t}$. Consider the following CDF of the numerator of (A.1):

$$\begin{aligned} & \Pr(y_t < y_t^*, S_t = i|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) \\ &= \int_{-\infty}^{y_t^*} f(y_t, S_t = i|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) dy_t \\ &= \int_{-\infty}^{\frac{y_t^* - x_t' \beta_i}{\sigma_i}} f(\varepsilon_t, S_t = i|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) d\varepsilon_t \\ &= \int_{-\infty}^{\frac{y_t^* - x_t' \beta_i}{\sigma_i}} \Pr(S_t = i|\varepsilon_t, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) f(\varepsilon_t|S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) d\varepsilon_t \\ &= \int_{-\infty}^{\frac{y_t^* - x_t' \beta_i}{\sigma_i}} \Pr(S_t = i|\varepsilon_t, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t) f(\varepsilon_t) d\varepsilon_t \end{aligned}$$

where the validity of moving to the last line in this derivation is ensured by the independence of ε_t over time, the exogeneity of \mathbf{Z}_t , and the independence of ε_t and S_{t-1} (see equation (8).)

Finally, differentiating this CDF with respect to y_t^* yields:

$$\begin{aligned} & f(y_t^*, S_t = i | S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, \theta) \\ &= \Pr\left(S_t = i \mid \left(\frac{y_t^* - x_t' \beta_i}{\sigma_i}\right), S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t\right) f\left(\frac{y_t^* - x_t' \beta_i}{\sigma_i}\right) \end{aligned} \quad (\text{A.2})$$

where $(y_t^* - x_t' \beta_i) / \sigma_i$ is a realization of the random variable ε_t . The first term in (A.2) is the conditional transition probability, $\tilde{p}_{ij,t}$. Given the marginal Gaussian distribution for ε_t , the second term in equation (A.2) is:

$$f\left(\frac{y_t^* - x_t' \beta_i}{\sigma_i}\right) = \frac{1}{\sigma_i} \phi\left(\frac{y_t^* - x_t' \beta_i}{\sigma_i}\right)$$

Combining the above results, we have:

$$f(y_t^* | S_t = i, S_{t-1} = j, \Psi_{t-1}, \mathbf{Z}_t, x_t) = \frac{\tilde{p}_{ij,t}}{p_{ij,t}} \left[\frac{1}{\sigma_i} \phi\left(\frac{y_t^* - x_t' \beta_i}{\sigma_i}\right) \right]$$

which is equation (14) evaluated at y_t^* .

