Chapter 5
Testing for a Markov-Switching Mean in Serially Correlated Data

James Morley and Zohra Rabah

Abstract When testing for Markov switching in mean or intercept of an autoregressive process, it is important to allow for serial correlation under the null hypothesis of linearity. Otherwise, a rejection of linearity could merely reflect misspecification of the persistence properties of the data, rather than any inherent nonlinearity. However, Monte Carlo analysis reveals that the Carrasco, Hu, and Ploberger (Optimal test for Markov Switching parameters, conditionally accepted at Econometrica, 2012) test for Markov switching has low power for empirically relevant data-generating processes when allowing for serial correlation under the null. By contrast, a parametric bootstrap likelihood ratio test of Markov switching has higher power in the same setting. Correspondingly, the bootstrap likelihood ratio test provides stronger support for a Markov-switching mean in an application to an autoregressive model of quarterly US real GDP growth.

Keywords Nonlinearity tests • Autoregressive processes • Markov switching • Parametric bootstrap • Real GDP dynamics

5.1 Introduction

Many macroeconomic time series such as the quarterly growth rate of US real GDP display positive serial correlation. An important question is what role nonlinear dynamics play in generating this persistence in the data. For example, estimates

J. Morley (✉)
School of Economics, Australian School of Business, University of New South Wales, Sydney, NSW 2052, Australia
e-mail: james.morley@unsw.edu.au

Z. Rabah
University of Lorraine, BETA, 6 rue des Michottes, 54000 Nancy Cedex, France
e-mail: zohra.rabah@univ-lorraine.fr

J. Ma and M. Wohar (eds.), *Recent Advances in Estimating Nonlinear Models: With Applications in Economics and Finance*, DOI 10.1007/978-1-4614-8060-0_5,

for the Markov-switching model of Hamilton (1989) suggest that much of the persistence in output growth is due to discrete changes in the mean growth rate that correspond to expansion and recession phases of the business cycle, rather than gradual changes in the conditional mean according to a simple linear time series process.

Formal tests of Markov switching are hindered by the presence of nuisance parameters that are only identified under the alternative hypothesis and by the problem of identically zero scores at the null hypothesis (see Hansen, 1992). However, Carrasco, Hu, and Ploberger (2012) (CHP hereafter) have recently proposed an information-matrix-based test that addresses these problems. The CHP test has been applied to macroeconomic data by Hamilton (2005) and Morley and Piger (2012), amongst others. Meanwhile, Kim, Morley, and Piger (2005) and Morley and Piger (2012) have considered a parametric bootstrap likelihood ratio (BLR) test of Markov switching. Di Sanzo (2009) conducts Monte Carlo analysis of tests for Markov switching and finds that a related semi-parametric BLR test has much higher power than the CHP test for some basic data-generating processes (DGPs), although the full theoretical justification for the BLR test remains to be established.

In this paper, we argue that, regardless of which test for Markov switching is being applied, it is crucial to allow for serial correlation under the null hypothesis of linearity when considering the alternative hypothesis of a Markov-switching mean for a stationary time series process. In particular, a Markov-switching mean implies autoregressive dynamics, albeit with highly nonstandard errors. Thus, any apparent evidence for Markov switching may just reflect the ability of the model under the alternative hypothesis to proxy for serial correlation omitted from the model under the null hypothesis of linearity. Specifically, tests for nonlinearity are based on a composite null hypothesis of linear dynamics and a particular model specification. To the extent that a Markov-switching mean is even considered in the first place, there is likely to be serial correlation in the data. Therefore, it is important to allow for serial correlation under the null hypothesis.

The point about the importance of allowing for serial correlation under the null hypothesis when testing for nonlinearity is straightforward enough. But it has the surprising and notable consequence that the CHP test discussed above can have very low power to detect nonlinearity in empirically relevant settings. Similar to Di Sanzo (2009), we employ Monte Carlo analysis to consider the small sample properties of the CHP test and find that it does not perform well in detecting Markov switching in mean or intercept of an autoregressive process when allowing for serial correlation under the null hypothesis. By contrast, the parametric BLR test considered in Kim, Morley, and Piger (2005) and Morley and Piger (2012) retains considerable power in this setting. We use these results to explain some previous findings in the empirical literature. Then we apply the tests to quarterly US real GDP growth and, corresponding to the Monte Carlo analysis, we find stronger support for nonlinearity based on the BLR test than the CHP test when taking serial correlation into account under the linear null hypothesis.

