Economics Bulletin

Volume 29, Issue 2

Misspecified Markov Switching Model

Youngki Shin
Department of Economics, University of Western Ontario

Abstract

I characterize the local power of an optimal test for a Markov Switching model under generalized alternatives. The result shows that the test still has power for the model with endogenous stochastic parameters unless they are orthogonal to the score functions.

I would like to thank Werner Ploberger and Marine Carrasco for their encouragement and guidance.
Citation: Youngki Shin, (2009) "Misspecified Markov Switching Model", *Economics Bulletin*, Vol. 29 no.2 pp. 957-963.
Submitted: Aug 22 2008. **Published:** May 12, 2009.

1 Introduction

Nonlinear dynamic models such as structural change, threshold, and markov switching models have received much interest from both econometric theory and empirical studies. They provide considerable flexibility in modeling, which gives more accurate predictions. Considering that regime switches or financial market crashes are easily observed in economic data sets, these nonlinear dynamic models can be a good analyzing tool in addition to existing linear models such as autoregressive models. In some cases, however, deciding whether the time series of the data is linear or not is not obvious, and requires a statistical testing procedure.

Most testing procedures for nonlinear models are directional tests, i.e., a test is designed for a single nonlinear model alternative and may not be consistent over all nonlinear models. Therefore, if a test has no power for a different nonlinear model and the true model is misspecified, then a linear model might be adopted even though the true one is nonlinear. For example, Carrasco (2002) shows that the test for a structural change has no power when the true data are generated either by threshold or markov switching models.

In this note, I characterize the performance of the test for a markov switching model under misspecification. Specifically, I consider the CHP test proposed by Carrasco, Hu, and Ploberger (2007) when the stochastic parameter has an endogenous factor, i.e., it depends on past time series like threshold models. The result reveals that the CHP test still has nontrivial local power under misspecified threshold models. Therefore, it can be applied as a pretest for detecting nonlinear stationary models. The result implies that, in practice, one can exclude the class of threshold models from possible candidates when the CHP test cannot reject the null of a linear model.

2 Main Result

First, I will briefly introduce the CHP test. The observations are given by $\{y_1, y_2, \cdots, y_T\}$. Let $l_t(\theta_t)$ denote the log of conditional density of y_t given $y_{t-1}, \cdots, y_1$, where θ_t is a k-dimensional parameter vector. The testing problem is

$$
\begin{aligned}
H_0 &: \theta_t = \theta_0 \\
H_1 &: \theta_t = \theta_0 + \eta_t
\end{aligned}
$$

where θ_0 is a constant and η_t is a markov chain that does not depend on $y_{t-1}, \cdots, y_1$. The stochastic parameter, η_t, is summarized by a vector of nuisance parameters, $\beta = (c^2, h, \rho)$, where a constant, c, and a k-dimensional vector, h, specify the amplitude and the direction of changes, and ρ denotes the correlation coefficient. See CHP (2007) for more detailed conditions.

Using the second Bartlett identity and information contained in the autocorrelation of η_t, CHP (2007) proposes the test statistic as follows:

$$
TS_T(\beta) = \Gamma_T(\beta) - \frac{1}{2T}\widehat{\varepsilon}(\beta)'\widehat{\varepsilon}(\beta)
$$

such that

$$\begin{aligned}\Gamma_T(\beta) &= \frac{1}{2\sqrt{T}}\sum_t c^2 h'\left[\left(l_t^{(2)}\left(\widehat{\theta}\right)+l_t^{(1)}\left(\widehat{\theta}\right)l_t^{(1)\prime}\left(\widehat{\theta}\right)\right)+2\sum_{s<t}\left(l_t^{(1)}\left(\widehat{\theta}\right)l_s^{(1)\prime}\left(\widehat{\theta}\right)\rho^{(t-s)}\right)\right]h \\ &\equiv \frac{1}{2\sqrt{T}}\sum_t \mu_{2,t}(\beta,\hat{\theta}). \qquad (1)\end{aligned}$$

where $\hat{\theta}$ denotes the maximum likelihood estimator of θ under the null, and $l_t^{(i)}$ denotes the i-th order derivative of l_t. A vector, $\widehat{\varepsilon}(\beta)$, is the residual from the OLS regression of $\frac{1}{2}\mu_{2,t}(\beta,\hat{\theta})$ on $l_t^{(1)}(\hat{\theta})$. To deal with the vector of nuisance parameters, CHP (2007) suggests applying sup-type tests as proposed by Davies (1987) or exponential-type tests with some prior distribution $J(\beta)$ over compact support $\bar{B}$ as in Andrews and Ploberger (1994):

$$\begin{aligned}\text{supTS} &= \sup_{\beta\in\bar{B}} TS_T(\beta) \\ \text{expTS} &= \int_{\bar{B}} \exp\left(TS_T(\beta)\right)dJ(\beta).\end{aligned}$$

CHP (2007) shows that the CHP test is admissible with the local alternatives of the order $T^{1/4}$.