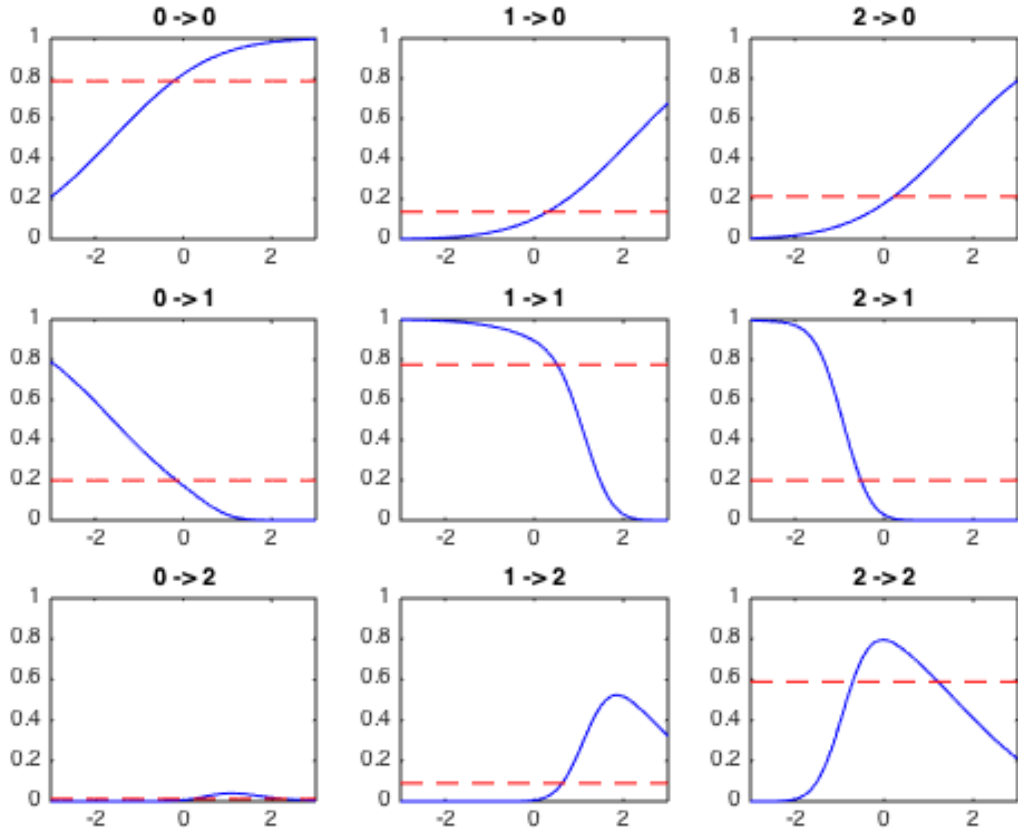
References

- Bekaert, G. and G. Wu (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies* 13(1), 1–42.
- Boldin, M. D. (1996). A check on the robustness of hamilton’s markov switching model approach to the economic analysis of the business cycle. *Studies in Nonlinear Dynamics and Econometrics* 1(1), 35–46.
- Burns, A. F. and W. C. Mitchell (1946). *Measuring business cycles*. New York: National Bureau of Economic Research. ID: 169122.
- Camacho, M. and G. Perez-Quiros (2007). Jump and rest effects of u.s. business cycles. *Studies in Nonlinear Dynamics and Econometrics* 11(4).
- Campbell, J. Y. and L. Hentschel (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31(3), 281–318.
- Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. In C. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, Advanced Texts and Econometrics, Oxford and New York, pp. 283–302. Oxford University Press.
- Fama, E. F. and K. R. French (2002). The equity premium. *Journal of Finance* 57(2), 637–659.
- Filardo, A. J. (1994). Business cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12(3), 299–308.
- French, K. R., W. G. Schwert, and R. F. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19(1), 3–29.
- Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* 78(1), 111–125.

- Goldfeld, S. M. and R. E. Quandt (1973). A markov model for switching regressions. *Journal of Econometrics* 1(1), 3–16.
- Guidolin, M. and A. Timmermann (2005). Economic implications of bull and bear regimes in uk stock and bond returns. *Economic Journal* 115(500), 111–143.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2), 357–384.
- Hamilton, J. D. (2005). What’s real about the business cycle? *Federal Reserve Bank of St. Louis Review* 87(4), 435–452.
- Hamilton, J. D. (2008). Regime switching models. In S. N. Durlauf and L. E. Blume (Eds.), *New Palgrave Dictionary of Economics, 2nd Edition*. Palgrave MacMillan.
- Kang, K. H. (2014). Estimation of state-space models with endogenous markov regime-switching parameters. *Econometrics Journal* 17(1), 56–82.
- Kim, C.-J. (1994). Dynamic linear models with markov switching. *Journal of Econometrics* 60(1-2), 1–22.
- Kim, C.-J., J. Morley, and J. Piger (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics* 20(2), 291–309.
- Kim, C.-J., J. C. Morley, and C. R. Nelson (2004). Is there a positive relationship between stock market volatility and the equity premium? *Journal of Money, Credit and Banking* 36(3), 339–360.
- Kim, C.-J. and C. J. Murray (2002). Permanent and transitory components of recessions. *Empirical Economics* 27(2), 163–183.
- Kim, C.-J. and C. R. Nelson (1999a). Has the u.s. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics* 81(4), 608–616.

