LUND UNIVERSITY

Transition Variables in the Markov-switching Model: Some Small Sample Properties

Erlandsson, Ulf

2005

Total number of authors:
1

Transition Variables in the Markov-switching Model: Some Small Sample Properties

Ulf G. Erlandsson[*]

March 21, 2005

Abstract

This paper researches small-sample properties of the Markov-switching model with time-varying transition probabilities. By mean of simulation, it is shown that the likelihood ratio statistic is over-sized for sample sizes relevant in many empirical applications. The number of regime switches occurring in the sample rather than the total number of observations is central to the magnitude of the distortion, with other factors such a persistence in transition equation variables and the precision at which states are inferred being influential on size. In an application to possible predictors of switches to recessions in U.S. data, it is shown that critical values for the likelihood ratio statistic need to be adjusted far upwards to reflect true confidence levels.

Keywords: regime switching, transition probability, small-sample
JEL classification: C13; C32; E32

———————————————

[*]Department of Economics, Lund University; P.O. Box 7082; 220 07 LUND; email: ulf.erlandsson@nek.lu.se; home-page: www.nek.lu.se/nekuer

Introduction

The Markov switching (MS) model of Hamilton (1989) remains a quite popular alternative to linear models. It has been used jointly with the (G)ARCH-family of models and has also been applied in the multivariate setting. Although the hypothesis underpinning these applications is the existence of different states with differing dynamics, few contributions in the literature focus on predicting states rather than the dynamics. At a first glance, this seems paradoxical.

The process of predicting states is referred to a using time-varying transition probabilities (TVP) in contrast to the constant transition probability (CTP) switching model. In econometrics, this perspective was first advocated by Diebold, Lee and Weinbach (1994). Except for Filardo (1998), there have been - to our knowledge - no direct studies of the statistical properties of the transition equation part of the TVP model. Filardo (1998) shows a few relatively non-restrictive assumptions that have to be fulfilled in order for maximum likelihood estimation of the MS-TVP model to be consistent.

This paper attempts to make an addition to this picture. Specifically, we study the small sample properties of the model. Since the model is non-linear in general and the transition equations in particular, one could suspect small sample effects to still exist in data with a relatively large number of observations. In the general regression setting, the econometrician seeks a relationship between an observed dependent variable on the left hand side, and the observed independent variables on the right hand side. At the surface, this is also what happens in the MS model. But the second part of the MS regression - that of the probabilities governing the states - differs from the general setting. The realizations of the states is only observed indirectly, so that the left hand side of the transition equation is inferred from the parameters of the model itself.

Our simulations suggest that the extra non-linearity has quite negative effects for standard asymptotics even in what can be considered quite large samples. We focus on the properties of the likelihood ratio statistic, and conclude that these apply only for data with many regime switches in relation to the noisiness of the data. In other cases, the standard χ^2 test has a larger size than suggested by ordinary asymptotics.

The distortion is large enough to have real impact in terms of model building. In order to derive the non-distorted distributions of the statistics, we opt for a simulation method where we generate a large number of likelihood ratio statistics based on the empirical parameter estimates. The method is straightforward although computationally demanding.

We illustrate the method by looking at variables suggested to predict U.S. recessions. When applying the χ^2 distribution we find 13 variables to significant at the 5% level. The corresponding figure for the simulated distribution is 10. At the 1% level, the ratio is 8 to 1. These findings have a considerable effect for further model building in a general-to-specific process.

In section 2, we introduce the general MS-TVP framework and the estimation

process. The short effective sample bias and the suggested solutions is discussed in section 3, followed by the empirical application in section 4. The last section concludes.

Model and Estimation Procedure

We base the discussion on a simple form of the Hamilton (1990) Markov regime switching model. The baseline model with constant transition probabilities (CTP):

$$\Delta y_t = \mu_{S_t} + \epsilon_t \tag{1}$$

where $\epsilon \sim N\left(0, \sigma_\epsilon^2\right)$ and S_t is a state variable that follows a first order Markov chain with transition probability matrix:

$$\mathbf{P} = \left[\begin{array}{cc} p_{11} & p_{12} \\ p_{21} & p_{22} \end{array}\right] \tag{2}$$

where, in turn, p_{ij} denotes the probability to go from state i to state j. The number of extensions made to this simple model, and the combinations thereof, can be counted in the hundreds. Most of these seek to engineer to model in a way as to have a better fit to the data, modifying the elements of equation 1, e.g. by introducing exogenous variables, auto-regressive parameters and ARCH effects. A smaller number of studies, e.g. Diebold, Lee and Weinbach (1994), have focused on modeling the probability to switch to other regimes, as in equation 2, noting that $\mathbf{P}$ by no means have to be constant. In the general 2 state case:

$$\mathbf{P}_t\left(\mathbf{Z}_t\right) = \left[\begin{array}{cc} g\left(\mathbf{Z}_t^1\right) & 1 - g\left(\mathbf{Z}_t^1\right) \\ 1 - h\left(\mathbf{Z}_t^2\right) & h\left(\mathbf{Z}_t^2\right) \end{array}\right] \tag{3}$$

where $g, h \to [0, 1]$.[1] This will be referred to the time-varying transition probability (TVP) model, as opposed to the CTP model. The functional form of f, g is usually chosen to be of probit or logit type. We will assume the logistic functional form for both h, g such that:

$$h\left(x\right) = g\left(x\right) = \frac{\exp\left(x\right)}{1 + \exp\left(x\right)} \tag{4}$$