From here, I relax the exogeneity assumption and consider a generalized alternative where the stochastic parameter depends on its past variables as well as an exogenous variable. Specifically, applying the multiplicative separability between exogenous and endogenous variables, I consider a class of alternatives in the following form:

$$\eta_t = c\cdot h\cdot\left(x_t\eta(\mathbf{y}_{t-d})\right) \qquad (2)$$

where $\mathbf{y}_{t-d} = (y_{t-1}, y_{t-2}, \cdots, y_{t-d})'$ is a vector of the past variables, x_t is an exogenous random variable, and $\eta(\cdot)$ is a mapping from $\mathbf{R}^d$ to $\mathbf{R}^1$. Nuisance parameters c and h reflect the amplitude of the change and the direction of the alternative, respectively. Note that this generalized alternative is a class of alternatives and includes various nonlinear models. For example, η_t becomes the markov chain alternative in CHP (2007) if x_t is a scalar markov chain and $\eta(\mathbf{y}_{t-d})$ is a constant function. It can be a structural change model if $x_t = 1(t > t_o)$ where $1(\cdot)$ is an indicator function. The threshold model is also an example of the previously detailed specification where $\eta(\mathbf{y}_{t-d}) = 1(y_{t-d} > c)$. In addition to these well known examples, this generalized alternative includes various nonlinear models depending on the possible combinations of x_t and $\eta(\cdot)$. Therefore, the CHP test would be more useful if it could have nontrivial power for this general class of alternatives.

I next characterize the local power of the CHP test under the generalized alternative. I consider local alternatives of order $T^{1/2}$:

$$\begin{aligned}H_0 &: \theta_t = \theta_0 \\ H_{1T} &: \theta_t = \theta_0 + \frac{1}{\sqrt{T}}\eta_t.\end{aligned}$$

where η_t now follows the form in (2). To investigate, I look at the limiting distributions of the test statistic under H_0 and H_{1T}. Lemma 4.1 in CHP (2007) implies that, under H_0,

$$TS_T(\beta) \overset{P_{\theta_0}}{\rightsquigarrow} G(\beta)$$

where $\overset{P}{\rightsquigarrow}$ denotes the weak convergence of stochastic processes under the probability measure P, and $G(\beta)$ is a Gaussian process over β with the following mean and covariance function:

$$\begin{aligned} E(G(\beta)) &= -\frac{1}{2}E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta,\theta_0)}{2} - d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right) \\ Cov(G(\beta_1), G(\beta_2)) &= E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta_1,\theta_0)}{2} - d(\beta_1)' l_t^{(1)}(\theta_0)\right)\left(\frac{\mu_{2,t}(\beta_2,\theta_0)}{2} - d(\beta_2)' l_t^{(1)}(\theta_0)\right)\right) \\ &\equiv k(\beta_1, \beta_2) \end{aligned}$$

where $d(\beta) = (I(\theta_0))^{-1} \cdot cov\Big(\frac{1}{2}\mu_{2,t}(\beta,\theta_0), l_t^{(1)}(\beta,\theta_0)\Big)$ and $I(\theta_0)$ is the information matrix. For notational simplicity, I denote $k(\beta_1,\beta_2)$ for $Cov(G(\beta_1), G(\beta_2))$. Note that $-1/2k(\beta,\beta) = E(G(\beta))$.

