University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 13/2017 (242)

BUSINESS CYCLE DATING AFTER THE GREAT MODERATION: A CONSISTENT TWO – STAGE MAXIMUM LIKELIHOOD METHOD

GILBERT MBARA

Warsaw 2017

Business Cycle Dating after the Great Moderation:
A Consistent Two – Stage Maximum Likelihood Method

GILBERT MBARA
Faculty of Economic Sciences
University of Warsaw
e-mail: gmbara@wne.uw.edu.pl

Abstract

The two-state Markov switching model of dating recessions breaks down when confronted with the low volatility macroeconomic time series of the post 1984 Great Moderation era. In this paper, I present a new model specification and a two--stage maximum likelihood estimation procedure that can account for the lower volatility and persistence of macroeconomic times series after 1984, while preserving the economically interpretable two--state boom--bust business cycle switching. I first demonstrate the poor finite sample properties (bias and inconsistency) of standard models then suggest a new specification and estimation procedure that resolves these issues. The suggested likelihood profiling method achieves consistent estimation of unconditional variances across volatility regimes while resolving the poor performance of models with multiple lag structures in dating business cycle turning points. Based on this novel model specification and estimation, I find that the nature of US business cycles has changed: economic growth has permanently become lower while booms last longer than before. The length and size of recessions however remain unchanged.

Keywords:

Regime Switching, Hidden Markov Models, Great Moderation, Maximum Likelihood Estimation

JEL:

C5, C51, C58, C32, E32

Acknowledgements

I would like to thank Jacek Suda and Ryszard Kokoszczynski as well as participants at the GRAPE Internal Seminars for their valuable comments. Usual disclaimer applies.

DOI:

https://doi.org/10.26405/WP/WNE/2017/242/013

1 Introduction

Since the publication of Hamilton's (1989) influential paper on the analysis of nonstationary time series, the two state regime switching model has become a predominant approach in the business cycle dating and forecasting literature while also finding multiple applications in different areas of macroeconomics.[1] However, as noted by Boldin (1996) and more recently by Chauvet and Su (2014), the model breaks down when the data are extended to periods after 1984; it fails to identify the transitions from booms to recessions that characterise business cycles across most of the worlds industrialized economies. Business cycle forecasters have identified the need to model the fall in volatility of many macroeconomic time series, now generally referred to as the "Great Moderation"[2] in order to maintain the model's empirical relevance using time series after 1984. The most interesting approach and the focus of this paper, has been the introduction of a second Markov chain to model the downward shift in volatility during Great Moderation era.[3]

While this solution is quite innovative and has been successfully used to address the problem of identifying business cycle phases using time series post 1984, the implied estimates of the unconditional variance of time series before and after the onset of moderation show very large biases. The bias in the unconditional variance estimates would not be of much consequence if researchers were merely interested in business cycle dating (in which case the estimated (un)conditional volatilities could be disregarded as nuisance or incidental parameters). However, these estimates have been used to calibrate macroeconomic models that explain the boom in asset prices (see e.g. Lettau, Ludvigson and Wachter 2008, Liu and Miao 2015) and to explain the volatility reduction itself (Bullard and Singh 2012) without careful evaluation of how close estimates are to the artefacts of interest. The statistical properties of maximum likelihood (ML) parameter estimates of Markov switching models are based on large sample asymptotic theory (see e.g. Franke 2012). However, simulation studies have shown that ML estimates of these models as characterized by large biases (see Psaradakis and Sola 1998, Ho 2001). The macroeconomics literature does not seem to have

[1] See Hamilton (2011), Barnett, Chauvet and Leiva-Leon (2016) for applications to forecasting recessions, Sims and Zha (2006) on monetary policy shifts, Lanne, Lütkepohl and Maciejowska (2010), Netsunajev (2013), Lütkepohl and Velinov (2016) on identification of structural vector auto-regressions via heteroskedasticity and Hamilton (2016) for a general survey of other applications.

[2] See e.g. Justiniano and Primiceri (2008) and the references therein.

[3] An early application of this approach to modelling economic growth with two independent Markov chains that captured the Great Moderation was implemented by McConnell and Perez Quiros (2000). In their textbook implementation of Hamilton's (1989) model, Kim and Nelson (1999) add a dummy variable for growth rates after 1984 to potentially capture "a change in the mean growth rates during boom or recession"(p. 80). Other implementations of this approach include Chen (2006), Bai and Wang (2011), Doornik (2013) and Chauvet and Su (2014) who also provide a comprehensive survey. Lettau et al. (2008) also use independent chains to model the lower volatility of consumption growth while examining how lower risk led to the stock market boom of the 1990s

given much attention to the potential bias inherent in estimates from these models, yet such biases could have considerable impacts on the interpretation of calibrated models.