In order to estimate the Markov regime switching model we follow Hamilton (1994) and iterate on the following equations:

$$\xi_{t|t} = \frac{\xi_{t|t-1} \odot \eta_t}{\mathbf{1}'\left(\xi_{t|t-1} \odot \eta_t\right)} \tag{5}$$

and

$$\xi_{t|t-1} = \mathbf{P}_{t-1}(\mathbf{Z}_{t-1})' \cdot \xi_{t-1|t-1} \tag{6}$$

---

[1] Extending the number of states to N is straightforward in theory, but harder in practice since the number of coefficients grows exponentially.

where η_t is a (NxT) matrix of each N states conditional density based on the parameter vector θ and ξ is a NxT matrix of inferred probabilities for each of N states to have occured. For the 2 state case:

$$\eta_t' = \left[\quad \frac{1}{\sqrt{2\pi}\sigma_1} \mathrm{e}^{\frac{-(y_t-\mu_1)^2}{2\sigma_1^2}}; \quad \frac{1}{\sqrt{2\pi}\sigma_2} \mathrm{e}^{\frac{-(y_t-\mu_2)^2}{2\sigma_2^2}} \quad \right] \tag{7}$$

The log likelihood to be maximized is given by:

$$L(\theta) = \sum_{t=1}^{T} \log \mathbf{1}' \left(\xi_{t|t-1} \odot \eta_t \right) \tag{8}$$

We note from the above that $\xi_{t|t-1}$ is the vector of inferred probabilities given time $t-1$ information that the process will be in state i at time t. Another measure of our inference on the state at time t can be useful; namely the *smoothed* probabilities that assign probabilities of each regime using full sample information T. They are given by the algorithm developed by Kim (1994):

$$\xi_{t|T} = \xi_{t|t} \odot \left\{ [\mathbf{P}_t]' \left[\xi_{t+1|T} \div \xi_{t+1|t} \right] \right\} \tag{9}$$

which is iterated from time $T - 1, T - 2, ..., 1$.

Short effective sample bias

As shown by Dacco and Satchell (1999), one condition for the Markov switching model to be efficient in terms of forecasting is considerable persistence in regimes. This is also seen in much of the empirical papers where the elements along the principal diagonal of $\mathbf{P}$ generally exceed 0.5 (see e.g Hamilton 1989, Gray 1996, Ang and Bekaert 2002) in the constant transition probability model. This implies that we observe relatively few regime switches in the data compared to the total number of observations.

For the transition equation parameters of TVP model, the dependent variable is in principle the number of regime switches - i.e. the number of transitions - rather than the total number of observations. Too see this, we assume that $\xi_{t+1|t}$ and $\xi_{t|t}$ are exogenously given and manipulate (6) to get

$$\mathbf{P}_t(\mathbf{Z}_t)' = \xi_{t+1|t}\xi_{t|t}' \left[\xi_{t|t}\xi_{t|t}' \right]^{-1} \tag{10}$$

For most of the time, if we have persistent Markov processes the distance between $\xi_{t+1|t}$ and $\xi_{t|t}$ will be small. If we let both of these matrices $\to [1 \ 0]$ ($N = 2$), we see that the right hand side of (10) converges to the identity matrix. The same happens when $\xi_{t+1|t}$ and $\xi_{t|t} \to [0 \ 1]$. When $\xi_{t+1|t}$ and $\xi_{t|t}$ instead goes towards the opposite positions (eg. $\xi_{t+1|t} \to [1 \ 0]$ and $\xi_{t|t} \to [0 \ 1]$)the right-hand side has 0s in its principal diagonal and 1s in the off diagonal elements. The latter opposite direction case happens when regime switches occur. Hence, we can see that all the variation in the right-hand side of (10) comes from the time periods where we observe regime

switches.

For purposes of the TVP parameters whose values essentially are found by observing the variation in the right-hand side of (10), one could have suspicion that the number of regime switches rather than the total sample size is the effective sample size. Once this line of thought is established, the questions about small sample properties of standard tests must also be raised.

To study whether this effect is actually in the data, we have simulated a large number of time-series with Markov switching dynamics with parameters $\mu_1 = -0.5$, $\mu_2 = -0.5$, $\sigma_\epsilon = 1$, $p_{11} = p_{22} = 0.95$. We have then estimated a CTP model on the each series and calculated the corresponding log likelihood value denoted ℓ^C. In the second step, we estimated a TVP model such that $\mathbf{Z}_t^1 = [1 \quad z_t]$ and $\mathbf{Z}_t^2 = [1]$. In order to investigate how the persistence of the z_t process influence results, the first order autoregression DGP was used:

$$z_t = \phi^i z_{t-1} + u_t \tag{11}$$

where $u_t \sim N\left(0, \sigma_u^2\right)$, $\sigma_u^2 = 1$. We conducted the experiment over a range of $\phi^i \in \Phi = [0, 0.25, 0.5, 0.75, 0.95]$ and for differing sample sizes of 100, 150, 250, 500, 750 and 1000. For brevity, we present the cases for $\phi^i \in \Phi = [0, 0.95]$ below. The log likelihood value of the TVP model was computed and denoted ℓ^T. The likelihood ratio statistic $LR = 2(\ell^T - \ell^C)$ should then be distributed according to a χ^2 distribution with 1 degree of freedom. The actual results in terms of probability values are plotted in figure 1.