To derive the limiting distribution under H_{1T}, I first look at the linear expansion of the log density ratio:

$$\log \frac{dP_{\theta_1,T}}{dP_{\theta_0}} - \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t' l_t^{(1)}(\theta_0) - \frac{1}{2}E_{\theta_0}\left(\eta_t' l_t^{(1)}(\theta_0)\right)^2\right) = o_p(1)$$

The central limit theorem and the contiguity property of $P_{\theta_1,T}$ and P_{θ_0} imply that, for any finite dimensional β,

$$\left(TS_T(\beta), \log\frac{dP_{\theta_1,T}}{dP_{\theta_0}}\right) \overset{P_{\theta_0}}{\rightsquigarrow} N\left(\begin{pmatrix} -\frac{1}{2}k(\beta,\beta) \\ -\frac{1}{2}E_{\theta_0}\left(\eta_t' l_t^{(1)}(\theta_0)\right)^2 \end{pmatrix}, \begin{pmatrix} k(\beta_1,\beta_2) & \tau(\beta) \\ \tau(\beta) & E_{\theta_0}\left(\eta_t' l_t^{(1)}(\theta_0)\right)^2 \end{pmatrix}\right)$$

where $\tau(\beta) \equiv Cov\Big(TS_T(\beta,\hat{\theta}), \log\frac{dP_{\theta_1,T}}{dP_{\theta_0}}\Big)$. Applying Le Cam's third lemma,

$$TS_T(\beta) \overset{P_{\theta_1,T}}{\rightsquigarrow} N\left(-\frac{1}{2}k(\beta,\beta) + \tau(\beta), k(\beta_1,\beta_2)\right).$$

Under Assumptions 1-4 given in CHP (2007), it can be shown that the class of functions, $\mathcal{F} = \left\{\frac{\mu_{2,t}(\beta,\theta)}{2} : (\beta,\theta) \in \mathcal{B}\otimes\Theta\right\}$, is a Donsker class. Therefore, the uniform result below follows from Theorem 3.10.12 in van der Vaart and Wellner (1996)[1]

$$TS_T(\beta) \overset{P_{\theta_1,T}}{\rightsquigarrow} G\Big(-\frac{1}{2}k(\beta,\beta) + \tau(\beta), k(\beta_1,\beta_2)\Big).$$

Based on this result, I can conclude that the test statistic has nontrivial power under the local alternatives if there is a correlation between $TS_T(\beta)$ and $\log\frac{dP_{\theta_1,T}}{dP_{\theta_0}}$, i.e., $\tau(\beta) \neq 0$. I

[1] See also Appendix A in Abadie (2002). He used the similar method in a different context.

characterize the property of $\tau(\beta)$ in the following proposition whose proof can be found in the appendix.

Proposition 1 *Under the generalized alternatives in (2) and Assumptions 1-4 in CHP (2007), the CHP test does not have any local power if and only if one of the following conditions is satisfied*
(1) $E_{\theta_0}(x_t) = 0$
(2) $E_{\theta_0}\left(h'\left(\sum_{s<t} l_t^{(1)}(\theta_0)\, l_s^{(1)\prime}(\theta_0)\, \rho^{(t-s)}\right) h \cdot l_t^{(1)\prime}(\theta_0)\, h\eta(\mathbf{y}_{t-d})\right) = 0$

The first condition implies that it does not have any power if the exogenous part has a mean of zero. The second condition implies that the CHP test does not have any local power when the score function is orthogonal to the space generated by the endogenous variables. Since the score function has a zero mean under the null, when the local alternative has the order of $T^{-1/2}$, the CHP test does not have any power for structural change models where $\eta(\cdot)$ is a constant. However, for any cases that $\eta(\mathbf{y}_{t-d})$ is not a constant and has some variation, it still has power and can detect the nonlinear property of the data series. Here is a simple example of the Threshold model:

$$y_t = \alpha + \alpha^* 1\{y_{t-1} \leq r^*\} + u_t$$

where $u_t \sim i.i.d.\ N(0, \sigma^2)$ for known σ^2. Then, a reparameterization of $\theta_0 = \alpha$ and $\eta_t = \alpha^* I\{y_{t-d} \leq r^*\}$ for $\alpha^* \neq 0$, shows that

$$\begin{aligned} E_{\theta_0}\left(l_{t-1}^{(1)}(\theta_0) \cdot \eta(y_{t-1})\right) &= E_{\theta_0}\left(\frac{y_{t-1} - \alpha}{\sigma^2} \cdot \alpha^* 1\{y_{t-1} \leq r^*\}\right) \\ &= \frac{\alpha^*}{\sigma} E_{\theta_0}\left(\frac{y_{t-1} - \alpha}{\sigma} 1\{\frac{y_{t-1} - \alpha}{\sigma} \leq \frac{r^* - \alpha}{\sigma}\}\right) \\ &= \frac{\alpha^*}{\sigma} \int_{-\infty}^{\frac{r^*-\alpha}{\sigma}} s\phi(s)\, ds \end{aligned}$$

where ϕ is the pdf of the standard normal distribution. Therefore, I can conclude that the CHP test has local power for any $\alpha^* \neq 0$ and $|r^*| < \infty$.

