Do Markov-Switching Models Capture Nonlinearities in the Data? Tests using Nonparametric Methods

R. Breunig[a] **and A.R. Pagan**[b]

[a]*Centre for Economic Policy Research, Research School of Social Sciences and School of Economics, The Australian National University, Canberra ACT 0200 Australia (Robert.Breunig@anu.edu.au)*
and
[b]*Research School of Social Sciences, The Australian National University, Canberra, Australia and Oxford University, Oxford, United Kingdom (arpagan@coombs.anu.edu.au)*

Abstract: Markov-switching models have become popular alternatives to linear autoregressive models. Many papers which estimate nonlinear models make little attempt to demonstrate whether the non-linearities they capture are of interest or if the models differ substantially from the linear option. By simulating the models and nonparametrically estimating functions of the simulated data, we can evaluate if and how the nonlinear and linear models differ.

Keywords: Markov-switching models; Nonparametric estimation; Simulation methods

1. INTRODUCTION

The purpose of this paper is to provide an example of how simulation methods, in combination with non-parametric density and regression estimation, may be used to evaluate the success of non-linear models in capturing (or failing to capture) important characteristics of the data. As an example, we look at Markov-switching (MS) models, popularized in the economics literature by Hamilton [1989]. MS models have become a popular alternative to linear autoregressive models in part due to their intuitive appeal, in that they allow a latent state process to cause `switching' between different linear regimes. This corresponds to the popular conception of underlying states of the economy, for example high-growth or low-growth states in GDP and high-volatility and low-volatility states in the stock market.

We consider a simple Markov-switching model estimated by Bodman [1998] using Australian unemployment data as an example of these techniques. Harding and Pagan [1998] provide an early example of the use of simulation and nonparametric techniques applied to MS models while Pagan [1999] and Clements and Galvao [2000] apply nonparametric conditional mean estimation to a selection of non-linear models. The idea of comparing nonparametric estimates of densities and models goes back to the late 1980s at least, and was used by Pagan [1996] and Ait-Sahalia [1996] in analyzing financial data.

The example in this paper is illustrative of the problems which often arise in estimating and using Markov-switching models. Breunig and Pagan [2001] explore numerous examples of MS models which fail to converge, which fail to accurately capture important features of the data, and which appear indistinguishable from simpler alternatives.

2. MARKOV-SWITCHING MODELS

Consider the following example of a Markov-switching model which describes the movement of some stationary economic variable x_t as a function of its p most recent lagged values and a state variable, s_t, which takes values 0 or 1

$$x_t = \mu_0(1-s_t) + \mu_1 s_t + \sum_{l=1}^{p} \phi_l \left(x_t - \mu_0(1-s_t) - \mu_1 s_t\right) + \sigma v_t \tag{1}$$

s_t evolves as a first-order Markov process with probability P_{00} of remaining in state 0 next period conditional on being in state 0 this period while P_{11} is defined analogously. Extensions of this simple model have been proposed: using more than two states; allowing dependence on the state of the error variance, σ, and the autoregressive parameters, ϕ_l; allowing the state variable to evolve as a higher-order Markov process; and allowing the transition probabilities to depend upon the duration of time spent in a given state. The model (1) may be estimated by Maximum Likelihood using the filter proposed by Hamilton [1989]. If $\mu_0 = \mu_1$, the model collapses back to a linear autoregressive model.

It should be incumbent upon researchers who propose the non-linear alternative to show (a) that the linear model, i.e. the case where $\mu_0 = \mu_1$, can be rejected; and (b) that the non-linear model captures important features of the data better than the linear model. In this paper, we use informal tests to address the second of these two issues. Garcia [1998] and Hansen [1992] have provided tests of the MS model against a one-state null hypothesis. These tests are difficult to implement for models which are more complicated than (1) and, at least to this point, have not been widely used in the literature. Breunig and Pagan [2001] propose a combination of moment based tests and informal tests akin to the ones discussed here and show how they can be used to test for convergence problems and general misspecification in MS models.