The size of the LR test increases significantly for the lower range of effective sample sizes as the variance is increased. For data with only one switch to the studied regime, the acceptance frequency is 25.7%/32.2% for the non-persistent/persistent case. With 10 switches to the investigated regime, we still find 8.7%/11.2% of the statistics exceeding the standard χ_1^2 5% critical value. We require almost 30 switches before the size of the test is close to its nominal level.

To illustrate the differences in the empirical distributions vis-a-vis the theoretical one, we calculated probability density functions for the simulated statistics using an Epanechikov kernel estimator. The plots of these density functions compared to a χ_1^2 can be seen in figure 1, panel (b). The solid line marks the pdf of the statistics based on 3 effective observations. It has markedly fatter tails than the pdf for the 25 effective observation case (broken line) and the χ_1^2 distribution. The pdf of the 25 switch distribution definitely converges towards the theoretical distribution relatively to the 3 observation case. A battery of test do however, for both the cases, reject the null hypothesis of the empirical distribution being χ_1^2.

To continue the investigation, it is useful to note that what essentially happens when the estimates of the regressors in the TVP equation are calculated, is a regression between the TVP independent variables and the inferred states. If we have strong inference on what state occurs when, it should be easier to quantify the relationship between the TVP variables and the unobserved states. If, however, the inference is weak, it seems logical that the relationship would be harder to quantify.

Our inference will be dependent upon how noisy the data is, so it is useful to

Figure 1: Case $\sigma_\epsilon = 1$, $\phi = 0$ full lines and $\phi = 0.95$ dotted lines. Panel (a): Percentage of simulated statistics exceeding the 5% critical value for different effective sample sizes proxied by the number of observed regime switches (horizontal axis). Panel (b): approximated probability density functions for the $\phi = 0$ case, solid line for the case of 3 effective observations, dotted line for 25 effective observations and the bold line indicates the pdf of a χ_1^2 distribution.

have a measure of the noisiness. The MS model is essentially a mixture of distributions where the timing of the draws is governed by a Markov process. Stronger inference on whether one draw belong to one distribution or the other will lead to stronger inference regarding the governing process. The following measure will be used to quantify the noisiness of data generated under a MS model:

$$Q^{mid} = 0.5 \cdot \left\{ \frac{\Phi\left(\mu_1, \sigma_2^2\right) + \left[1 - \Phi\left(\mu_2, \sigma_1^2\right)\right]}{2} + 0.5 \right\}^{-1} \tag{12}$$

where $\mu_1 < \mu_2$, and Φ is the cumulative distribution function of the normal distribution. This measures the mass of the right-hand distribution to the left of the mean of left-hand distribution added with the mass of the left-distribution to the right of the right-hand distribution relative to the total mass. A richer measure would include the effect of high or low transition probabilities on the noisiness, but we leave that for future work. For now, the noisiness measures are conditional on a given set of transition parameters.

A similar measure is based on making confidence intervals for a given significance level α and measuring the magnitude of the intersection of the probability distribution functions inside the intervals. This answers the question "Given the $1 - \alpha$ significance level, how often will we erroneously classify the generating distribution of any given draw?". The relevant statistic is then:

$$Q^{\alpha} = \Theta\left(\alpha\right) \cdot \left\{ \Phi\left(\mu_1 + \Theta\left(\alpha\right)\sigma_1, \sigma_2^2\right) + \left[1 - \Phi\left(\mu_2 + \Theta\left(1 - \alpha\right)\sigma_2, \sigma_1^2\right)\right] + \Theta\left(\alpha\right) \right\}^{-1} \tag{13}$$

where Θ is the probability distribution functions of the standard normal distribution.

For the data generated so far, the first measure Q^{mid} of noisiness equals 75.9% meaning that if we pick an observation which is either to left of the intercept of the left-hand distribution or to the right of the right-hand distribution, we have a 75.9% chance to classify it correctly. If we instead pick a draw that is within the 95% left-hand mass of the left-hand distribution, we have a 54% chance to be correct, and analogously for the right hand distribution since they are assumed to be symmetrical with $\mu_1 = -\mu_2; \ \sigma_1 = \sigma_2$.

To test whether a reduced noisiness has effect on the size distortion of the likelihood ratio test, we conducted a similar Monte Carlo but with the variance parameters set to $\sigma_\epsilon = 0.25$. In other words, we let the simulated data be less noisy with $Q^{mid} > 0.999$ and $Q^{0.05} = 0.979$ as yardsticks. Thus our inference on the hidden states is sharper. The results of this simulation is presented in figure 2.