This result reveals that the CHP test can be applied as a pretest for the nonlinear model selection problem. Since it has a nontrivial power for various popular threshold models such as Threshold Autoregressive (TAR), Self Exciting Threshold Autoregressive (SETAR), and Smooth Threshold Autoregressive (STAR) to name a few, one can exclude both markov switching and a class of threshold models when the CHP test cannot reject the null of a linear model.

I conclude this note by suggesting some potential related research. It has been shown in this note that the CHP test does not have any local power when the endogenous stochastic parameter is orthogonal to the score function. Developing an optimal test for a more generalized parameter stability problem would be ideal. The optimal property of the integrated conditional moment test found in Bierens and Ploberger (1997) may have potential, but I leave this generalization for future research.

A Appendix

Proof of Proposition 1: For notational simplicity, I suppress the dependency of functions $\mu_{2,t}(\beta,\theta)$ and $l_t^{(i)}(\theta)$ on parameters, and denote $\mu_{2,t}$ and $l_t^{(i)}$, respectively. I expand the $\tau(\beta)$ and show that it can be written as a product of those two conditions listed in the Proposition.

$$\begin{aligned}
\tau(\beta) &= Cov\Big(TS_T(\beta), \log \frac{dP_{\theta_{1,T}}}{dP_{\theta_0}}\Big) \\
&= E_{\theta_0}\Big(\Big(\frac{1}{2}\mu_{2,t} - d' l_t^{(1)}\Big) \cdot \eta_t' l_t^{(1)}\Big) \\
&= E_{\theta_0}\left(\frac{1}{2}\mu_{2,t} \cdot l_t^{(1)\prime}\eta_t\right) - E_{\theta_0}\left(d' l_t^{(1)} \cdot l_t^{(1)\prime}\eta_t\right) \\
&= E_{\theta_0}\left(\frac{1}{2}\mu_{2,t} \cdot l_t^{(1)\prime}\eta_t\right) - E_{\theta_0}\left(\frac{1}{2}\mu_{2,t} \cdot l_t^{(1)\prime}\right) E_{\theta_0}(\eta_t) \\
&= E_{\theta_0}\left(\frac{1}{2}\mu_{2,t} \cdot l_t^{(1)\prime}\eta_t\right).
\end{aligned}$$

The fourth equality follows from the definitions of d and the information matrix, and the fifth equality holds since $E_{\theta_0}(\eta_t)$ is normalized as zero. See CHP (2007) for further explanation. I expand it further by substituting (1) for $\frac{1}{2}\mu_{2,t}$:

$$\tau(\beta) = E_{\theta_0}\left[c^2 h' \left(l_t^{(2)} + l_t^{(1)} l_t^{(1)\prime}\right) h \cdot l_t^{(1)\prime}\eta_t\right] + E_{\theta_0}\left[c^2 h' \sum_{s<t} l_t^{(1)} l_s^{(1)\prime} \rho^{(t-s)} h \cdot l_t^{(1)\prime}\eta_t\right]. \quad (3)$$

Note that η_t depends on either past variables $\mathbf{y}_{t-d}$ or an exogenous variable x_t. Thus, the first term in (3) is again zero because of the normalization, and I get the final expression by substituting (2) for η_t :

$$\begin{aligned}
\tau(\beta) &= E_{\theta_0}\left[c^2 h' \sum_{s<t} l_t^{(1)} l_s^{(1)\prime} \rho^{(t-s)} h \cdot l_t^{(1)\prime}\eta_t\right] \\
&= c^3 E_{\theta_0}\left[h' \left(\sum_{s<t} l_t^{(1)} l_s^{(1)\prime} \rho^{(t-s)}\right) h \cdot l_t^{(1)\prime} h\eta(\mathbf{y}_{t-d})\right] E_{\theta_0}(x_t),
\end{aligned}$$

which is equal to zero uniformly over β if either term of the two expectations is zero. ■

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97(457), 284–292.

Andrews, D., and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Bierens, H., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Carrasco, M. (2002): "Misspecified Structural Change, Threshold, and Markov-switching models," *Journal of Econometrics*, 109, 239–273.

Carrasco, M., L. Hu, and W. Ploberger (2007): "Optimal Test for Markov Switching," manuscript.

Davies, R. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33–43.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer, New York.