Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 143 (2008) 263–273
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Estimation of Markov regime-switching regression models with endogenous switching

Chang-Jin Kim[a,b,*], Jeremy Piger[c], Richard Startz[b]

[a]*Korea University, Seoul, South Korea*
[b]*Department of Economics, University of Washington, Box 353330, Seattle, WA 98195-3330, USA*
[c]*Department of Economics, 1285 University of Oregon, Eugene, OR 97403-1285, USA*

Available online 3 December 2007

Abstract

Following Hamilton [1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384], estimation of Markov regime-switching regressions typically relies on the assumption that the latent state variable controlling regime change is exogenous. We relax this assumption and develop a parsimonious model of endogenous Markov regime-switching. Inference via maximum likelihood estimation is possible with relatively minor modifications to existing recursive filters. The model nests the exogenous switching model, yielding straightforward tests for endogeneity. In Monte Carlo experiments, maximum likelihood estimates of the endogenous switching model parameters were quite accurate, even in the presence of certain model misspecifications. As an application, we extend the volatility feedback model of equity returns given in Turner et al. [1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. Journal of Financial Economics 25, 3–22] to allow for endogenous switching.

JEL classification: C13; C22; G12

Keywords: Endogeneity; Regime-switching

1. Introduction

Recent decades have seen extensive interest in time-varying parameter models of macroeconomic and financial time series. One notable set of models are regime-switching regressions, which date to at least Quandt (1958). Goldfeld and Quandt (1973) introduced a particularly useful version of these models, referred to in the following as a Markov-switching model, in which the latent state variable controlling regime shifts follows a Markov-chain, and is thus serially dependent. In an influential article, Hamilton (1989) extended Markov-switching models to the case of dependent data, specifically an autoregression.

The vast literature generated by Hamilton (1989) typically assumes that the regime shifts are exogenous with respect to all realizations of the regression disturbance. In this paper we work with Markov-switching

*Corresponding author. Department of Economics, University of Washington, Box 353330, Seattle, WA 98195-3330, USA.
E-mail addresses: changjin@u.washington.edu (C.-J. Kim), jpiger@u.oregon.edu (J. Piger), startz@u.washington.edu (R. Startz).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.10.002

regressions of the type considered by Hamilton (1989) and various extensions, but relax the exogenous switching assumption. We develop a Gaussian model of endogenous Markov regime switching based on a probit specification for the realization of the latent state. The model is quite parsimonious, and admits a test for endogenous switching as a simple parameter restriction. The model parameters can be estimated via maximum likelihood with relatively minor modifications to the recursive filter in Hamilton (1989).

Why are we motivated to investigate Markov-switching regressions with endogenous switching? Many of the model's applications are in macroeconomics or finance in situations where it is natural to assume the state is endogenous. As an example, it is often the case that the estimated state variable has a strong business cycle correlation. This can be seen in recent applications of the regime-switching model to identified monetary VARs, such as Sims and Zha (2006) and Owyang (2002). It is not hard to imagine that the shocks to the regression, such as the macroeconomic shocks to the VAR, would be correlated with the business cycle. As another example, some applications of the model contain parameters that represent the reaction of agents to realization of the state, as is the case in the model of equity returns given in Turner et al. (1989) (TSN hereafter). However, it is likely that agents do not observe the state, but instead draw inference based on some information set, the contents of which are unknown to the econometrician. Use of the actual state to proxy for this inference leads to a regression with measurement error in the explanatory variables, and thus endogeneity.

To evaluate the sensitivity of maximum likelihood estimation based on the Gaussian endogenous switching model-to-model misspecification, we conduct a battery of Monte Carlo experiments in which the true data generating process is a non-Gaussian endogenous switching model. These experiments suggest that quasi-maximum likelihood estimation produces accurate estimates of the parameters of the endogenous switching model, at least for the particular model misspecifications considered. We conduct additional Monte Carlo experiments to evaluate the finite sample performance of tests for endogenous switching, and find that the likelihood ratio test has close to correct size for all cases considered.

As an application, we extend the "volatility feedback" model of equity returns given in TSN to allow for endogenous switching. As discussed above, this model provides a setting in which we might reasonably expect the Markov-switching state variable to be endogenous. We find marginal statistical evidence of endogenous switching in the model and that allowing for endogeneity has substantial effects on parameter estimates.

The model of endogenous switching developed in this paper has much in common with an earlier literature using switching regressions. This literature, such as Maddala and Nelson (1975), was often concerned with endogenous switching, as the primary applications were in limited dependent variable contexts such as self-selection and market disequilibrium settings. The model we have presented here can be interpreted as an extension of the Maddala and Nelson (1975) approach, which was a model of independent switching, to the Hamilton (1989) regime-switching model, in which the state process is serially dependent.

In the next section we lay out a two-regime Markov-switching regression model with endogenous switching and discuss maximum likelihood estimation. Section 3 generalizes this model to the N-regime case. Section 4 gives the results of Monte Carlo experiments evaluating the performance of parameter inference and tests for endogenous switching. Section 5 presents the empirical example to the "volatility feedback" model of TSN. Section 6 concludes.