We note that the size distortion is drastically reduced compared to the earlier experiment. For effective sample sizes exceeding 10, the size seems to be close to its nominal 5% level. For smaller samples it still is oversized. Again, we observe a convergence of distributions towards the theoretical one as the effective sample size is increased. In this case, we cannot reject the null of the empirical distribution being a χ_1^2 distribution for either the 3 or the 25 sample size. The tendency is clear: as the variance in the non-TVP part of the model decreases, the closer the properties
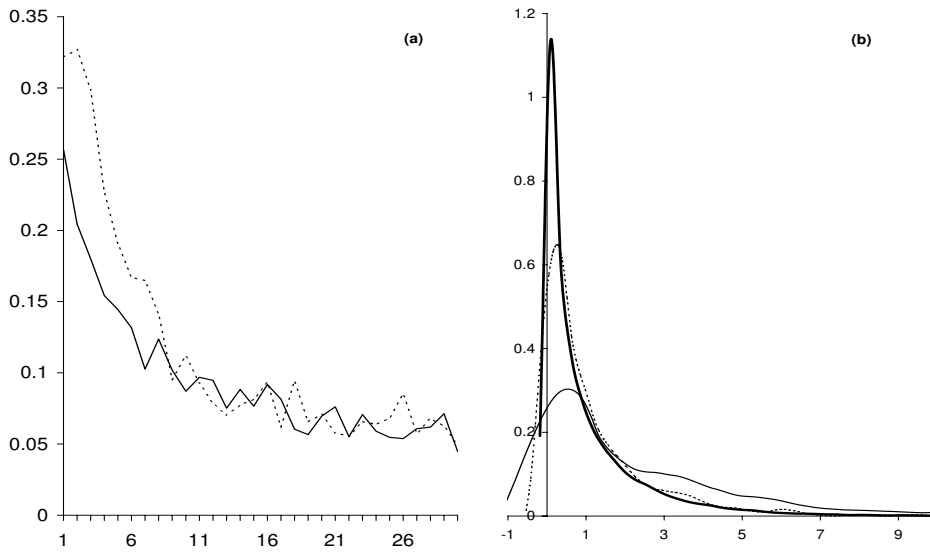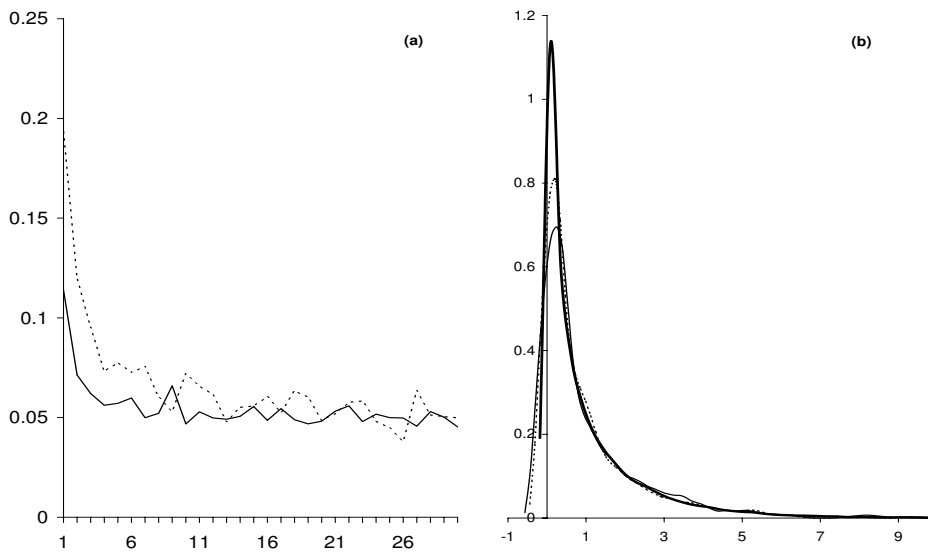
Figure 2: Case $\sigma_\epsilon = 0.25, \phi = 0$ full lines and $\phi = 0.95$ dotted lines. Panel (a): Percentage of simulated statistics exceeding the 5% critical value for different effective sample sizes proxied by the number of observed regime switches (horizontal axis). Panel (b): approximated probability density functions for the $\phi = 0$ case, solid line for the case of 3 effective observations, dotted line for 25 effective observations and the bold line indicates the pdf of a χ_1^2 distribution.

Figure 3: Size distortions for differing levels of volatility in the TVP equation, $\sigma_u^2 = [0.5, 1, 2, 3]$, depicted by the flat line, squares, crosses and plus signs in the same order.
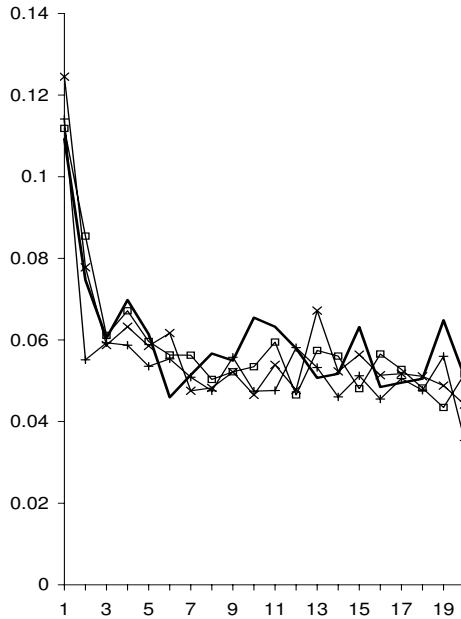
of the likelihood ratio test are to the theoretical ones.