3. INFORMAL TESTS

The central idea of this paper is to simulate from the proposed model taking the parameter estimates, $\hat{\theta}$, as given. Using that simulated data, we can use nonparametric techniques to compute quantities such as $f_M\left(x_t \mid \hat{\theta}\right)$, the implied distribution of x_t from the model, and $E_M\left(x_t \mid x_{t-j};\hat{\theta}\right)$, the implied conditional mean from the model, and compare these to the respective quantities from the data. In order to minimize the degree of error in estimating these functions, we take a very large number of simulated draws (60,000 in the application below.) Comparisons are made using simple graphs, providing a very powerful tool for quickly identifying problems with the models and the ways in which the non-linear and linear models differ. The visual check of the 'fit' between the data and the implied quantity from the model gives the informal test.

The intended use of the model should inform the choice of functions to be examined. In the empirical example below we consider two aspects of the data: the empirical distribution of changes in the unemployment rate and the conditional mean of the unemployment rate changes on their one-period lagged values. Density comparisons such as the former will be particularly important where one would like to compare MS and alternative models to determine which better replicates the empirical density function (as in models of financial data where a criticism of commonly used models such as GARCH is their failure to capture the concentration of data around zero). The latter choice will be important for forecasting, frequently the main interest in univariate models.

3.1 Nonparametric Estimation

Once we have generated a large number of simulated draws, n, from the model, we estimate the density using the non-parametric kernel density estimate introduced by Rosenblatt [1956] and Parzen [1962]

$$\hat{f}_M(x) = \frac{1}{nh}\sum_{s=1}^{n} K\left(\frac{x_s - x}{h}\right) \qquad (2)$$

where K( ) is a smooth kernel function and *h* is a window width which controls the smoothness of the estimate. Pagan and Ullah [1999] provide a general survey of nonparametric estimation, including discussion of choice of kernel and window width for a wide range of problems. We estimate the density using a normal kernel at each point in the data set for both the simulated series and the actual data. This allows us to create a graphical comparison of the distribution of the data and the implied distribution from the model. In the density estimates presented below, we choose the bandwidth, h, for the actual data by leave-one-out cross-validation whereas for the simulated data, we use $h = 1.06\sigma n^{-1/5}$ as a simple plug-in bandwidth. With 60000 observations, the technique is not sensitive to bandwidth selection.

Making further use of the simulated values, we estimate the conditional mean on values lagged one period using Nadaraya-Watson kernel regression

$$\hat{E}_M(x_t \mid x_{t-1};\hat{\theta}) = \frac{\sum_{s=2}^{n} x_s K\left(\frac{x_{s-1} - x}{h}\right)}{\sum_{s=2}^{n} K\left(\frac{x_{s-1} - x}{h}\right)} \qquad (3)$$

We use a normal kernel and the robust bandwidth suggested by Silverman [1986]: $h = .9 * \min(IQR/1.34, \sigma)$ where IQR is the inter-quartile range. We can again use graphical techniques, this time comparing the conditional mean implied by the model to a cross plot of the data against its lagged values.

4. APPLICATION

Bodman [1998] fitted the two state MS model (1) to the first differences of the quarterly Australian unemployment rate x_t for the time period 1959:3 to 1997:3. The second column of Table 1 provides his estimates of the parameters.

Using these parameter estimates, we generated 63000 observations from the implied model. We drop the first 3000 observations to remove the effect of starting values. Consequently, estimates of $f_M(x_t \mid \hat{\theta})$ and $E_M(x_t \mid x_{t-j}; \hat{\theta})$ are based upon sample sizes of 60,000. Figure 1 presents the nonparametric density estimates from the data, the linear AR(4) model, and the MS model.