2. A two-regime endogenous switching model

2.1. Model specification

Consider the following Gaussian regime-switching model for the sample path of a time series, $\{y_t\}_{t=1}^{\mathrm{T}}$:

$$y_t = x_t'\beta_{S_t} + \sigma_{S_t}\varepsilon_t,$$
$$\varepsilon_t \sim i.i.d.\ \mathrm{N}(0,1), \tag{2.1}$$

where y_t is scalar, x_t is a $(k \times 1)$ vector of observed exogenous or predetermined explanatory variables, which may include lagged values of y_t, and $S_t = i$ is the state variable. Both y_t and x_t are assumed to be covariance-stationary variables. Denote the number of regimes by N, so that $i = 1, 2, \ldots, N$. We begin with the case where $N = 2$. In addition to aiding intuition, the two-regime case is a popular specification in applied work.[1]

The state variable is unobserved and is assumed to evolve according to a first-order Markov chain with transition probabilities:

$$P(S_t = i|S_{t-1} = j, z_t) = P_{ij}(z_t). \tag{2.2}$$

In (2.2), the transition probabilities are influenced by a $(q \times 1)$ vector of covariance-stationary exogenous or predetermined variables z_t, where z_t may include elements of x_t. The Markov chain is assumed to be stationary, and to evolve independently of all observations of those elements of x_t not included in z_t.[2]

To model the influence of z_t on the [0,1] transition probabilities in (2.2) we use a probit specification for S_t:

$$S_t = \begin{Bmatrix} 1 & \text{if} & \eta_t < a_{S_{t-1}} + z_t' b_{S_{t-1}} \\ 2 & \text{if} & \eta_t \geqslant a_{S_{t-1}} + z_t' b_{S_{t-1}} \end{Bmatrix},$$

$$\eta_t \sim i.i.d.\ \mathrm{N}(0, 1). \tag{2.3}$$

The transition probabilities are then:

$$p_{1j}(z_t) = P(\eta_t < a_j + z_t' b_j) = \Phi(a_j + z_t' b_j), \tag{2.4}$$

$$p_{2j}(z_t) = P(\eta_t \geqslant a_j + z_t' b_j) = 1 - \Phi(a_j + z_t' b_j),$$

where Φ is the standard normal cumulative distribution function.[3]

To model endogenous switching, assume that the joint density function of ε_t and η_t is bivariate normal:

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathrm{N}(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{2.5}$$

where ε_t and η_{t-h} are uncorrelated $\forall h \neq 0$. Regime-switching models found in time-series applications nearly always make the assumption that ε_t is independent of S_{t-h}, $\forall h$, which corresponds to the restriction that $\rho = 0$ in the model presented here.[4]

2.2. Maximum likelihood estimation

Let $\Omega_t = (x_t', x_{t-1}', \ldots, x_1', z_t', z_{t-1}', \ldots, z_1')'$ and $\xi_t = (y_t, y_{t-1}, \ldots, y_1)'$ be vectors containing observations observed through date t, and $\theta = (\beta_1, \sigma_1, a_1, b_1, \beta_2, \sigma_2, a_2, b_2, \rho)$ be the vector of model parameters. The conditional likelihood function for the observed data ζ_t is constructed as $L(\theta) = \prod_{t=1}^{T} f(y_t|\Omega_t, \xi_{t-1}; \theta)$, where:

$$f(y_t|\Omega_t, \xi_{t-1}; \theta) = \sum_i \sum_j f(y_t|S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$$
$$Pr(S_t = i, S_{t-1} = j|\Omega_t, \xi_{t-1}; \theta). \tag{2.6}$$

[1] As the regime ordering is arbitrary, we assume that the model in (2.1) is appropriately normalized. See Hamilton et al. (2007) for detailed discussion of this issue.

[2] Several special cases of (2.2) are worth mentioning. The unrestricted model is the time-varying transition probability Markov-switching model of Goldfeld and Quandt (1973), Diebold et al. (1994) and Filardo (1994). When the transition probabilities are not influenced by S_{t-1}, we have the time-varying transition probability independent switching model of Goldfeld and Quandt (1972). When the transition probabilities are not influenced by z_t, we have the fixed transition probability Markov-switching model of Goldfeld and Quandt (1973) and Hamilton (1989). When the transition probabilities are influenced by neither z_t or S_{t-1}, we have the fixed transition probability independent switching model of Quandt (1972).

[3] Alternatively, a logistic specification could be used to describe the transition probabilities as in Diebold et al. (1994) or Filardo (1994). The probit specification is used here because it provides a straightforward approach to model endogenous switching.

[4] In recent work, Chib and Dueker (2004) develop a non-Markov regime switching model in which observable variables are related to the sign of a Gaussian autoregressive latent state variable, the innovations to which are allowed to be correlated with the model residual through a bivariate normal specification as in (2.5). The authors develop Bayesian procedures to estimate this model.

The weighting probability in (2.6) is computed recursively by applying Bayes' rule:

$$\begin{aligned}
&Pr(S_t = i, S_{t-1} = j|\Omega_t, \xi_{t-1}; \theta) = P_{ij}(z_t)Pr(S_{t-1} = j|\Omega_t, \xi_{t-1}; \theta),\\
&Pr(S_t = i|\Omega_{t+1}, \xi_t; \theta) = Pr(S_t = i|\Omega_t, \xi_t; \theta)\\
&\quad = \frac{1}{f(y_t|\Omega_t, \xi_{t-1}; \theta)} \sum_j f(y_t|S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)\\
&\qquad \times Pr(S_t = i, S_{t-1} = j|\Omega_t, \xi_{t-1}; \theta).
\end{aligned} \tag{2.7}$$

To initialize (2.7), the usual practice is to approximate $P(S_0 = j|\Omega_1, \xi_0; \theta)$ with the unconditional probability, $P(S_0 = j; \theta)$. Alternatively, this initial probability can be treated as an additional parameter to be estimated.