A relevant question is whether the size distortion in the persistent cases are related to the persistence itself, or the increased variance. The variance of an autoregressive process such as (11) has the following well known property:

$$\sigma_z^2 = \frac{\sigma_u^2}{1 - \phi}$$

To determine if the persistence has effect directly or through the indirect volatility increase, we conduct a similar experiment but allow σ_u^2 to vary between $[0.5, 1, 2, 3]$. From figure 3, we conclude that changing the variance σ_u^2 does not systematically influence the size of the LR test. Consequently, the sole factor explaining the differences in the size of the tests should be the persistence of the process.

For empirical purposes, these findings are discomforting. One can suspect that the problem is exacerbated by the introduction additional variables in the transition equation. If so, once runs a very large risk of introducing spurious inference about the suggested transition variables and the true transition process. Inference on possible explanatory variables for the transition probabilities may well come from overfitting rather than a true causality. In the end, this would also be reflected in non-sensible point estimates and poor out-of-sample forecasting properties. Since

the MS model generally is less than trivial to estimate, a too large number of variables in a general-to-specific methodology comes with considerable estimation difficulties.

As shown above, short effective samples for many cases distorts the size of the likelihood ratio statistic. Moreover, the magnitude of this distortion seems to be dependent both upon the dependent variable in question, as well as the regressor variable in the transition equation.[2] Therefore, it is prohibitively burdensome to produce tables of relevant critical values to correct for the bias.

Instead, we propose a generic parametric bootstrap method to produce the distribution under the data-specific conditions:

1. Estimate the set of empirical parameter values for the CTP and TVP models, collected in the vectors $\hat{\theta}^{CTP}$ and $\hat{\theta}^{TVP}$. Compute the log likelihood values ℓ^C and ℓ^T, and compute the empirical likelihood ratio $\mathrm{LR}^{EMP} = 2\left(\ell^T - \ell^C\right)$. Calculate the effective sample size proxy by computing the sum of absolute first differences of the series of smoothed state probabilities.

2. Estimate the persistence of the transition variable as the AR(1) coefficient in equation (11), denoted $\hat{\phi}^{EMP}$.

3a. Generate a new series of observations based on $\hat{\theta}$, with the same effective sample size as the empirical series. Generate the transition variable as an AR(1) process with persistence coefficient $\hat{\phi}^{EMP}$. Estimate the CTP and TVP models on the simulated data, and compute the corresponding likelihood ratio statistic, denoted LR_i^{SIM}.

or

3b. Generate a new series of observations based on $\hat{\theta}$, with the same effective sample size as the empirical series. Generate TVP regressor series by simulating a set of AR processes with a grid over values of ϕ. Estimate the CTP and TVP models on the simulated data, and compute the corresponding likelihood ratio statistic, denoted .

4. Perform M repetitions of step 3.

5. Calculate the p-value of the empirical likelihood ratio test as the number of $\mathrm{LR}_i^{SIM}/\mathrm{LR}_{i,\phi}^{SIM}$ that exceed LR^{EMP} divided by the total number of simulations.

When the number of TVP variables to test is small, one may consider using step 3a., whereas when a larger number of variables should be tested, step 3b. is to prefer. In the latter case, one compares the empirical likelihood ratios with the corresponding α percentiles in the grid; preferably with the one closest above the empirical persistence level to avoid Type I errors.

Earlier we noted that the effective sample size, that is, the number of switches to the regime under study, has a a large effect on the size distortion. Since the Markov chain is generally assumed to be hidden, this number has to be estimated