5.2 Markov-Switching Models in Mean

In this section, we discuss two variants of Markov-switching models. The first specification is the popular model of Hamilton (1989), denoted as MSM-AR, which allows the mean of an autoregressive process to switch between regimes. The second model is denoted as MSI-AR, which allows the intercept of an autoregressive process to switch between regimes. See Krolzig (1997) for a full discussion of this subclass of models. For simplicity, we focus our discussion on the particular case of specifications with two regimes and an autoregressive order of one.

5.2.1 The MSM(2)-AR(1) Model

Let x_t denote a stationary time series such as the first difference of log real GDP. An MSM(2)-AR(1) model is given by

$$x_t = \mu_t + \phi(x_{t-1} - \mu_{t-1}) + e_t, \tag{5.1}$$

where $e_t \tilde{} NID(0, \sigma_e^2)$. The time-varying mean μ_t is allowed to switch between regimes as follows:

$$\mu_t = \gamma_0 + \gamma_1 S_t \tag{5.2}$$

with $S_t = \{0, 1\}$ denoting a two-state Markov-switching state variable with constant transition probabilities Pr $[S_t = 0|S_{t-1} = 0] = p_{00}$ and Pr $[S_t = 1|S_{t-1} = 1] = p_{11}$. The regime-dependent mean of x_t is γ_0 if $S_t = 0$ and $\gamma_0 + \gamma_1$ otherwise.

Now, let ξ_t denote the zero-mean Markov state variable such that $\xi \equiv S_t - \pi$ where $\pi \equiv E(S_t) = (1 - p_{00})/(2 - p_{00} - p_{11})$ is the unconditional probability of being in the $S_t = 1$ regime. The dynamics of the Markov chain can be expressed as

$$\xi_t = \rho\xi_{t-1} + v_t, \tag{5.3}$$

where $\rho = p_{00} + p_{11} - 1$ represents the persistence parameter and v_t follows a martingale difference sequence. Consistent with positive persistence in regimes such as expansions and recessions, we assume $\rho > 0$.

Given (5.1)–(5.3), the MSM-AR model can be rewritten as the sum of two independent processes as follows:

$$x_t - \mu = \gamma_1\xi_t + z_t, \tag{5.4}$$

where μ is the unconditional mean of x_t such that $\mu = \gamma_0 + \gamma_1\pi$. While the process $z_t = \phi z_{t-1} + e_t$ is Gaussian, the term $\gamma_1\xi_t$ reflects the contribution of the Markov process. The variances of the two terms depend on σ_e^2 and γ_1^2, respectively.

The first thing to notice from (5.4) is that, even if the Gaussian process z_t displays no persistence (i.e., $\phi = 0$) or its variance is trivially small (i.e., $\sigma_e^2 / \gamma_1^2 \rightarrow 0$), the time series x_t will still display serial correlation due to the Markov process ξ_t. Thus, a necessary condition for even considering a Markov-switching mean should be that a time series displays some serial correlation. Meanwhile, given that a time series displays serial correlation, it is crucial to allow for serial correlation under the null hypothesis of linearity when testing for the presence of a Markov-switching mean. Otherwise, any apparent evidence of nonlinearity may simply reflect omitted linear dynamics under the null hypothesis.

5.2.2 The MSI(2)-AR(1) Model

Next, we consider a first-order autoregressive process with a switching intercept. The MSI(2)-AR(1) is given as follows:

$$x_t = \mu_t + \phi x_{t-1} + e_t. \tag{5.5}$$

To simplify notation, we use the same function μ_t for the switching intercept term as was used for the switching mean in the previous subsection.

Similar to before, the dynamics of an MSI(2)-AR(1) can be rewritten as follows:

$$x_t - \mu = \gamma_1 \xi_t + \phi(x_{t-1} - \mu) + e_t, \tag{5.6}$$

where μ, the unconditional mean of x_t, is $(\gamma_0 + \gamma_1 \pi)(1 - \phi)^{-1}$.