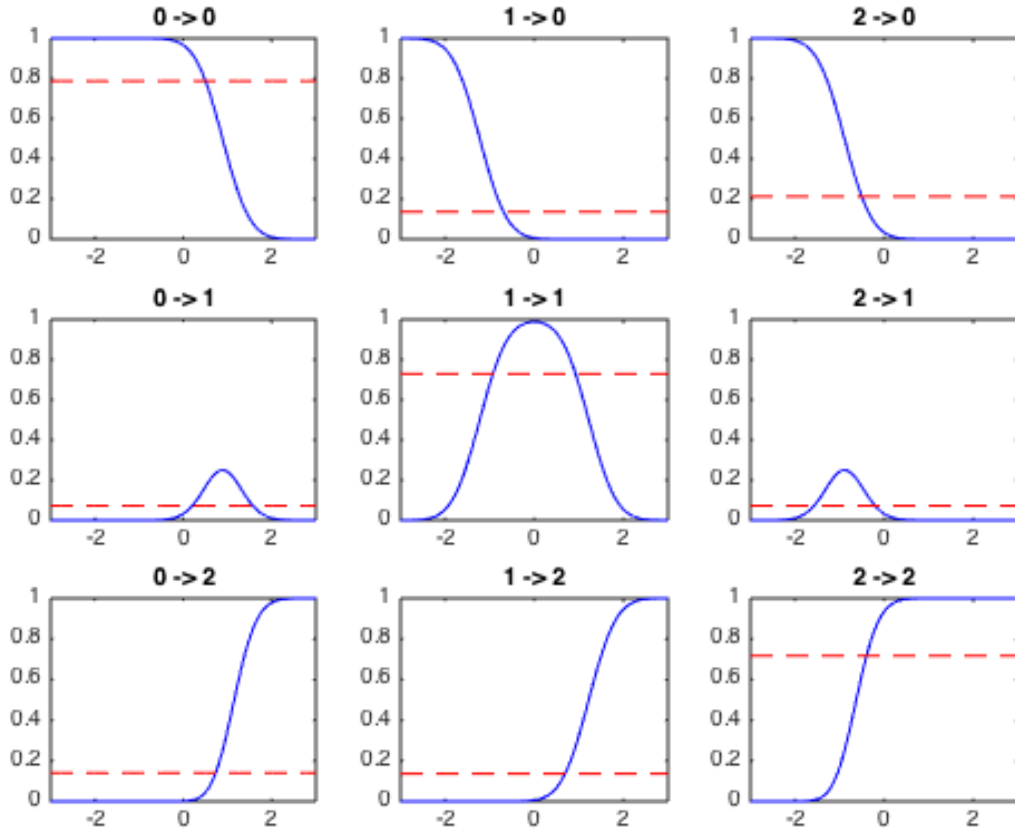
- Kim, C.-J. and C. R. Nelson (1999b). *State-Space Models with Regime Switching*. Cambridge, MA: The MIT Press.
- Kim, C.-J., J. Piger, and R. Startz (2008). Estimation of markov regime-switching regressions with endogenous switching. *Journal of Econometrics* 143(2), 263–273.
- McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980’s? *American Economic Review* 90(5), 1464–1476.
- Piger, J. (2009). Econometrics: Models of regime changes. In B. Mizrach (Ed.), *Encyclopedia of Complexity and System Science*, New York. Springer.
- Sichel, D. E. (1994). Inventories and the three phases of the business cycle. *Journal of Business and Economic Statistics* 12(3), 269–277.
- Sims, C. A. and T. Zha (2006). Were there regime switches in u.s. monetary policy? *American Economic Review* 96(1), 54–81.
- Turner, C. M., R. Startz, and C. R. Nelson (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* 25(1), 3–22.
- Vazquez-Leal, H., R. Castaneda-Sheissa, U. Filobello-Nino, A. Sarmiento-Reyes, and J. S. Orea (2012). High accurate simple approximation of normal distribution integral. *Mathematical Problems in Engineering* 2012(2012).

Figure 1
 $Pr(S_t = i | S_{t-1} = j)$ vs. $Pr(S_t = i | S_{t-1} = j, \varepsilon_t)$
 $\rho_1 = -0.5, \rho_2 = 0.9$