Clearly the linear Gaussian model fails to capture certain key aspects of the data including the high peak just below zero and the rapid decline in the density between zero and one-half. However, the degree to which the MS model fails to capture the distribution of the data is shocking. Here we are making a simple informal visual comparison. Since the densities are so clearly different, no formal test seems necessary. However, it is possible to derive formal tests, as done by Ait-Sahalia [1996]. Breunig and Pagan [2001] construct formal tests based upon testing moments of the data against the simulated moments of the model. They show that failure to match the simple moments such as the mean and variance may be used as a check of model convergence.

Table 1. Comparison of Bodman and Alternative Estimates for an MS model.

	Bodman estimate	Revised estimate
μ_0	.486	.6952
μ_1	-.361	-.0399
ϕ_1	.022	.4174
ϕ_2	.147	.1163
ϕ_0	-.079	.1664
ϕ_0	-.185	-.3273
σ	.326	.2148
P_{00}	.83	.48
P_{11}	.89	.927

Many researchers, e.g. Goodwin [1993, p.332] and Hamilton's web page, report convergence problems with these models. To investigate the possibility that this failure to match the data arose from convergence problems we re-estimated the model using the data provided to us by Bodman and a variety of programs, including that on Hamilton's web site. The revised parameter estimates are given in the third column of Table 1. Figure 2 shows that the model based upon these revised estimates implies a distribution much closer to the data density and superior to the linear model. Hence it is clear that any conclusions drawn from Bodman's original model must be treated with caution due to the failure to converge.

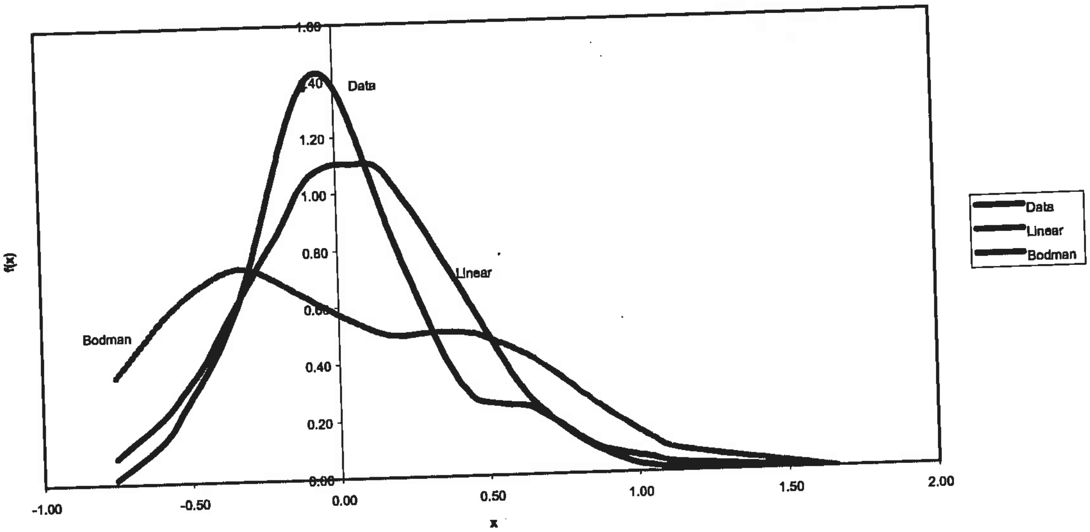

Figure 1. Density implied by Bodman MS Model and AR(4).