Again, even if $\phi = 0$ or $\sigma_e^2 / \gamma_1^2 \rightarrow 0$, the time series x_t will display first-order serial correlation due to the Markov process ξ_t. So, as before, it is crucial to allow for serial correlation under the null hypothesis of linearity when testing for a Markov-switching intercept.

5.2.3 ARMA Representations and Forecasts

As shown by Krolzig (1997), the subclass of Markov-switching autoregressive models given above can be represented as ARMA processes. Consider again the MSM(2)-AR(1) model, temporarily setting the unconditional mean to zero for simplicity and using lag notation:

$$(1 - \phi L)x_t = (1 - \phi L)(1 - \rho L)^{-1} \gamma_1 v_t + e_t. \tag{5.7}$$

Multiplying both sides by $(1 - \rho L)$ gives the following:

$$(1 - \rho L)(1 - \phi L)x_t = (1 - \phi L)\gamma_1 v_t + (1 - \rho L)e_t, \tag{5.8}$$

which is an ARMA(2,1) process with highly nonstandard errors as long as $\phi \neq \rho$. Otherwise, if $\phi = \rho$, it is just an AR(1) process with highly nonstandard errors:

$$(1 - \rho L)x_t = \gamma_1 v_t + e_t. \tag{5.9}$$

Meanwhile, consider the MSI(2)-AR(1) model in lag notation:

$$(1 - \phi L)x_t = \gamma_1 (1 - \rho L)^{-1} v_t + e_t. \tag{5.10}$$

Multiplying both sides with $(1 - \rho L)$ gives the following:

$$(1 - \rho L)(1 - \phi L)x_t = \gamma_1 v_t + (1 - \rho L)e_t, \tag{5.11}$$

which is an ARMA(2,1) process with highly nonstandard errors as long as $\rho \neq 0$. Otherwise, if $\rho = 0$, it is an AR(1) process with highly nonstandard errors:

$$(1 - \phi L)x_t = \gamma_1 v_t + e_t. \tag{5.12}$$

What this analysis reveals is that part of the reason why standard Markov-switching models are different from a linear AR(1) model is that they have more complicated autocorrelation structures. As shown by Krolzig (1997), the autocorrelation function (ACF) of an MSM(2)-AR(1) process is given by

$$\Gamma_x(h) = \rho^h \gamma_1 \pi (1 - \pi) + \phi^h (1 - \phi^2)^{-1} \sigma_e^2 \tag{5.13}$$

for $h \geq 0$ and $\Gamma_x(h) = \Gamma_x(-h)$ for $h < 0$. If $\phi \neq 0$, $\rho \neq 0$, and $\phi \neq \rho$, (5.13) corresponds to the ACF of an ARMA(2,1) process. Also, consider the ACF of order one of an MSI(2)-AR(1) model:

$$\Gamma_x(1) = \frac{\phi + \rho}{1 + \phi\rho}\Gamma_x(0) - \frac{\rho}{1 + \phi\rho}\sigma_e^2, \tag{5.14}$$

with $\Gamma_x(0) = \frac{1}{1-\phi^2}\left[\pi(1-\pi) + \sigma_e^2\right]$. For $h > 1$, the ACF can be calculated recursively as follows:

$$\Gamma_x(h) = (\phi + \rho)\Gamma_x(h-1) - \phi\rho\Gamma_x(h-2). \tag{5.15}$$

These ACFs are clearly more complicated than for an AR(1) model, where

$$\Gamma_x(h) = \phi^h (1 - \phi^2)^{-1} \sigma_e^2. \tag{5.16}$$

They are also different from each other, with the ACF for the switching intercept model being the most complicated.