To complete the recursion in (2.6)–(2.7), we require the regime-dependent conditional density function, $f(y_t|S_t = i,\ S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$. For the exogenous switching case (i.e. when $\rho = 0$) this density function is Gaussian:

$$f(y_t|S_t = i,\ S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta) = \frac{1}{\sigma_i}\phi\left(\frac{y_t - x_t'\beta_i}{\sigma_i}\right), \tag{2.8}$$

where ϕ is the standard normal probability density function. However, for non-zero values of $\rho \in (-1,1)$, $f(y_t|S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$ is given by[5]

$$f(y_t|S_t = 1, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta) = \frac{\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\Phi\left(\frac{a_j + z_t'b_j - \rho((y_t - x_t'\beta_1)/\sigma_1)}{\sqrt{1-\rho^2}}\right)}{\sigma_1 p_{1j}(z_t)}, \tag{2.9}$$

$$f(y_t|S_t = 2, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta) = \frac{\phi\left(\frac{y_t - x_t'\beta_2}{\sigma_2}\right)\Phi\left(\frac{-(a_j + z_t'b_j) + \rho((y_t - x_t'\beta_2)/\sigma_2)}{\sqrt{1-\rho^2}}\right)}{\sigma_2 p_{2j}(z_t)}.$$

When S_t is endogenous, maximum likelihood estimation assuming S_t is exogenous, and thus based on the distribution in (2.8), is inconsistent in general. To see this, note that:

$$\begin{aligned}
&\mathrm{E}(\varepsilon_t|S_t = 1, S_{t-1} = j; \theta) = \mathrm{E}(\varepsilon_t|\eta_t < a_j + z_t'b_j) = -\rho\frac{\phi(a_j + z_t'b_j)}{\Phi(a_j + z_t'b_j)},\\
&\mathrm{E}(\varepsilon_t|S_t = 2, S_{t-1} = j; \theta) = \mathrm{E}(\varepsilon_t|\eta_t \geqslant a_j + z_t'b_j) = \rho\frac{\phi(a_j + z_t'b_j)}{1 - \Phi(a_j + z_t'b_j)}.
\end{aligned} \tag{2.10}$$

Thus, when $\rho \neq 0$, the regime-dependent conditional mean of ε_t is non-zero, implying that maximum likelihood estimates based on (2.8) suffer from the ordinary problem of omitted variables. Another, less obvious, source of inconsistency arises because $f(y_t|S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$ is non-Gaussian when $\rho \neq 0$, as is clear from (2.9). In this case maximum likelihood estimation based on (2.8) is quasi-maximum likelihood estimation, which, as pointed out in Campbell (2002), is inconsistent for regime-switching models in general.

2.3. Testing for endogeneity

In the model of endogenous switching presented above, the null hypothesis that S_t is exogenous is equivalent to the scalar restriction $\rho = 0$. Thus, a test for exogeneity can be carried out by any suitable test of this restriction. One obvious choice is based on the t-statistic:

$$t = \frac{\hat{\rho}}{se(\hat{\rho})}, \tag{2.11}$$

[5]The density (2.9) belongs to the "skew-normal" family of density functions, which are commonly credited to Azzalini (1985). See Arnold and Beaver (2002) for a survey of this literature.

where $se(\hat{\rho})$ is an estimate of the standard error of $\hat{\rho}$. Assuming the likelihood function is correctly specified, an appropriate $se(\hat{\rho})$ can be constructed from an estimate of the inverse of the information matrix, such as that based on the negative of the second derivative of the log-likelihood function. Alternatively, one could test for endogeneity using the likelihood ratio statistic, constructed as

$$LR = 2(L(\hat{\theta}) - L(\hat{\theta}_R)), \tag{2.12}$$

where $L(\hat{\theta})$ is the maximized value of the likelihood function, and $L(\hat{\theta}_R)$ is the maximized value of the likelihood function under the restriction that $\rho = 0$. If the likelihood function is correctly specified, both t and LR have their usual asymptotic distributions when $\rho = 0$. For further details, see Hamilton (1994).

3. An N-regime endogenous switching model

In this section we generalize the two-regime Gaussian endogenous-switching model presented in Section 2 to N regimes. We begin by modifying the probit specification of the transition probabilities given in (2.3). Suppose the realization of S_t is now determined by the outcome of $\eta_t \sim i.i.d.$ N(0, 1) as follows:

$$S_t = \left\{ \begin{array}{llcccl} 1 & \text{if} & -\infty & \leqslant \eta_t & < & a_{1,j} + z_t' b_{1,j} \\ 2 & \text{if} & a_{1,j} + z_t' b_{1,j} & \leqslant \eta_t & < & a_{2,j} + z_t' b_{2,j} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ N-1 & \text{if} & a_{N-2,j} + z_t' b_{N-2,j} & \leqslant \eta_t & < & a_{N-1,j} + z_t' b_{N-1,j} \\ N & \text{if} & a_{N-1,j} + z_t' b_{N-1,j} & \leqslant \eta_t & < & \infty \end{array} \right\}. \tag{3.1}$$

The transition probabilities, $p_{ij}(z_t)$, are then given as follows:

$$p_{ij}(z_t) = \Phi(c_{i,j,t}) - \Phi(c_{i-1,j,t}), \tag{3.2}$$

where $c_{0,j,t} = -\infty$, $c_{N,j,t} = \infty$, and $c_{i,j,t} = a_{i,j} + z_t' b_{i,j}$ for $0 < i < N$.