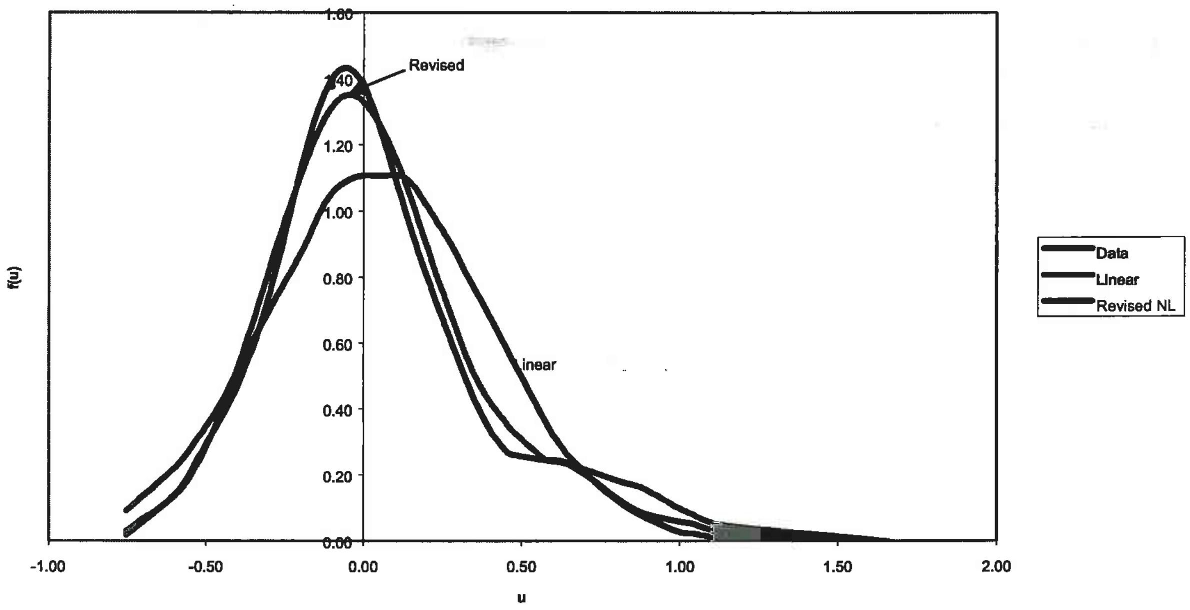

Figure 2. Density estimates for Revised MS Model.

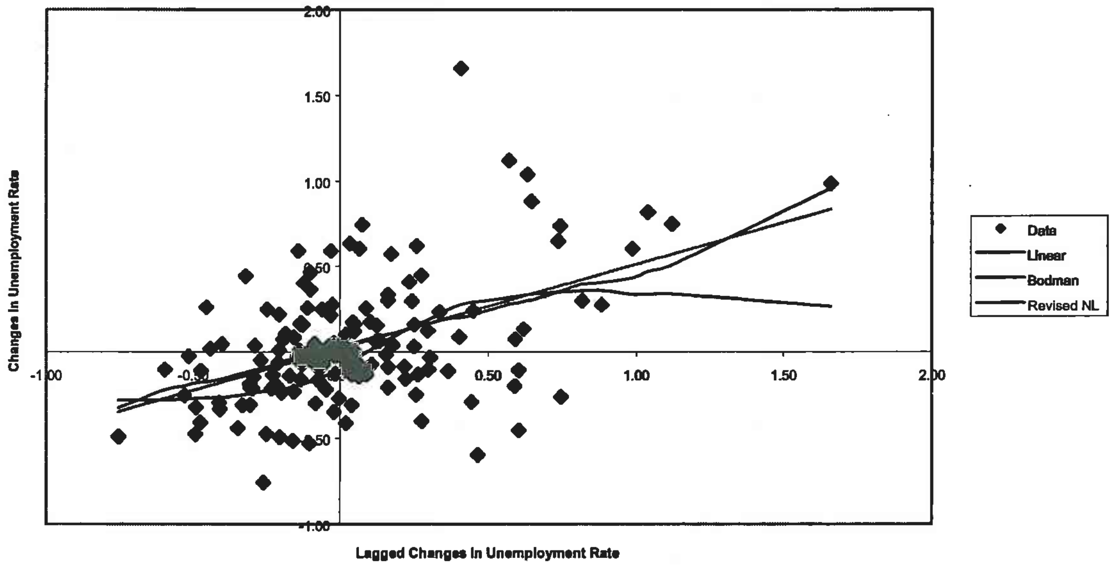

Figure 3. Change in Australian Unemployment Rate vs. Unemployment Rate Lagged.