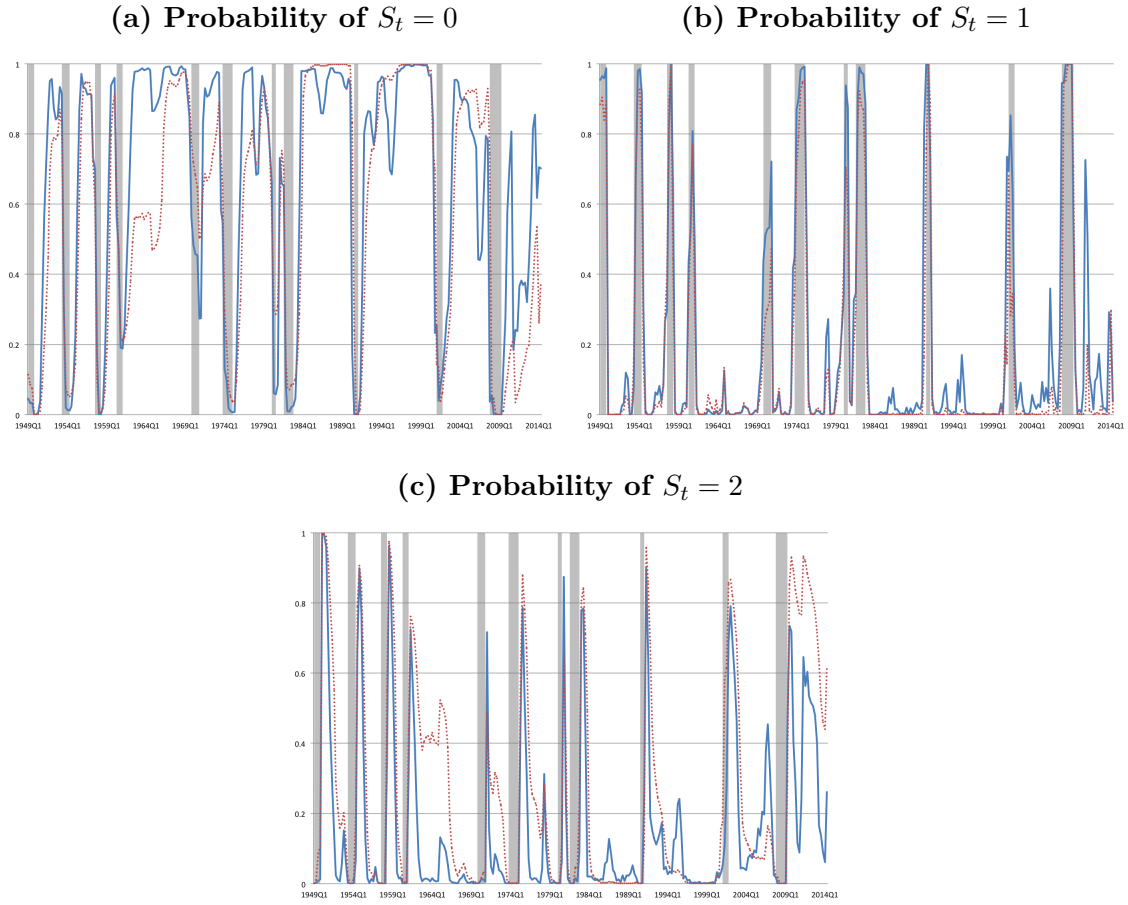
Notes: These graphs show the unconditional transition probability, $Pr(S_t = i | S_{t-1} = j)$ (horizontal dashed line), as well as the transition probability conditional on the continuous disturbance term in equation (2), $Pr(S_t = i | S_{t-1} = j, \varepsilon_t)$ (solid line). In all panels, $j \rightarrow i$ indicates transitions from state j to state i , and the x-axis measures alternative values of ε_t .

Figure 2
 $Pr(S_t = i | S_{t-1} = j)$ vs. $Pr(S_t = i | S_{t-1} = j, \varepsilon_t)$
 $\rho_1 = 0.9, \rho_2 = 0.9$