Again, to model endogenous switching, assume that the joint density of ε_t and η_t is bivariate normal as in (2.5). Let the vector of model parameters be $\theta = (\theta_1', \theta_2', \ldots, \theta_N', \rho)'$, where $\theta_i = (\beta_i, \sigma_i, a_i, b_i)'$. Given $f(y_t | S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$, the likelihood function, $L(\theta)$, can again be constructed using the recursion in (2.6)–(2.7). It can be shown that[6]:

$$f\left(y_t | S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta\right) = \frac{\phi\left(\frac{y_t - x_t'\beta_i}{\sigma_i}\right)\left(\Phi\left(\frac{c_{i,j,t} - \rho\left(\frac{y_t - x_t'\beta_i}{\sigma_i}\right)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{c_{i-1,j,t} - \rho\left(\frac{y_t - x_t'\beta_i}{\sigma_i}\right)}{\sqrt{1-\rho^2}}\right)\right)}{\sigma_i p_{ij}(z_t)}. \tag{3.3}$$

Finally, as with the two-regime endogenous-switching model, a test of the null hypothesis that S_t is exogenous is equivalent to a test of the restriction $\rho = 0$.

4. Monte Carlo analysis

In this section we provide Monte Carlo evidence regarding the sensitivity of maximum likelihood estimation based on the joint normality assumption in (2.5) to departures from this Gaussian assumption in the data generating process. Such a departure renders the estimator based on (2.5) a quasi-maximum likelihood (QML) estimator, which is inconsistent for Markov-switching models in general (Campbell, 2002). Our Monte Carlo experiments then provide some limited evidence of how badly the QML estimator performs in practice.[7] We

[6] We provide a derivation of (3.3) in an unpublished appendix, available at: http://www.uoregon.edu/~jpiger/.

[7] In untabulated results, available from the authors, we have also conducted Monte Carlo experiments in which the data generating process maintains the joint normality assumption given in (2.5). These results suggest that maximum likelihood estimation of the

also present Monte Carlo evidence regarding the finite sample performance of the t and likelihood ratio tests for endogenous switching.

Given its prominence in the applied literature, we focus on the two-regime model with fixed, Markov-switching transition probabilities, so that $b_1 = b_2 = 0$. For each Monte Carlo experiment, 1000 simulated series are generated from the model given in (2.1)–(2.3). We consider two sample sizes for the simulated series, $T = 200$ and 500. For each simulation, the vector of exogenous explanatory variables is set to $x_t = [1 \quad x_t^*]$, where $x_t^* \sim i.i.d.\ \mathrm{N}(0, 2)$, and the vector of regime-switching parameters is set to $\beta_1 = (\beta_{0,1}, \beta_{1,1})' = (1.0, 1.0)'$, $\beta_2 = (\beta_{0,2}, \beta_{1,2})' = (-1.0, -1.0)'$, $\sigma_1 = 0.33$, and $\sigma_2 = 0.67$. We consider three different sets of transition probabilities corresponding to moderate persistence ($p_{11} = 0.7$, $p_{22} = 0.7$), high persistence ($p_{11} = 0.9$, $p_{22} = 0.9$), and differential persistence ($p_{11} = 0.7$, $p_{22} = 0.9$). We also consider three different values for ρ, corresponding to high correlation ($\rho = 0.9$), moderate correlation ($\rho = 0.5$), and zero correlation ($\rho = 0$), where the zero correlation case is used to evaluate the size performance of tests for endogenous switching. Finally, to produce a non-Gaussian joint density for ε_t and η_t, we generate ε_t as a standard normal random variable, and η_t as a weighted sum of ε_t and a t-distributed random variable with four degrees of freedom. The weighting is calibrated so that $(\varepsilon_t, \eta_t)'$ has covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & \rho\gamma_4 \\ \rho\gamma_4 & \gamma_4^2 \end{bmatrix},$$

where $\gamma_4^2 = 2$ is the variance of a t-distributed random variable with four degrees of freedom.

For each simulated time series, two sets of maximum likelihood estimates are computed.[8] The first, which we label the "exogenous" estimator, assumes that $\rho = 0$, and is thus based on the recursion in (2.6)–(2.7), using (2.8) to measure $f(y_t | S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$. The second, which we label the "endogenous" estimator, allows for $\rho \neq 0$, and is thus based on the recursion in (2.6)–(2.7), using (2.9) to measure $f(y_t | S_t = i, S_{t-1} = j, \Omega_t, \xi_{t-1}; \theta)$. Finally, we also record the outcome of 5% nominal size t and likelihood ratio tests of the null hypothesis $\rho = 0$.[9] For those cases where $\rho = 0$ in the data generating process, these tests document the empirical size of the 5% nominal size tests. For those cases where $\rho \neq 0$, we use size-adjusted critical values, taken from the Monte Carlo simulations generated with $\rho = 0$, to measure the power of the tests.