To be more concrete about the differences between the models, it is illustrative to consider their implied point forecasts. Following Krolzig (2000), the MSM(2)-AR(1) optimal h-step predictor is given by

$$x_{t+h|t} - \mu = \phi^h (x_t - \mu) + \gamma_1 (\rho^h - \phi^h) \xi_{t|t}, \tag{5.17}$$

where $x_{t+h|t} = E_t[x_{t+h}]$ and $\xi_{t|t} = E_t[\xi_t]$. The first term on the right-hand side of (5.17) represents the optimal linear predictor for an AR(1) model and the second one reflects the Markov chain prediction. The latter depends on both the magnitude of shift γ_1 and on the persistence of the regimes, ρ, relative to the persistence of the Gaussian process, ϕ. Of course, when $\gamma_1 = 0$ or $\rho = \phi$, the optimal prediction rule reduces to the linear predictor. Meanwhile, the MSI(2)-AR(1) optimal h-step predictor can be written as follows:

$$x_{t+h|t} - \mu = \phi^h (x_t - \mu) + \gamma_1 \left(\sum_{i=1}^{h} \rho^i \phi^{h-i} \right) \xi_{t|t}. \tag{5.18}$$

As before, the second term on the right-hand side of (5.18) reflects the nonlinearity, which depends on the magnitude of the parameter change γ_1, on the persistence parameters ϕ and ρ. It collapses to a linear predictor if there is no intercept change, $\gamma_1 = 0$, or the regimes are not persistent, $\rho = 0$. Note that, as long as the persistence of the Gaussian and Markov processes are similar, the MSI-AR model will differ more from the linear AR model than the MSM-AR model in terms of the implied point forecast.

The above analysis implies that it is important to test for a Markov-switching mean rather than simply comparing out-of-sample predictive accuracy of point forecasts for linear and nonlinear autoregressive models. Specifically, point forecasts may be quite similar for the linear and nonlinear models, with the linear models having the advantage of tighter parameterization. Indeed, several studies have reported that linear models produce better point forecasts than Markov-switching models (see, for example, Clements and Krolzig, 1998, Clements et al., 2004, Siliverstovs and van Dijk, 2003, and Ferrara, Marcellino, and Mogliani 2012). Also, studies such as Dacco and Satchell (1999) and Teräsvirta (2006) show that, even when nonlinear models such as those with Markov-switching parameters are consistent with the DGP, a linear model can produce more accurate point forecasts out of sample.

Density forecasts will differ more than point forecasts given the highly non-standard distributional assumption about the errors in the ARMA representation of the Markov-switching model versus the standard assumption underlying the nested linear model. Indeed, this difference provides motivation for why we would care about testing for nonlinearity in the first place given similar point forecasts between linear and nonlinear models. Here, the literature is somewhat more supportive of nonlinearity (see, for example, Siliverstovs and van Dijk, 2003). However, out-of-sample comparisons are likely to be somewhat sensitive to the holdout sample considered, suggesting the importance of testing nonlinearity using the whole available sample. Yet, despite the importance of testing for Markov switching when considering serially correlated data, it turns out that the recently proposed CHP test can have very little power to detect nonlinearity in this setting. Fortunately, the BLR test performs much better. We turn to the issue of testing for Markov switching next.

5.3 Monte Carlo Analysis

In this section, we present the results of several Monte Carlo experiments that are designed to compare the small sample performance of the CHP test for Markov switching with the performance of the BLR test for Markov switching. To fully assess the effect of allowing for serial correlation under the null hypothesis, we also consider the case with no serial correlation for both specifications discussed in Section 5.2 (i.e., MSM(2)-AR(0) and MSI(2)-AR(0), which are equivalent given the AR(0) specification).

In all cases, we generate 1,000 artificial series of length $T + 50$ with $T \in \{100, 200, 300\}$ to assess the small sample properties of the tests. The 50 initial observations are thrown out to minimize the effect of initial conditions. The parameters we use are the maximum likelihood estimates (MLE) obtained by fitting the models to the quarterly US real GDP growth rate data (measured as 100 times the first differences of the natural logarithms of the level data) for the sample period of 1984Q2 to 2010Q3. Then, both tests are applied to the artificial series generated under the null and alternative hypotheses. For the CHP test, parametric bootstrap experiments are required to compute the critical values. Thus, $B = 1,000$ bootstrap samples are generated based on the MLE under the null hypothesis. For each bootstrap sample, the MLE are calculated under the null and the statistic is maximized numerically with respect to $\rho \in (0.02, 0.98)$ to capture persistent business cycle phases. The bootstrap critical value for a nominal size α can be found by sorting the bootstrap test statistics from smallest to largest and finding the $(1 - \alpha)B$ test statistic. For the BLR test, $B = 500$ experiments are conducted. Following Kim, Morley, and Piger (2005), we address the problems of numerous local optima and unstable estimation under the alternative in conducting simulation experiments by considering a grid of possible values for the transition probabilities. In what follows, all tests are evaluated for a nominal 5% significance level.