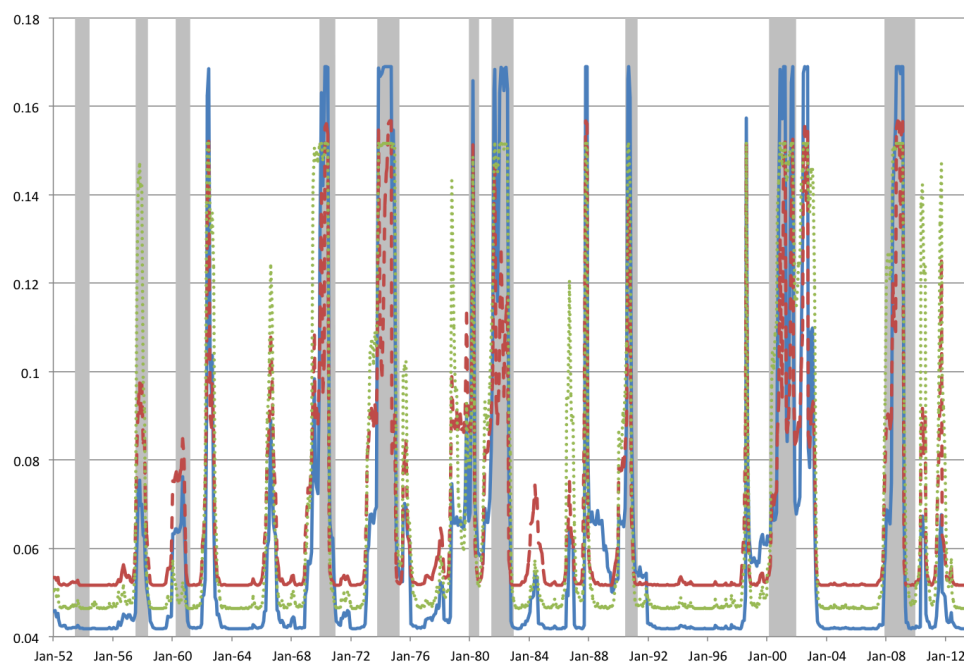
Notes: These graphs show the unconditional transition probability, $Pr(S_t = i | S_{t-1} = j)$ (horizontal dashed line), as well as the transition probability conditional on the continuous disturbance term in equation (2), $Pr(S_t = i | S_{t-1} = j, \varepsilon_t)$ (solid line). In all panels, $j \rightarrow i$ indicates transitions from state j to state i , and the x-axis measures alternative values of ε_t .

Figure 3
Smoothed State Probabilities for Three Regime Model of Real GDP Growth



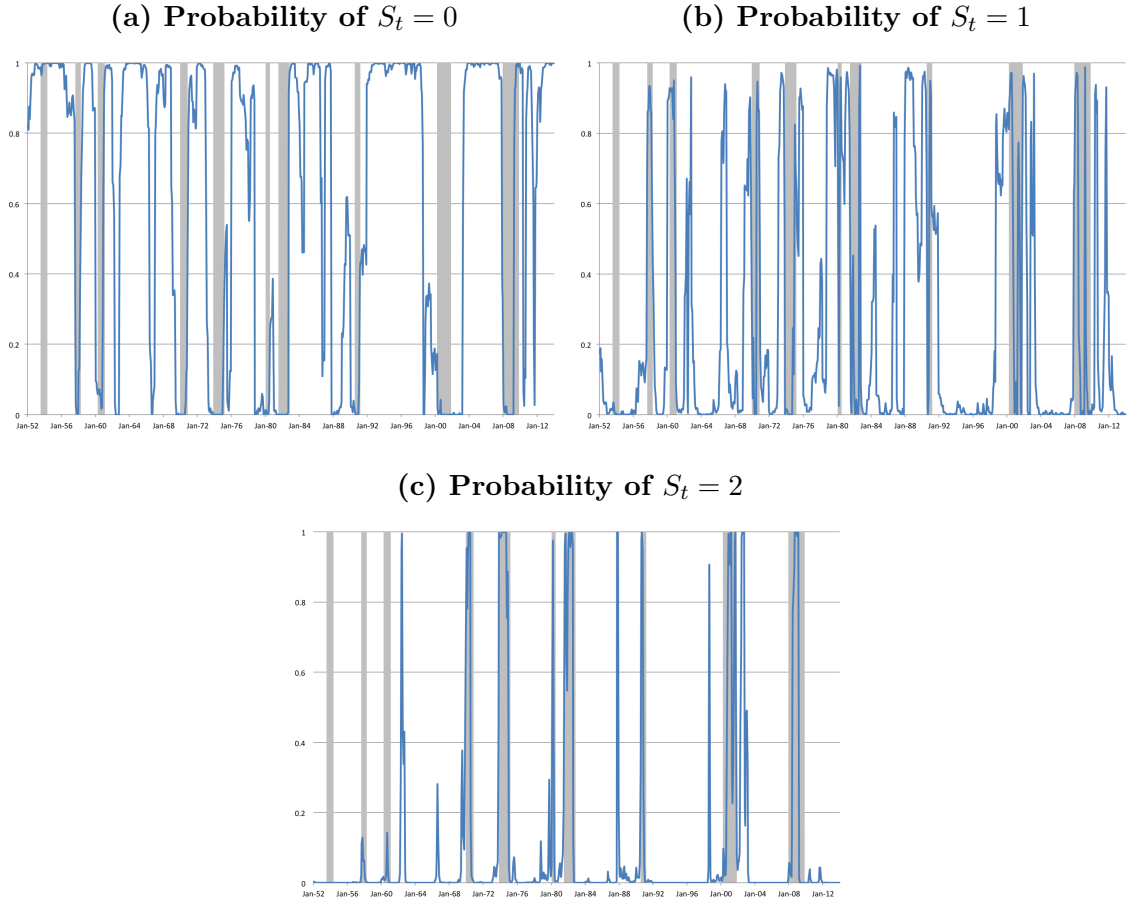
Notes: Smoothed probability of mature expansion phase ($S_t = 0$), recession phase ($S_t = 1$), and post-recession recovery phase ($S_t = 2$). Dotted lines denote the regime probability estimated by the exogenous switching model, and solid line represents the regime probability estimated by the endogenous switching model. NBER recessions are shaded.