---

[2]The size of the test is also decreasing in the total sample size *given* a specific effective sample size. This is related to the noisiness of the data; more switches in a given sample size yields noisier data. Hence, we do not probe further into this issue. Details on this is available upon request.
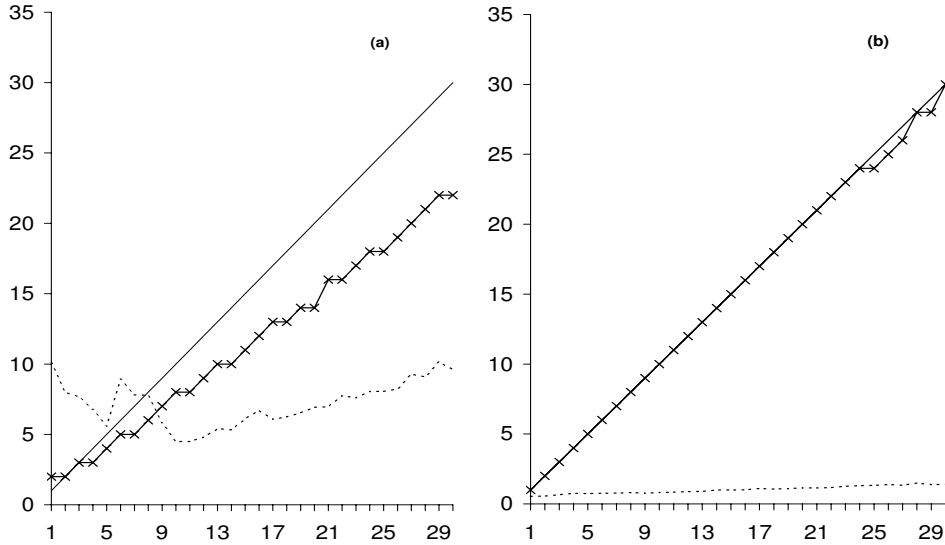
Figure 4: Smoothed switches estimator of true number of regime switches. Panel (a) depicts the $\sigma_\epsilon = 1$ case, panel (b) the $\sigma_\epsilon = 0.25$ case. Solid line indicates the 45 degree line, crosses the median of the estimates $\hat{\kappa}$, broken line the root mean squared error of $\hat{\kappa}$ vis-a-vis the real number of switches.

from the data. The question is whether the smoothed probability estimate of the number of regime switches is a good one. As a byproduct of earlier simulations, we have calculated a measure of the number of regime switches as

$$\hat{\kappa} = \sum_{t=1}^{T} \left\{ \begin{array}{c} 1 \;\; \text{if} \;\; \xi_{t|T}^2 \geq 0.5 \;\; and \;\; \xi_{t-1|T}^2 < 0.5 \\ 0 \;\; \text{otherwise} \end{array} \right. \tag{14}$$

We see that for the less noisy case, the estimate of the actual number of switches works quite well. For the noisier data, performance deteriorates markedly and cannot be considered particularly reliable. Our suggestion in this case is to use visual inspection of the data and the smoothed probabilities to determine the effective number of switches, and to use a conservative (low) estimate. Experience also suggests that as noisiness increases, tests are less inclined to reject a null test of no MS dynamics, so that at some level of noisiness the above analysis is rendered irrelevant since no evidence of MS dynamics is found at all.

Empirical Application

Now we will illustrate this testing procedure with an empirical example. Data have been obtained from the *Economagic* data base and the National Bureau of Economic Research (NBER).[3] Similarily to Hamilton (1989) and subsequently Filardo and Gordon (1998), we analyze quarterly GDP figures over the sample 1964:I-2004:IV. We consider a simple model of Markov switching with the level equation following

$$\Delta y_t = \mu_{S_t} + \epsilon_{R_t,t} \tag{15}$$

where $\epsilon_{R_t,t} \sim N\left(0, \sigma^2_{R_t}\right)$. S_t and R_t are two possibly different Markov processes. The first step is to validate that we have regime switching at all, which is tested by comparing a model where $R_t = S_t$ with the simple single-regime benchmark process

$$\Delta y_t = \mu + \epsilon_t \tag{16}$$

As pointed out in earlier literature (e.g. Hansen (1992)), testing this is a non-standard statistical problem. Under the null of (16), the transition matrix parameters in the alternative are not identified, and standard statistical procedures are no longer applicable. The likelihood ratio statistic does consequently no longer have its ordinary χ^2 distribution. Cheung and Erlandsson (2004) propose a testing procedure based on simulation to obtain the empirical distribution of the statistic both under the null and the alternative. Holding model (16) as the null, and (15) with $R_t = S_t$ as the alternative, we reject the null with a p-value of 0.004. When reversing the null and the alternative, we are unable to reject the null based on a p-value of 0.574. Hence, we conclude that there are Markov switching dynamics with at least 2 states in the data.

A correct classification of states is essential to the empirical results in this model so we continue to test the null of model (15) with $R_\tau = S_\tau$, $\tau = \{1, 2, ..., T-1, T\}$ to the alternative of $R_\tau \neq S_\tau$. This is a test of a 2 state model vis-a-vis a restricted 4 state model, where we again obtain the empirical distribution of the likelihood ratio statistic via simulation. We reject the null of $R_\tau = S\tau$ with a p-value of 0.008, and are unable to reject the null of $R_\tau \neq S_\tau$ with a p-value exceeding 0.99. These findings lead us to proceed with model (15) in its unconstrained form.

The resulting model replicates the NBER recession dates almost exactly in terms of the S_t state process. The volatility process R_t seems to indicate a much more volatile behavior of real GDP up till the mid-eighties. After that, the movements have been much smaller. We find neither any significant auto-correlation in the (standardized) residuals, nor ARCH effects.