We start with the case where the DGP has no linear autoregressive dynamics and analyze the size and power of the tests. To evaluate the rejection rates when the null is true, the data are generated according to a linear DGP as follows:

$$\text{DGP}_0 : x_t = \mu + e_t \text{ with } \mu = 0.681 \text{ and } \sigma = 0.607.$$

Table 5.1 Rejection rates, 5% nominal level

	DGP_0			$DGP_{MS\|0}$		
	$T=100$	$T=200$	$T=300$	$T=100$	$T=200$	$T=300$
BLR test	3.8	3.6	3.5	74.2	93.9	98.2
CHP test	5.1	5.0	5.7	54.1	74.1	87.6

To investigate the power of the tests, we generate the data under the alternative hypothesis of nonlinearity. We use the following MSM(2)-AR(0) model:

$$DGP_{MS|0} : x_t = \gamma_0 + \gamma_1 S_t + e_t \text{ with } \gamma_0 = -0.744,\ \gamma_1 = 1.532,\ p_{00} = 0.698,$$
$$p_{11} = 0.978, \text{ and } \sigma = 0.465.$$

Rejection rates obtained from DGP_0 and$DGP_{MS|0}$ are reported in Table 5.1. The results suggest that the CHP test has an empirical size close to the nominal level, while the BLR test is slightly undersized for each sample size. The low rejection rates for the BLR test could be due to the coarseness of the grid search, which serves to keep the computational burden manageable.

Both tests have a good power. However, the BLR test has a slightly higher power than the CHP test for this DGP.

So far, both tests work well when the null is no serial correlation. Of course, it seems unlikely that a researcher would actually consider a Markov-switching mean if the data appear serially uncorrelated, as they would under DGP_0. At the same time, given serial correlation under $DGP_{MS|0}$, a researcher might also be hesitant to impose no serial correlation under the null hypothesis when testing for Markov switching, as the imposed null would clearly be at odds with the apparent serial correlation in the data.

Next, we consider the more realistic scenario of data with serial correlation under both the null and alternative hypotheses. We explore the size and power of the tests for both specifications already mentioned, MSM(2)-AR(1) and MSI(2)-AR(1). We evaluate empirical size when the DGP is a linear AR(1) process in demeaned form as follows:

$$DGP_{M|1} : x_t = \mu + \phi(x_{t-1} - \mu) + e_t \text{ with } \mu = 0.663,\ \phi = 0.442, \text{ and } \sigma = 0.537,$$

and when the DGP is a linear AR(1) process with an intercept term given by

$$DGP_{I|1} : x_t = \mu + \phi x_{t-1} + e_t \text{ with } \mu = 0.370,\ \phi = 0.442, \text{ and } \sigma = 0.537.$$

To compute the empirical power, we generate the data according to the MSM(2)-AR(1) process:

Table 5.2 Monte Carlo size, 5% nominal level

	$DGP_{M\|1}$			$DGP_{I\|1}$		
	$T = 100$	$T = 200$	$T = 300$	$T = 100$	$T = 200$	$T = 300$
BLR test	4.6	3.9	4.7	1.8	1.9	2.3
CHP test	5.0	2.1	1.4	6.3	5.1	3.8

Table 5.3 Monte Carlo power, 5% nominal level

	$DGP_{MSM\|1}$			$DGP_{MSI\|1}$		
	$T = 100$	$T = 200$	$T = 300$	$T = 100$	$T = 200$	$T = 300$
BLR test	65.4	90.0	97.1	56.5	83.3	94.9
CHP test	6.3	2.7	3.0	7.2	8.5	12.4