Figure 4
Risk Premium from Alternative Volatility Feedback Models



Notes: Risk premium implied by different Markov-switching volatility feedback models. The red dashed line reports the risk premium produced by the exogenous switching model with three states, the green dotted line reports the risk premium produced by the endogenous switching model with two states, and the blue solid line reports the risk premium produced by the endogenous switching model with three states. NBER recessions are shaded.

Figure 5
Smoothed State Probabilities from Three Regime
Volatility Feedback Model with Endogenous Switching



Notes: Smoothed probability of low volatility phase ($S_t = 0$), medium volatility phase ($S_t = 1$), and high volatility phase ($S_t = 2$). NBER recessions are shaded.

Table 1
Monte Carlo Simulation Results
Performance of Misspecified Maximum Likelihood Estimator

	$\rho_1 = \rho_2 = 0.9$					
	$\mu_0 = -1$	$\mu_1 = 0$	$\mu_2 = 1$	$\sigma_0 = 0.33$	$\sigma_1 = 0.67$	$\sigma_2 = 1$
$T = 300$						
<i>High Persistence</i>	-1.10 (0.10)	-0.01 (0.22)	1.34 (0.36)	0.30 (0.03)	0.57 (0.12)	1.15 (0.16)
<i>Low Persistence</i>	-1.16 (0.29)	-0.05 (0.25)	1.67 (0.69)	0.28 (0.06)	0.43 (0.24)	1.02 (0.10)
$T = 500$						
<i>High Persistence</i>	-1.10 (0.10)	-0.02 (0.06)	1.25 (0.12)	0.36 (0.04)	0.60 (0.11)	1.03 (0.06)
<i>Low Persistence</i>	-1.24 (0.24)	0.01 (0.06)	1.62 (0.62)	0.33 (0.08)	0.45 (0.25)	0.89 (0.12)
	$\rho_1 = \rho_2 = 0.5$					
	$\mu_0 = -1$	$\mu_1 = 0$	$\mu_2 = 1$	$\sigma_0 = 0.33$	$\sigma_1 = 0.67$	$\sigma_2 = 1$
$T = 300$						
<i>High Persistence</i>	-1.06 (0.09)	-0.01 (0.12)	1.17 (0.25)	0.32 (0.03)	0.62 (0.09)	1.01 (0.11)
<i>Low Persistence</i>	-1.11 (0.17)	-0.02 (0.22)	1.42 (0.52)	0.31 (0.04)	0.57 (0.17)	0.94 (0.16)
$T = 500$						
<i>High Persistence</i>	-1.08 (0.09)	-0.02 (0.07)	1.23 (0.26)	0.37 (0.03)	0.63 (0.09)	0.96 (0.06)
<i>Low Persistence</i>	-1.18 (0.19)	-0.01 (0.09)	1.49 (0.51)	0.34 (0.06)	0.51 (0.20)	0.92 (0.11)

Notes: This table contains summary results from 1000 Monte Carlo simulations when the true data generating process is given by $y_t = \mu_{S_t} + \sigma_{S_t}\epsilon_t$ and S_t evolves according to the endogenous switching model detailed in Section 2 with N=3 states. “High Persistence” indicates high state persistence, with transition probabilities $p_{00} = p_{11} = p_{22} = 0.9$, while “Low Persistence” indicates transition probabilities $p_{00} = p_{11} = p_{22} = 0.7$. Each cell contains the mean of the 1000 maximum likelihood point estimates for the parameter listed in the column heading, as well as the root mean squared error of the 1000 point estimates from that parameter’s true value (in parentheses). The maximum likelihood estimator is applied to the incorrectly specified model that assumes the state process is exogenous.