The final objective is to evaluate the usefulness of a number of exogenously suggested predictors of the business cycle on the data. These indicators are given as "business cycle indicators" in the Economagic data base, and are presented in table 2. In terms of policy, it seems most reasonable to focus on predicting recession. Hence, we let the transition probability from the $\mu > 0$ (denoted $S_t = 1$) state to
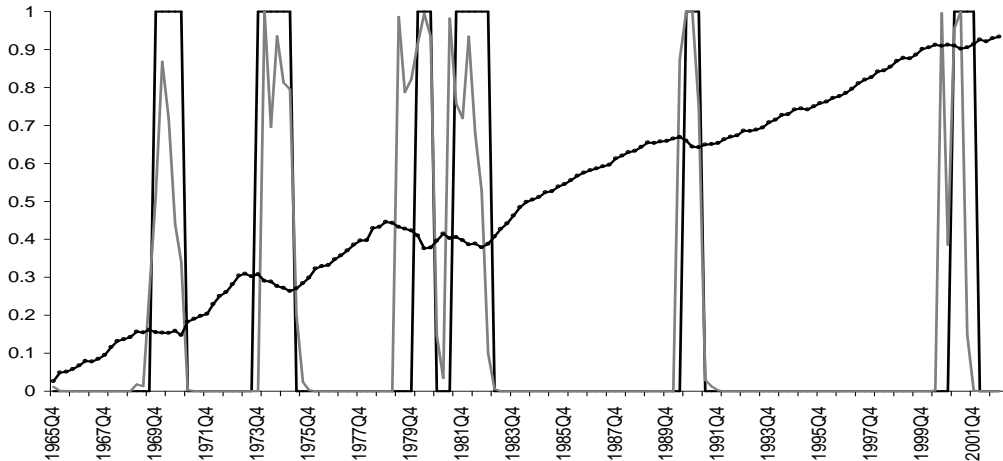
---

[3]www.economagic.com

Figure 5: NBER dates (black) and smoothed probabilities of the contractionary state (grey) using the constant transition probability model. U.S. real GDP is plotted with dots (normalized to 0 in 1965Q4).

the $\mu < 0$ ($S_t = 2$) state be time varying. In terms of the transition probability matrix:

$$\mathbf{P}_t\left(\mathbf{Z}_t\right) = \begin{bmatrix} g\left(\mathbf{Z}_t^1\right) & 1 - g\left(\mathbf{Z}_t^1\right) \\ p_{21} & p_{22} \end{bmatrix} \tag{17}$$

where $\mathbf{Z}_t$ is the set of predictor variables to be evaluated.

One advantage of analyzing this data is the fact that the number of regime switches is exogenously given by the NBER dating. But one should also note that the model itself actually replicates the same number of regime switches to the recession states, which is 6. The variance of the two states is 0.4778 for the low volatility states and 1.0645 for the high volatility state. Conditioning on the low volatility state, noise-measures are $Q^{mid} = .996$ and $Q^{0.05} = .838$, and for the high volatility state $Q^{mid} = .832$ and $Q^{0.05} = .564$. For the latter state, we cannot expect the smoothed estimator of the number of regime switches to be particularly precise, and should complement it with a qualitative analysis. From figure 5, we can infer that the estimate of 6 switches to the recessionary state of the economy - as indicated by our estimate - seems to be correct. Given the earlier simulation exercises, we can suspect the size of the likelihood ratio test to be distorted in this setting. Hence, we conduct the bootstrap procedure proposed above to obtain non-distorted distributions of the likelihood ration test. The 5% critical values for different levels of persistence are presented in column 2 of table 1. In the third column, the numbers presented are the counterpart χ^2_1 probability values based on the empirical 5% critical values. Rather than being their nominal 5%, they are between 1.1% and 1.7%.

13

ϕ	5% Crit. Val.	$(\chi_1^2)^{-1}(5\%)$
0	5.743	0.017
0.2	5.709	0.017
0.4	5.783	0.016
0.6	5.864	0.015
0.7	6.142	0.013
0.8	6.148	0.013
0.9	6.178	0.013
0.925	6.405	0.011
0.95	6.331	0.012
0.975	6.336	0.012

Table 1: Critical values and corresponding χ_1^2 probability values for different ϕ.

In table 2, we present the probability values for each of the suggested indicator both in terms of the χ^2 distribution and calculated based on the simulated probability density function. We note that 13 variables are significant using the χ_1^2 distribution and the 5% significance level, whereas only 10 are using the empirical likelihood ratio distributions. At the 1% significance level, the problem is exacerbated, only 1 variable is found significant with the empirical test in contrast to 8 using the standard statistic. The figure is 14/14 at the 10% level. It seems that for purposes of further investigating the inter-relationships between these variables, deriving the empirical distribution for the tests has marked effects at the lower traditional significance levels.

Conclusion

In this paper, the properties of the likelihood ratio test on the variables in the transition equation of the Markov switching model are investigated. It is argued that the non-linear features of the model itself, and the regression of a set of variables on an unobserved dependent variable may give arise to small sample effects even in relatively large sample. This is shown to be the case via a number of simulation exercises. In practice, we tend to reject the null of no significance of TVP variables too often using the widely applied likelihood ratio test.

To what degree the test is oversized is shown to depend upon the noisiness of the MS process, the number of regime switches in the data, and the persistence of the TVP regressor in question. The magnitude of the size distortion is quite large in the not too unrealistic cases we study. We propose a simulation procedure to generate empirical distributions of the likelihood ratio statistic and in that way obtain statistics with their actual size equalling the nominal.