$$\begin{aligned} \text{DGP}_{\text{MSM}|1} : x_t &= \gamma_0 + \gamma_1 S_t + \phi(x_{t-1} - \gamma_0 - \gamma_1 S_{t-1}) + e_t \\ &\text{with } \gamma_0 = -0.737,\ \gamma_1 = 1.507,\ \phi = 0.101, \\ &p_{00} = 0.684,\ p_{11} = 0.979,\ \text{and } \sigma = 0.458, \end{aligned}$$

and the MSI(2)-AR(1) process:

$$\begin{aligned} \text{DGP}_{\text{MSI}|1} : x_t &= \gamma_0 + \gamma_1 S_t + \phi x_{t-1} + e_t \\ \text{with } \gamma_0 = -0.718,\ \gamma_1 &= 1.347,\ \phi = 0.186,\ p_{00} = 0.648, \\ p_{11} &= 0.977,\ \text{and}\, \sigma = 0.455. \end{aligned}$$

We report the empirical size and power provided by the different DGPs in Tables Table 5.2 and Table 5.3. It appears from Table 5.2 that the size is less than or close to the nominal level for both tests. In terms of power, the results in Table 5.3 suggest that the BLR test has a reasonable power across different specifications. Its power increases as the sample size increases. However, the CHP test has a very little power when the null involves autoregressive terms. It is particularly poor in the case of a switching mean, as the power is remarkably close to the size (i.e., the test is uninformative). Although the CHP test does better in the case of a switching intercept, it still suffers a significant loss of power compared to the case of no serial correlation.

As already mentioned in Section 5.2, the MSM version of the model corresponds to an AR(1) process with highly nonstandard errors when $\rho = \phi$ and the MSI version corresponds to an AR(1) process when $\rho = 0$, so we might not expect power in these cases. Note, however, that there is considerable difference in the persistence parameter ρ and the persistence of either the autoregressive process ϕ or zero in the DGPs considered here. Despite this, there is little power for the CHP test, similar to how Bessec and Bouabdallah (2005) found that different persistence parameters did not significantly affect the forecasting performance of the Markov-switching models compared to a linear benchmark. By contrast, it should also be noted that, according to the results in Di Sanzo (2009), the CHP test has much better power when the alternative allows for Markov switching in the autoregressive and variance parameters.

To summarize, the BLR test works well in all cases. But the CHP test only has power when the tested null assumes no serial correlation. If instead serial correlation is allowed for under the null, as we argue it should be when testing for Markov switching, the CHP test is particularly poor at finding evidence of nonlinearity. In the case of switching intercept, it has a little more power, but is still much less powerful than the BLR test.

5.4 Discussion

The preceding analysis can be used to better understand some previous findings in the empirical literature. Here, we consider the previously mentioned studies by Hamilton (2005) and Morley and Piger (2012). We also discuss why these results might be expected.

Hamilton (2005) employs the CHP test to check for Markov-switching nonlinearity in the US monthly unemployment rate for 1948–2004, the US short-term commercial paper interest rate for 1957–1913, and the US 6-month Treasury bill interest rate for 1957–2004. He is able to reject linearity in all three cases. However, the test statistics are much larger for the interest rate series than for the unemployment rate. One reason for the difference in the size of the test statistics is that the AR(2) model for the interest rates assumes a Markov-switching variance rather than a Markov-switching mean and the CHP test clearly has higher power in the context of Markov-switching variance (again, see the Monte Carlo results in Di Sanzo, 2009). By contrast, Hamilton considers an MSI-AR(2) model with t-distributed errors for the unemployment rate. Given the results in the previous section, we might expect the power to be much lower for this model, although there appears to be such strong nonlinearity in the unemployment rate that he is still able to reject despite the relatively low power of the CHP test in this setting. It is notable, however, that an MSM-AR model was not considered and it is possible that the CHP test would not have been significant for the specification given the results in the previous section.

Morley and Piger (2012) apply the CHP test for the US quarterly real GDP from 1947 to 2006 for the alternative hypotheses of Hamilton's (1989) model with an MSM-AR(2) specification and Kim, Morley, and Piger's (2005) bounceback model with a linear AR(2) component. They consider both Gaussian and t-distributed errors and are unable to reject linearity for the MSM-AR(2) model alternative, consistent with the low power of the CHP test reported in the previous section. By contrast, they are able to reject linearity for bounceback model alternative, which represents a more fundamental departure from linearity than the basic MSM model (for example, the point forecasts will be considerably more complicated for the bounceback model and much more different than the linear case than with the MSM model). They are also able to reject linearity based on a BLR test for an alternative bounceback model that links the strength of recovery to the depth of recession.