Table 2
Monte Carlo Simulation Results
Performance of Correctly Specified Maximum Likelihood Estimator

	$\rho_1 = \rho_2 = 0.9$					
	$\mu_0 = -1$	$\mu_1 = 0$	$\mu_2 = 1$	$\sigma_0 = 0.33$	$\sigma_1 = 0.67$	$\sigma_2 = 1$
$T = 300$						
<i>High Persistence</i>	-1.00 (0.02)	0.01 (0.06)	1.04 (0.09)	0.32 (0.02)	0.68 (0.05)	1.09 (0.10)
<i>Low Persistence</i>	-0.98 (0.03)	-0.03 (0.05)	1.00 (0.08)	0.33 (0.02)	0.70 (0.06)	1.09 (0.11)
$T = 500$						
<i>High Persistence</i>	-0.99 (0.03)	0.00 (0.05)	0.99 (0.13)	0.38 (0.01)	0.67 (0.03)	0.98 (0.05)
<i>Low Persistence</i>	-1.00 (0.02)	0.00 (0.04)	1.00 (0.10)	0.37 (0.02)	0.69 (0.04)	0.98 (0.06)
	$\rho_1 = \rho_2 = 0.5$					
	$\mu_0 = -1$	$\mu_1 = 0$	$\mu_2 = 1$	$\sigma_0 = 0.33$	$\sigma_1 = 0.67$	$\sigma_2 = 1$
$T = 300$						
<i>High Persistence</i>	-1.00 (0.05)	-0.01 (0.12)	1.01 (0.18)	0.33 (0.03)	0.66 (0.08)	0.99 (0.09)
<i>Low Persistence</i>	-1.00 (0.08)	-0.03 (0.24)	1.10 (0.37)	0.33 (0.04)	0.67 (0.14)	1.01 (0.15)
$T = 500$						
<i>High Persistence</i>	-1.00 (0.03)	0.00 (0.06)	1.01 (0.12)	0.38 (0.03)	0.69 (0.05)	0.98 (0.06)
<i>Low Persistence</i>	-1.00 (0.04)	0.00 (0.08)	1.01 (0.14)	0.38 (0.04)	0.69 (0.07)	0.99 (0.08)

Notes: This table contains summary results from 1000 Monte Carlo simulations when the true data generating process is given by $y_t = \mu_{S_t} + \sigma_{S_t}\epsilon_t$ and S_t evolves according to the endogenous switching model detailed in Section 2 with N=3 states. “High Persistence” indicates high state persistence, with transition probabilities $p_{00} = p_{11} = p_{22} = 0.9$, while “Low Persistence” indicates transition probabilities $p_{00} = p_{11} = p_{22} = 0.7$. Each cell contains the mean of the 1000 maximum likelihood point estimates for the parameter listed in the column heading, as well as the root mean squared error of the 1000 point estimates from that parameter’s true value (in parentheses). The maximum likelihood estimator is applied to the correctly specified model that assumes the state process is endogenous.

Table 3
Monte Carlo Simulation Results
Size and Size-Adjusted Power of Tests of $\rho_1 = \rho_2 = 0$

	Size $\rho_1 = \rho_2 = 0$		Power $\rho_1 = \rho_2 = 0.5$		Power $\rho_1 = \rho_2 = 0.9$	
	Wald	LR	Wald	LR	Wald	LR
$T = 300$						
High Persistence	20.4	4.7	55.7	95.5	100	100
Low Persistence	24.2	6.8	25.9	86.0	100	100
$T = 500$						
High Persistence	11.3	4.1	97.7	99.7	100	100
Low Persistence	11.8	3.6	80.8	93.3	100	100