Tables 1 and 2 show the results of the Monte Carlo experiments investigating maximum likelihood estimation of the endogenous-switching model, with Table 1 holding results for experiments in which $\rho = 0.5$ and Table 2 holding results for experiments in which $\rho = 0.9$. For the parameters β_1, β_2, σ_1, σ_2, each table shows the mean of the 1000 maximum likelihood point estimates, as well as the root mean squared error (RMSE) of the 1000 maximum likelihood point estimates from the true value of the parameter.[10] The results suggest that for the particular data generating process considered, the approximation provided by the normality assumption in (2.5) is quite good. For both sample sizes and all values of the transition probabilities and ρ considered, the mean parameter estimates from the endogenous estimator are very close to their true values. While this result may not generalize to non-normal distributions more generally, it is suggestive that the quality of the endogenous estimator is not hyper-sensitive to the joint-normality assumption.

Tables 1 and 2 also demonstrate the estimation bias that occurs when the endogenous state variable is treated as exogenous in estimation. When the exogenous estimator is used, the mean estimates of $\beta_{0,1}$ and $\beta_{0,2}$ are far from their true values, with the bias larger for higher values of ρ. The mean estimates of σ_1 and σ_2 are also biased downward. Note that the mean estimates are nearly identical in the $T = 200$ and 500 cases, suggesting the bias is not a small sample phenomenon. Also note that the estimates of $\beta_{1,1}$ and $\beta_{1,2}$ are close to

(*footnote continued*)
endogenous switching model performs quite well, producing accurate model parameter estimates for all parameterizations and sample sizes considered.

[8]All computations were performed in GAUSS 8.0 using the QNewton numerical optimization package.

[9]The t-tests were constructed using a standard error estimate based on the second derivative of the log-likelihood function. Results when the standard error estimate is alternatively based on the outer product of the gradient are very similar, and are available from the authors.

[10]Model estimation also produces estimates of the transition probabilities, and, in the case of the endogenous estimator, the correlation parameter ρ. Although not reported, results for these parameter estimates are qualitatively similar to those for the conditional mean and variance parameters of the regression model.

Table 1
Monte Carlo results $\rho = 0.5$

	$\beta_{0,1} = 1.0$	$\beta_{0,2} = -1.0$	$\beta_{1,1} = 1.0$	$\beta_{1,2} = -1.0$	$\sigma_1 = 0.33$	$\sigma_2 = 0.67$
$T = 200$						
$p_{11} = 0.7$, $p_{22} = 0.7$						
Exog. estimator	0.87 (0.13)	−0.73 (0.27)	1.00 (0.02)	−1.00 (0.03)	0.30 (0.04)	0.61 (0.08)
Endog. estimator	1.00 (0.07)	−1.00 (0.14)	1.00 (0.02)	−1.00 (0.03)	0.33 (0.04)	0.67 (0.07)
$p_{11} = 0.7$, $p_{22} = 0.9$						
Exog. estimator	0.85 (0.16)	−0.90 (0.11)	1.00 (0.03)	−1.00 (0.03)	0.31 (0.04)	0.64 (0.05)
Endog. estimator	1.00 (0.09)	−1.00 (0.07)	1.00 (0.03)	−1.00 (0.03)	0.33 (0.04)	0.67 (0.05)
$p_{11} = 0.9$, $p_{22} = 0.9$						
Exog. estimator	0.94 (0.07)	−0.88 (0.14)	1.00 (0.02)	−1.00 (0.04)	0.32 (0.03)	0.65 (0.05)
Endog. estimator	1.00 (0.04)	−1.00 (0.09)	1.00 (0.02)	−1.00 (0.03)	0.33 (0.03)	0.67 (0.05)
$T = 500$						
$p_{11} = 0.7$, $p_{22} = 0.7$						
Exog. estimator	0.87 (0.13)	−0.74 (0.26)	1.00 (0.01)	−1.00 (0.02)	0.30 (0.03)	0.61 (0.06)
Endog. estimator	1.00 (0.04)	−1.00 (0.08)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.02)	0.67 (0.04)
$p_{11} = 0.7$, $p_{22} = 0.9$						
Exog. Estimator	0.85 (0.15)	−0.90 (0.11)	1.00 (0.02)	−1.00 (0.02)	0.31 (0.03)	0.65 (0.03)
Endog. estimator	1.00 (0.05)	−1.00 (0.04)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.03)	0.67 (0.03)
$p_{11} = 0.9$, $p_{22} = 0.9$						
Exog. estimator	0.95 (0.06)	−0.89 (0.12)	1.00 (0.01)	−1.00 (0.02)	0.32 (0.02)	0.66 (0.03)
Endog. estimator	1.00 (0.02)	−1.00 (0.05)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.02)	0.67 (0.03)

Notes: This table contains summary results from 1000 Monte Carlo simulations when the true data generating process is characterized by endogenous switching (detailed in Section 4) with $\rho = 0.5$. 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). Exog. estimator refers to the maximum likelihood estimator assuming the state process is exogenous, so that $\rho = 0$. Endog. estimator refers to the maximum likelihood estimator allowing the state process to be endogenous, so that $\rho \in (-1,1)$.

their true values. The accuracy of these parameter estimates can be traced to the model assumption, maintained in the Monte Carlo samples, that x_t^* is independent of the endogenous state variable S_t.