Figure 3 presents the cross-plot of the data against the conditional means implied by the linear model, Bodman's original estimates, and our revised estimates. Again, the obvious problems with Bodman's model are apparent in that the conditional mean is substantially different than the data. When we simulate Bodman's original model, we observe a negative relationship between unemployment changes and their lagged values above values of .8. This negative relationship is not supported by the cross plot.

It is also clear that there is very little difference between the conditional expectations found from the revised MS model and a linear model, at least in that part of the variable space where there are a reasonable number of data points. Thus, for an exercise such as one-step ahead forecasting, where it is the conditional mean that is used, we would not expect any gains from using the MS model. Of course, estimating the MS model is more complicated and involves estimating three additional parameters. The gain from this extra complexity seems rather small.

In this application, whether the nonlinear model withstands scrutiny depends upon one's stance regarding the importance of the density estimate relative to the conditional mean. For problems where the central concern is matching the unconditional distribution of the data, the MS model would appear superior. For conditional mean modelling, however, the more parsimonious linear model has an equal claim to the MS model.

It is worth noting that the comparison here is based upon a linear model with normal errors. This is the assumption behind the standard tests of the number of states in MS models which are effectively testing whether the densities from the MS model and the linear model are the same. These tests may lead one to accept the MS model, even in cases where neither is a good representation of the data. In this example, the densities from the MS model appear to be a good match with the data. If forecasting is the objective, as it most likely is with quarterly unemployment data, and the choice is between the linear model and the MS model, then the linear model would appear to be preferable on grounds of simplicity.

5. CONCLUSIONS

The techniques outlined here have the advantage of being easy to apply, even for very complicated models. Clearly the choice of which conditional moments to examine and which densities are important depend upon the particular application. In the example presented here, the techniques uncovered a convergence problem with the algorithm that was used to estimate the model. Even after correcting the estimates, there is little evidence that the non-linear model performs better than the linear alternative in ways that would be important for most economic applications. Breunig and Pagan [2001] show that convergence problems and a wide range of specification problems can be uncovered using these techniques. It should be required that authors submit some conditional moment estimates to justify the use of non-linear models.

6. REFERENCES

Ait-Sahalia, Y., Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies,* 9, 385-426, 1996.

Bodman, P., Asymmetry and Duration Dependence in Australian GDP and Unemployment, *Economic Record*, 74, 399-411, 1998.

Breunig, R. and A.R. Pagan, Some Simple Methods for Assessing Markov Switching Models, mimeo, Australian National University (available at http://econrsss.anu .edu.au/~breunig/), 2001.

Clements, M.P. and A.B.C. Galvao, Conditional Mean Functions of Non-Linear Models of US Output, mimeo, University of Warwick, 2000.

Garcia, R., Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, *International Economic Review*, 39(3), 763-788, 1998

Goodwin. T.H., Business-Cycle Analysis with a Markov-Switching Model, *Journal of Business and Economic Statistics*, 11, 231-339, 1993.

Hamilton, J.D., A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357-384, 1989.

Hansen, B., The Likelihood Ratio Test Under Non-standard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics*, 7, S61-S82, 1992.

Harding. D. and A.R. Pagan, Dissecting the Cycle, *Journal of Monetary Economics*, forthcoming, 1998.

Pagan, A.R., The Econometrics of Financial Markets, *Journal of Empirical Finance*, 3, 15-102, 1996.

Pagan, A.R., The Getting of Macroeconomic Wisdom, Invited Address to the World Economic Congress of the International Economic Association in Beunos Aires (available at http://econrsss.anu.edu.au/ staff/adrian/index.html), 1999.

Pagan, A.R. and A. Ullah, *Nonparametric Econometrics*, Cambridge University Press, New York, 1999.

Parzen, E., On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics,* 33, 1065-1076, 1962.

Rosenblatt, M., Remarks on Some Nonparametric Estimates of Density Function, *Annals of Mathematical Statistics,* 27, 832-837, 1956.

Silverman, B., *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, London, 1986.