In terms of why the CHP test has lower power when allowing for serial correlation under the null, one issue is that the test does not actually distinguish between whether the time-varying mean under the alternative follows a discrete Markov-switching process or a continuous AR(1) process with Gaussian or t-distributed errors. This point is discussed in Section 6.1 of Carrasco, Hu, and Ploberger (2012). By contrast, a likelihood ratio test definitely makes this distinction in the sense that the likelihood of the alternative model will differ depending on the specification of a Markov-switching time-varying mean or a Gaussian time-varying mean. Then, following our analysis in Section 5.2, the CHP test really amounts to a test of whether an AR(1) model can be rejected in favor of an ARMA(2,1) model, with the alternative model being close to the null model in the case of an MSM(2)-AR(1) when the time-varying mean and the deviations from it have similar persistence (i.e., $\phi \approx \rho$) or an MSI(2)-AR(1) model when there is little persistence in the time-varying mean (i.e., $\rho \approx 0$). The BLR test will still have power in these cases because the fit of the alternative models will still be better than the linear null. Meanwhile, even when $\phi \neq \rho \neq 0$, the likelihood ratio test will have more power against a Markov-switching alternative because it will take into account the improved fit of the Markov-switching specification above and beyond any support for more complicated dynamics than an AR(1) process.

5.5 Application

In addition to discussing some previous findings, we briefly consider an application to the US real GDP growth for the sample period of 1984Q2 to 2010Q3, estimates for which were used as the DGPs in the Monte Carlo analysis in Section 5.3. The 1984Q2 start date for the sample period is designed to avoid any confusion about whether a rejection of the linear null reflects a structural break in volatility corresponding to the "Great Moderation" (see, for example, Kim and Nelson, 1999) rather than a Markov-switching mean growth rate. Additionally, the short sample makes it particularly important to consider a test with reasonable power. The test results for the BLR and CHP tests are reported in Table 5.4.

As with the Monte Carlo analysis, we begin with the AR(0) null and an MS(2)-AR(0) alternative. In this case, we can strongly reject the null of linearity for both tests. However, it is not clear that the rejection of the AR(0) null actually reflects the presence of nonlinearity or whether it reflects serial correlation in the data. If we consider an AR(1) model, the autoregressive coefficient is 0.44, with a t-statistic of 5.09. Thus, it is not surprising that the AR(0) model is rejected, regardless of whether there are actually nonlinear dynamics.

When we consider the more realistic AR(1) linear null, we continue to reject linearity when considering the BLR test. Meanwhile, the evidence is not as strong

Table 5.4 Tests for Markov-switching mean in the US real GDP growth (1984Q2–2010Q3)

AR(0)/MS(2)-AR(0)	BLR statistic = 32.75		CHP statistic=20.63	
	1%	5%	1%	5%
Critical value	10.38	7.10	7.04	3.43
AR(1)/MSM(2)-AR(1)	BLR statistic = 11.57		CHP statistic = 7.25 (p-value = 0.12)	
	1%	5%	1%	5%
Critical value	12.03	7.99	16.22	10.88
AR(1)/MSI(2)-AR(1)	BLR statistic = 13.27		CHP statistic = 5.45	
	1%	5%	1%	5%
Critical value	11.74	8.53	8.14	4.73

for the CHP test, including a failure to reject at the conventional 5% level for the MSM(2)-AR(1) alternative. These results are consistent with the power properties of the CHP test discussed above. In particular, it appears that the persistence of the Markov-switching mean is reasonably high (estimated to be around 0.66) and at least somewhat closer to the persistence of the deviations from the mean (estimated to be around 0.11–0.18) than to zero. Thus, we would expect the CHP test based on the MSI(2)-AR(1) alternative to have higher power than the test based on the MSM(2)-AR(1) alternative.