Table 3 reports the size and size-adjusted power of the 5% nominal size t and likelihood ratio tests of the null hypothesis that $\rho = 0$ for the data generating processes considered in Tables 1 and 2. When the null hypothesis is true, the t-test is somewhat oversized, with rejection rates close to 13% when $T = 200$. However, this appears to be a small sample phenomena, as the t-test has roughly correct size when $T = 500$.[11] In contrast, the likelihood ratio test has roughly correct size for all cases considered. When the alternative hypothesis is true, the t-test and likelihood ratio test have similar size-adjusted power for most of the alternatives considered. The one exception is when $T = 200$ and $p_{11} = p_{22} = 0.7$, in which case the likelihood ratio test has significantly higher size-adjusted power than the t-test.[12]

Overall, the Monte Carlo experiments suggest that maximum likelihood estimates using the endogenous estimator are quite accurate, even in the presence of a specific departure in the data generating process from the joint normality assumption in (2.5), while the exogenous estimator produces substantially biased parameter estimates when the true process has endogenous switching. Also, the likelihood ratio test appears to be a fairly reliable test for endogenous switching. In the next section we turn to an empirical application of the endogenous-switching model.

[11]The poor performance of the t-test in small samples is consistent with a literature investigating the finite sample properties of tests for sample selection bias, which are closely related to the tests for endogenous switching considered here. In particular, Nawata and McAleer (2001) present Monte Carlo evidence that the t-test for sample selection bias can be significantly oversized in small samples, while the likelihood ratio test has approximately correct size. They trace the source of the small sample distortions to inaccuracies with standard asymptotic variance estimators when the estimate of the correlation parameter driving the extent of sample selection bias falls close to a boundary value. In our case, this corresponds to an estimate of ρ that is close to the boundary of $|\rho| = 1$.

[12]The size and power performance of 1% and 10% nominal size tests (not reported) was very similar to that for the 5% nominal size tests. In particular, the t-test is oversized when $T = 200$, the likelihood ratio test has close to correct size in all cases, and the tests have similar size-adjusted power for most of the alternatives considered.

Table 2
Monte Carlo results $\rho = 0.9$

	$\beta_{0,1} = 1.0$	$\beta_{0,2} = -1.0$	$\beta_{1,1} = 1.0$	$\beta_{1,2} = -1.0$	$\sigma_1 = 0.33$	$\sigma_2 = 0.67$
$T = 200$						
$p_{11} = 0.7$, $p_{22} = 0.7$						
Exog. estimator	0.79 (0.21)	−0.58 (0.42)	1.00 (0.01)	−1.00 (0.03)	0.25 (0.08)	0.52 (0.16)
Endog. estimator	1.00 (0.04)	−0.99 (0.08)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.03)	0.67 (0.07)
$p_{11} = 0.7$, $p_{22} = 0.9$						
Exog. estimator	0.75 (0.26)	−0.83 (0.18)	1.00 (0.02)	−1.00 (0.03)	0.28 (0.06)	0.59 (0.08)
Endog. estimator	1.00 (0.06)	−1.00 (0.06)	1.00 (0.02)	−1.00 (0.02)	0.33 (0.04)	0.67 (0.05)
$p_{11} = 0.9$, $p_{22} = 0.9$						
Exog. estimator	0.90 (0.10)	−0.80 (0.21)	1.00 (0.02)	−1.00 (0.03)	0.31 (0.03)	0.63 (0.06)
Endog. estimator	1.00 (0.04)	−1.00 (0.07)	1.00 (0.02)	−1.00 (0.03)	0.33 (0.02)	0.67 (0.05)
$T = 500$						
$p_{11} = 0.7$, $p_{22} = 0.7$						
Exog. estimator	0.80 (0.20)	−0.57 (0.43)	1.00 (0.01)	−1.00 (0.02)	0.25 (0.08)	0.52 (0.16)
Endog. estimator	0.99 (0.03)	−0.99 (0.05)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.02)	0.67 (0.04)
$p_{11} = 0.7$, $p_{22} = 0.9$						
Exog. estimator	0.75 (0.25)	−0.83 (0.18)	1.00 (0.01)	−1.00 (0.02)	0.29 (0.04)	0.60 (0.08)
Endog. estimator	1.00 (0.04)	−1.00 (0.04)	1.00 (0.01)	−1.00 (0.01)	0.33 (0.02)	0.67 (0.03)
$p_{11} = 0.9$, $p_{22} = 0.9$						
Exog. estimator	0.90 (0.10)	−0.79 (0.21)	1.00 (0.01)	−1.00 (0.02)	0.31 (0.03)	0.63 (0.05)
Endog. estimator	1.00 (0.02)	−1.00 (0.04)	1.00 (0.01)	−1.00 (0.02)	0.33 (0.02)	0.67 (0.03)

Notes: This table contains summary results from 1000 Monte Carlo simulations when the true data generating process is characterized by endogenous switching (detailed in Section 4) with $\rho = 0.9$. 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). Exog. estimator refers to the maximum likelihood estimator assuming the state process is exogenous, so that $\rho = 0$. Endog. estimator refers to the maximum likelihood estimator allowing the state process to be endogenous, so that $\rho \in (-1,1)$.