In an empirical setting, we study the U.S. business cycle based on quarterly real GDP. Our model captures the NBER business recession dates quite well. Cal-

ibrating our simulation procedure the model, we show that the probability for an orthogonal TVP regressor to be significant at the 5% level using the standard χ_1^2 distribution lies around 10.5%-13.5%, so the small sample problem indeed exists even if the data exceeds 150 observations. When the empirical distributions are applied to a set of business cycle indicators, we see that at the 5% level, 10 rather than 13 variables are deemed significant. At the 1% level, the corresponding figures are 1 vis-a-vis 8 variables significant. For further modeling purposes, this should have considerable implications.

References

Ang, A. and Bekaert, G. (2002). Regime Switches in Interest Rates. *Journal of Business and Economic Statistics*, 20:2:163–82.

Cheung, Y.-W. and Erlandsson, U. G. (2004). Exchange rates and Markov switching dynamics. Forthcoming in Journal of Business and Economic Statistics.

Dacco, R. and Satchell, S. (1999). Why Do Regime-Switching Models Forecast So Badly. *Journal of Forecasting*, 18:1–16.

Diebold, F. X., Lee, J.-H., and Weinbach, G. (1994). Regime switching with time-varying transition probabilities. In Hargreaves, C., editor, *Nonstationary Time Series Analysis and Cointegration*, pages 283–302, Oxford. Oxford University Press.

Filardo, A. J., , and Gordon, S. F. (1998). Business cycle durations. *Journal of Econometrics*, 85:99–123.

Filardo, A. J. (1998). Choosing Information Variables for Transition Probabilities in a Time-Varying Transition Probability Markov Switching Model. Federal Reserve Bank of Kansas City Working Paper 98-09.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62.

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. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 45:39–70.

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

Hansen, B. E. (1992). The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics*, 7:S61–S82.

Kim, C.-J. (1994). Dynamic Linear Models with Markov–Switching. *Journal of Econometrics*, 60:1–22.

Indicator	Frequency	Index number	Emp. p	χ^2 p
Total private: Indexes of Aggregate Weekly Hours, SA	M	1	.730	.663
Average Weekly Hours; Private Nonagricultural Establishments; SA	M	2	.244	.170
Total Borrowings at Federal Reserve Banks; Billions of Dollars; NSA	M	3	.540	.488
Change in Business Inventories; SAAR Billions of Dollars	Q	4	.435	.339
Corporate Profits After Tax with IVA and CCAdj; Billions; SAAR	Q	5	.589	.540
Consumer Price Index All Urban Consumers: Total; 1982-84=100; SA	Q	6	.017	.003
Consumer Price Index All Urban Consumers: Less Food and Energy; 1982-84=100, SA	M	7	.030	.006
Employment Ratio; Civilian Employment/Civilian Non. Inst. Pop.; Percent SA	M	8	.613	.518
Gross Savings; Billions of Dollars SAAR	Q	9	.491	.436
Index of Help Wanted Advertising; in Newspapers; 1987=100; SA	M	10	.036	.010
Total Industrial Production Index; 1992=100 SA	M	11	.081	.032
M2 Money Stock; Billions of Dollars; SA	M	12	.568	.485
Bank Prime Loan Rate	M	13	.030	.008
Nonfarm Business Sector: Output Per Hour of All Persons; SA, 1992=100	Q	14	.325	.253
Private Business Sector: Output Per Hour of All Persons; SA, 1992=100	Q	15	.535	.482
Payroll Employment; of Wage and Salary Workers; Thousands; SA	M	16	.387	.284
Personal Consumption Expenditures; Billions of Dollars SAAR	M	17	.761	.741
Personal Income; Billions of Dollars SAAR	M	18	.532	.482
PPI - Capital Equipment; 1982=100 SA	M	19	.086	.029
PPI - Crude Materials for Further Processing; 1982=100 SA	M	20	.043	.014
PPI - Finished Consumer Foods; 1982=100 SA	M	21	.186	.114
PPI - Finished Goods; 1982=100 SA	M	22	.021	.005
PPI - Intermediate Materials; 1982=100 SA	M	23	.022	.004
Personal Saving; Billions of Dollars SAAR	Q	24	.090	.039
Civilian Unemployed for 15 Weeks and Over; Thousands; SA	M	25	.696	.653
Manufacturing Sector: Unit Labor Cost; SA, 1992=100	Q	26	.003	.000
Nonfarm Business Sector: Unit Labor Cost; SA, 1992=100	Q	27	.014	.003
Consumer Sentiment; University of Michigan; 1966Q1=100; NSA	Q	28	.056	.019
Unemployment Level; All Civilian Workers; Thousands; SA	Q	29	.392	.312
Capacity Utilization: Manufacturing (SIC); SA	M	30	.380	.261
Industrial Production Index: Consumer goods; 1997=100; SA	M	31	.026	.006

Table 2: Evaluated predictors of the transition probability to the contraction state.