Based on these results, we conclude that there is support for nonlinearity in the form of a Markov-switching mean in the US real GDP, with the powerful BLR test providing the strongest support. Meanwhile, it should be mentioned that an AR(1) model may still be insufficient to capture the serial correlation properties of the data under the null hypothesis given statistical significance of the AR(2) and AR(3) coefficients for this particular sample period. However, even for an AR(4) model that has no apparent serial correlation in the residuals, we can still reject the MSI(2)-AR(4) alternative at the 5% level when considering the BLR test, but not for the CHP test. Meanwhile, we cannot reject the MSM(2)-AR(4) alternative with either test, but the significance level is still much higher for the BLR test than it is for the CHP test.

5.6 Conclusion

When testing for Markov switching in mean for an autoregressive process, it is important to allow for serial correlation under the null hypothesis of linearity to make sure that the power of a test actually reflects nonlinearity rather than a severe misspecification of the persistence properties of the data. However, when allowing for serial correlation under the linear null, we find that the Carrasco, Hu, and Ploberger (2012) test has very little power to detect nonlinearity in small samples for empirically relevant DGPs. By contrast, a parametric BLR test displays considerable power in this setting. The power properties are related to the fact that the BLR test takes into account the entire fit of a model, while the CHP test is related more narrowly to deviations from the autocorrelation structure implied by the null model. These properties are confirmed in an empirical application that rejects linearity for US real GDP growth with the BLR test, but not with the CHP test. Given these properties, we recommend empirical researchers consider the BLR test when testing for Markov switching in mean for autoregressive processes, although we note that further work needs to be done on the full theoretical justification for this test.

References

Bessec, M., and Bouabdallah, O., 2005, What causes the forecasting failure of Markov-Switching models? A Monte Carlo study, *Studies in Nonlinear Dynamics & Econometrics* 9, Article 6.

Carrasco, M., Hu, L., and Ploberger, W., 2012, Optimal test for Markov Switching parameters, conditionally acceptedat Econometrica. Working Paper, available at https://www.webdepot.umontreal.ca/Usagers/carrascm/MonDepotPublic/carrascm/chp.pdf, conditionally accepted at *Econometrica*.

Clements, M.P., Franses, P.H., and Swanson, N.R., 2004, Forecasting economic and financial time-series with non-linear models, *International Journal of Forecasting* 20, 169–183.

Clements, M.P. and Krolzig, H.-M., 1998, A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *Econometrics Journal* 1, 47–75.

Dacco, R. and Satchell, S., 1999, Why do regime-switching models forecast so badly? *Journal of Forecasting* 18, 1–16.

Di Sanzo, S., 2009, Testing for linearity in Markov switching models: A bootstrap approach, *Statistical Methods and Applications* 18, 153–168.

Ferrara, L., Marcellino, M., and Mogliani, M., 2012, Macroeconomic forecasting during the Great Recession: The return of non-linearity? Working paper, available at http://www.banque-france.fr/uploads/tx_bdfdocumentstravail/DT383_01.pdf.

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., 2005, What's real about the business cycle?*Federal Reserve Bank of St. Louis Review* 87, 435–452.

Hansen, B.E., 1992,The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP, *Journal of Applied Econometrics* 7, 61–82.

Morley, J. and Piger J., 2012, "The asymmetric business cycle, *Review of Economics and Statistics*94, 208–221.

Kim, C.-J., Morley J., and Piger, J., 2005, Nonlinearity and the permanent effects of recessions, *Journal of Applied Econometrics* 20, 291–309.

Kim, C.-J., and Nelson, C.R., 1999, 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, 608–616.

Krolzig, H.-M., 1997, Markov-Switching Vector Autoregressions: Modelling, StatisticalInference, and Application to Business Cycle Analysis, Berlin Springer.

Krolzig, H.-M., 2000, Predicting Markov-Switching Vector Autoregressive Processes, Working Paper, Department of Economics, University of Oxford.

Siliverstovs, B. and van Dijk, D., 2003, Forecasting industrial production with linear, nonlinear, and structural change models, Econometric Institute Report EI 2003-16, Erasmus University Rotterdam.

Teräsvirta, T., 2006, Forecasting economic variables with nonlinear models, In Elliot, G.,Granger, C. and Timmermann, A., *Handbook of Economic Forecasting* (Elsevier:Amsterdam) Chapter 8.