Table 3
Monte Carlo results size and size adjusted power of tests of $\rho = 0$

	Size $\rho = 0$		Power $\rho = 0.5$		Power $\rho = 0.9$	
	t	LR	t	LR	t	LR
$T = 200$						
$p_{11} = 0.7$, $p_{22} = 0.7$	12.7	6.7	48.0	57.2	99.9	100
$p_{11} = 0.7$, $p_{22} = 0.9$	7.9	5.7	72.1	72.5	100	99.9
$p_{11} = 0.9$, $p_{22} = 0.9$	7.7	6.3	82.0	83.5	100	100
$T = 500$						
$p_{11} = 0.7$, $p_{22} = 0.7$	7.0	5.7	94.6	95.6	100	100
$p_{11} = 0.7$, $p_{22} = 0.9$	5.7	5.0	97.3	97.6	100	100
$p_{11} = 0.9$, $p_{22} = 0.9$	6.2	5.6	99.8	99.8	100	100

Notes: Each cell of the table contains the percentage of 1000 Monte Carlo simulations for which the t-test or likelihood ratio (LR) test described in Section 2.3 rejected the null hypothesis that $\rho = 0$ 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 the 1000 simulated test statistics from the corresponding Monte Carlo experiment in which $\rho = 0$. The data generating process used to simulate the Monte Carlo samples is detailed in Section 4.

5. Application: measurement error and estimation of the volatility feedback effect

A stylized fact of US equity return data is that the volatility of realized returns is time-varying and predictable. Given this, classic portfolio theory would imply that the equity risk premium should also be time-varying and respond positively to the expectation of future volatility. However, the data suggest that realized returns and realized volatility, as measured by squared returns, are negatively correlated.[13]

One explanation for the observed data is that while investors do require an increase in expected return in exchange for expected future volatility, they are often surprised by news about realized volatility. This "volatility feedback effect" creates a reduction in prices in the period in which the increase in volatility is realized. If the volatility feedback effect is strong enough, it may create a negative contemporaneous correlation between realized returns and volatility in the data. The volatility feedback effect has been investigated extensively in the literature, see for example French et al. (1987), TSN, Campbell and Hentschel (1992), Bekaert and Wu (2000) and Kim et al. (2004).

TSN model the volatility feedback effect with a Markov-switching model:

$$\begin{aligned} r_t &= \theta_1 \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1}) + \theta_2(\mathrm{E}(\sigma^2_{S_t}|\psi^*_t) - \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1})) + \sigma_{S_t}\varepsilon_t, \\ \varepsilon_t &\sim i.i.d.\ \mathrm{N}(0,1), \end{aligned} \tag{5.1}$$

where S_t is a discrete Markov-switching variable taking on values 1 or 2, with transition probabilities p_{ij} parameterized as in Eq. (2.4). For normalization we assume $\sigma^2_2 > \sigma^2_1$, so that $S_t = 2$ is the high volatility state.

The model in (5.1) is motivated as follows. At the beginning of period t, the risk premium, $\theta_1 \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1})$, is determined based on the expectation of period t volatility formed with information available at the end of period $t-1$. During period t additional information regarding volatility is observed. By the end of period t, this information is collected in the information set ψ^*_t. When $\mathrm{E}(\sigma^2_{S_t}|\psi^*_t) \neq \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1})$, information about volatility revealed during the period has surprised agents. If $\theta_2 < 0$, surprises that reveal greater probability of the high-variance state are viewed negatively by investors, and thus reduce the contemporaneous return.

One estimation difficulty with the model in (5.1) is that there exists a discrepancy between the investors' and the econometrician's data set. In particular, while ψ_{t-1} may be summarized by all data up to $t-1$, that is $\psi_t = \{r_{t-1}, r_{t-2}, \ldots\}$, the information set ψ^*_t includes information that is not summarized in the researcher's data set on observed returns. This is because, while the researcher's data set is collected discretely at the beginning of each period, the market participants continuously observe trades that occur during the period.

To handle this estimation difficulty, TSN use the actual volatility, $\sigma^2_{S_t}$, as a proxy for $\mathrm{E}(\sigma^2_{S_t}|\psi^*_t)$. That is, they estimate:

$$\begin{aligned} r_t &= \theta_1 \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1}) + \theta_2(\sigma^2_{S_t} - \mathrm{E}(\sigma^2_{S_t}|\psi_{t-1})) + \sigma_{S_t} u_t \\ u_t &\sim \mathrm{N}(0,1). \end{aligned} \tag{5.2}$$

In essence, this approximation replaces the estimated probability of the state, $P(S_t = i|\psi^*_t)$, with one if $S_t = i$ and zero otherwise. Assuming these differ, this introduces classical measurement error into the state variable in the estimated equation, thus rendering it endogenous.

The existing literature estimates (5.2) assuming the state variable is exogenous. However, the techniques developed in Section 2 can be used to estimate the volatility feedback model allowing for endogeneity, as well as to test for endogeneity. Here we estimate (5.2) using monthly returns for a value-weighted portfolio of all NYSE-listed stocks in excess of the one-month Treasury Bill rate over the sample period January 1952 to December 1999, the same data as used in Kim et al. (2004). Table 4 summarizes the results.

The first panel of Table 4 shows estimates when endogeneity is ignored. These estimates, which are similar to those in TSN, are consistent with both 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, that is θ_1 is very small relative to θ_2. The second panel shows the estimates when endogeneity is allowed, so that the correlation parameter ρ is estimated. The estimate of ρ is substantial, equaling -0.40. The likelihood ratio test, which recorded reliable finite sample size performance in the Monte Carlo experiments discussed in Section 4, provides marginal evidence against the null hypothesis that $\rho = 0$ (p-value $= 0.081$).[14] The primary difference in the parameter estimates is for the volatility feedback coefficient

[13]For a recent discussion of this result, see Brandt and Kang (2004).