Notes: Each cell of the table contains the percentage of 1000 Monte Carlo simulations for which the Wald test or likelihood ratio (LR) test rejected the null hypothesis that $\rho_1 = \rho_2 = 0$ at the 5% significance level. For columns labeled “Size”, critical values are based on the asymptotic distribution of the test-statistic. For columns labeled “Power”, size adjusted critical values are calculated from 1000 simulated test statistics from the corresponding Monte Carlo experiment in which $\rho_1 = \rho_2 = 0$. The data generating process used to simulate the Monte Carlo samples is given by $y_t = \mu_{S_t} + \sigma_{S_t}\epsilon_t$, where S_t evolves according to the endogenous switching model detailed in Section 2 with N=3 states and ρ_1 and ρ_2 given by the column headings.

Table 4
U.S. Real GDP Growth Rate in Quarters
Following Post-War U.S. Recessions

Quarters After Recession	Average Growth	Observations
1	6.45	11
2	6.32	11
3	5.50	11
4	5.79	10
5	4.14	10
6	4.29	10
7	3.66	10
8	3.48	9
Full Sample	3.15	262

Notes: Average growth rates are measured as annualized percentages. The sample period is 1949:Q1 to 2014:Q2. For four quarters and longer, one observation is lost due to the termination of the expansion following the 1980 recession. For eight quarters, another observation is lost due to the termination of the expansion following the 1957-1958 recession.

Table 5
Regime Switching Model for Real GDP Growth Rate

	Exogenous Switching		Endogenous Switching	
μ_0	0.89	(0.06)	1.00	(0.11)
μ_1	-0.46	(0.18)	-0.37	(0.18)
$\mu_{2,1}$	1.71	(0.45)	2.21	(0.42)
$\mu_{2,2}$	0.56	(0.15)	0.35	(0.18)
p_{00}	0.93	(0.03)	0.92	(0.04)
p_{11}	0.72	(0.09)	0.72	(0.08)
p_{22}	0.84	(0.13)	0.64	(0.10)
σ_1	0.93	(0.08)	0.90	(0.08)
σ_2	0.47	(0.03)	0.54	(0.05)
$\rho_{1,1}$			-0.29	(0.51)
$\rho_{1,2}$			0.91	(0.11)
$\rho_{2,1}$			-0.19	(0.38)
$\rho_{2,2}$			0.14	(0.43)
Likelihood	-313.44		-308.67	
	<i>Q</i> -statistic	p-value	<i>Q</i> -statistic	p-value
$Q(k = 1)$	5.59	0.02	0.08	0.78
$Q(k = 2)$	7.81	0.02	1.27	0.53
$Q(k = 4)$	8.61	0.07	2.38	0.67

Notes: This table reports maximum likelihood estimates of the three state switching mean model of U.S. real GDP growth given in equation (22). The sample period is 1952:Q1 to 2014:Q2. Standard errors, reported in parentheses, are based on second derivatives of the log-likelihood function in all cases. $Q(k)$ stands for the Ljung-Box test statistic for serial correlation in the standardized disturbance term calculated by smoothed probabilities up to k lags.

Table 6
Volatility Feedback Model with Three Volatility States

	Exogenous Switching		Endogenous Switching	
θ_1	0.38	(0.12)	0.33	(0.13)
θ_2	-5.52	(2.04)	-4.11	(1.03)
p_{00}	0.97	(0.01)	0.96	(0.01)
p_{01}	0.04	(0.04)	0.03	(0.01)
p_{10}	0.08	(0.02)	0.08	(0.01)
p_{11}	0.79	(0.04)	0.84	(0.03)
p_{21}	0.38	(0.04)	0.28	(0.05)
p_{22}	0.48	(0.11)	0.72	(0.05)
σ_0	0.37	(0.01)	0.36	(0.01)
σ_1	0.48	(0.03)	0.45	(0.02)
σ_2	0.64	(0.07)	0.72	(0.08)
ρ_1			-0.52	(0.16)
ρ_2			0.44	(0.34)
Likelihood	-487.52		-485.01	

Notes: This table reports maximum likelihood estimates of the volatility feedback model in equation 23 with three volatility states, estimated assuming both exogenous and endogenous Markov switching. The sample period is 1952:M1 to 2013:M12. Standard errors, reported in parentheses, are based on second derivatives of the log-likelihood function in all cases.