[14]It is worth emphasizing that the validity of the likelihood ratio test for exogenous switching relies on the correct specification of the model likelihood function. Evidence in favor of endogenous switching should therefore be interpreted conditional on this maintained hypothesis.

Table 4
Maximum likelihood estimates of the Turner et al. (1989) volatility-feedback model

Parameter	Ignoring endogeneity	Accounting for endogeneity
θ_1	0.31 (0.10)	0.36 (0.10)
θ_2	−1.55 (0.45)	−1.07 (0.45)
σ_1	0.40 (0.02)	0.40 (0.02)
σ_2	0.75 (0.07)	0.74 (0.07)
a_1	2.05 (0.20)	2.05 (0.17)
a_2	−1.09 (0.21)	−1.16 (0.22)
ρ	–	−0.40 (0.18)
Log likelihood	−372.41	−370.89

Notes: This table reports maximum likelihood estimates of the "volatility feedback" model of excess equity returns given in Turner et al. (1989) and detailed in Eq. (5.2). Excess returns are measured using monthly returns in excess of the one-month Treasury Bill rate generated from a value-weighted portfolio of all NYSE-listed stocks. The sample period is January 1952 through December 1999. The column labeled "Ignoring Endogeneity" holds estimates in which the Markov-switching state variable is assumed exogenous of the regression error term. The column labeled "Accounting for Endogeneity" holds estimates in which the Markov-switching state variable is allowed to be endogenous using the approach detailed in Section 2. Standard errors, reported in parentheses, are based on second derivatives of the log-likelihood function in all cases.

θ_2, which is estimated to be about one-third smaller when endogeneity is allowed than when it is ignored. Thus, while there is still evidence of a strong volatility feedback effect, it is substantially smaller than that implied by the model with no allowance for endogeneity.

6. Conclusion

We have developed a model of Markov-switching in which the latent state variable controlling the regime shifts is endogenously determined. The model is quite parsimonious, and admits a test for endogenous switching as a simple parameter restriction. The model parameters can be estimated via maximum likelihood with relatively minor modifications to the recursive filter in Hamilton (1989). In Monte Carlo experiments, maximum likelihood estimation of the endogenous-switching model and the likelihood ratio test for endogeneity performed quite well, even in the presence of certain model misspecifications. We apply the model to test for endogenous switching in the volatility feedback model of equity returns given in Turner et al. (1989).

Acknowledgments

Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington. Startz acknowledges financial support from the Cecil and Jane Castor Professorship at the University of Washington. We thank the editor, an associate editor, three anonymous referees, Barry Arnold, Robert Beaver, Michael Dueker, James Morley, Charles Nelson and seminar participants at USC and the 2003 NBER Summer Institute for helpful comments. Parts of this paper were written while Piger was a Senior Economist at the Federal Reserve Bank of St. Louis. The views expressed in this paper should not be interpreted as those of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

References

Arnold, B.C., Beaver, R.J., 2002. Skewed multivariate models related to hidden truncation and/or selective reporting. Sociedad de Estadistica e Investigacion Operativa 11, 7–54.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Brandt, M., Kang, Q., 2004. On the relationship between the conditional mean and volatility of stock returns. A Latent VAR Approach 72, 217–257.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Campbell, S.D., 2002. Specification Testing and Semiparametric Estimation of Regime Switching Models: An Examination of the US Short Term Interest Rate, Brown University Department of Economics Working Paper #2002-26.

Chib, S., Dueker, M., 2004. Non-Markovian Regime Switching with Endogenous States and Time Varying State Lengths, Federal Reserve Bank of St. Louis working paper #2004-030A.

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

Filardo, A.J., 1994. Business-cycle phases and their transitional dynamics. Journal of Business and Economic Statistics 12, 299–308.

French, K.R., Schwert, G., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Goldfeld, S.M., Quandt, R.E., 1972. Nonlinear methods in econometrics. North Holland, Amsterdam.

Goldfeld, S.M., Quandt, R.E., 1973. A Markov model for switching regressions. Journal of Econometrics 1, 3–16.

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

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hamilton, J.D., Waggoner, D.F., Zha, T., 2007. Normalization in econometrics. Econometric Reviews 26, 221–252.

Kim, C.J., Morley, J.C., Nelson, C.R., 2004. Is there a positive relationship between stock market volatility and the equity premium? Journal of Money, Credit and Banking 36, 339–360.

Maddala, G.S., Nelson, F., 1975. Switching regression models with exogenous and endogenous switching. Proceedings of the American Statistical Association, 423–426.

Nawata, K., McAleer, M., 2001. Size characteristics of tests for sample selection bias: a Monte Carlo comparison and empirical example. Econometric Reviews 20, 105–112.

Owyang, M., 2002. Modeling Volcker as a Non-Absorbing State: Agnostic Identification of a Markov-Switching VAR, working paper, Federal Reserve Bank of St. Louis.

Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association 53, 873–880.

Quandt, R.E., 1972. A new approach to estimating switching regressions. Journal of the American Statistical Association 67, 306–310.

Sims, C., Zha, T., 2006. Were there regime switches in US monetary policy? American Economic Review 96, 54–81.

Turner, C.M., Startz, R., Nelson, C.R., 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. Journal of Financial Economics 25, 3–22.