In this paper, I first document the bias in unconditional variance estimates implied by ML estimators from Markov switching models (see Section 2.2). I link this general bias to an old statistical result due to Neyman and Scott (1948) on the inconsistency of ML estimators when dealing with partially consistent observations/data.[4] I then propose a new model specification and estimation procedure that exploits the Bayesian structure of Hidden Markov Models to implement an idea originally due to Basu (2011) on dealing with incidental/nuisance parameters. One outcome of the "Double Mixture Autoregressive" model I propose and its estimation procedure is that it allows for the lag order or autocorrelation structure of the data generating process (DGP) to vary across variance regimes. This is a novel implementation of a Markov switching regression that has not be attempted in the economics or related statistical literature. By accounting for the change in the lag length of time series across the different volatility periods, I am able to obtain more precise estimates of the unconditional moments. Finally, my model allows for the intercepts, persistence and state duration of the business cycle related growth phases to change across volatility regimes, which as noted by Kim and Nelson (1999) has been an important feature of the data.[5] The remainder of the paper is organized as follows: in the next section (2), I revisit the standard Markov switching model and document the size of the bias in estimates of the unconditional variance in the literature. In section 3, I specify a new type of model, "A Double Mixture Autoregressive Model" and describe a likelihood profiling method that affords estimation and filtering. In section 4, I give results of estimating the model for using 3 time series: US GNP, Industrial Production and Consumption Growth. The final section concludes with a discussion of the results and potential applications.

2 The Markov Switching Model

Hamilton (1989) modelled the growth rate in United States GNP as the outcome of a first order hidden Markov chain with states $S_t = \{0, 1\}$ defined by transition probabilities:

$$p_{j|i} = \text{Prob}\left[S_t = j | S_{t-1} = i\right], \quad i, j = 0, 1 \tag{1}$$

[4] Markov Switching models of the business cycle fit this framework as we are essentially dealing with observation sequences that obey a probability law with different parameter values.

[5] There has been a narrowing of the gap between growth rates during booms and recessions post 1984. See Filardo and Gordon (1998) or Chang, Choi and Park (2017, sec 5.2) for alternative takes on the time varying duration issue.

The time series y_t, follows the auto-regression:

$$y_t = \mu_{S_t} + \sum_{l=1}^{\ell} \phi_l \left(y_{t-l} - \mu_{S_{t-l}} \right) + e_t, \quad e_t \sim \text{i.i.d } N(0, \sigma_e^2) \tag{2}$$

where mean growth rate μ_{S_t} depends on current state S_t and realizations of the state going back to a maximum of ℓ lags. For ease of exposition, I denote this specification as the Markov Switching Mean - Autoregressive [MSM - AR(ℓ)] model. The presence of lags in the dynamic regression can be dealt with by defining a new hidden state variable together with a transition matrix. For example, if $\ell = 1$ so that we have only one lag, we can define a new 4 - state Markov chain with states S_t^* and transition matrix $\mathbf{P}$:

$$S_t^* = \begin{cases} 0 & \text{if } S_t = 0,\ S_{t-1} = 0 \\ 1 & \text{if } S_t = 1,\ S_{t-1} = 0 \\ 2 & \text{if } S_t = 0,\ S_{t-1} = 1 \\ 3 & \text{if } S_t = 1,\ S_{t-1} = 1 \end{cases} \quad ; \quad \mathbf{P}^* := \begin{pmatrix} \mathrm{p}_{0|0} & \mathrm{p}_{1|0} & 0 & 0 \\ 0 & 0 & \mathrm{p}_{0|1} & \mathrm{p}_{1|1} \\ \mathrm{p}_{0|0} & \mathrm{p}_{1|0} & 0 & 0 \\ 0 & 0 & \mathrm{p}_{0|1} & \mathrm{p}_{1|1} \end{pmatrix} \tag{3}$$

The state variable S_t^* now keeps track of the previous period's state but is still the outcome a first-order Markov chain whose transition matrix $\mathbf{P}^*$ preserves the normalization $\sum_{j=0}^{3} \mathrm{p}_{j|i} = 1$ for $i = \{0, 1, 2, 3\}$. Estimating this model requires a specification of the likelihood function and which can then be efficiently evaluated using the filtering procedure of Hamilton (1990) after expanding the state-space and transition matrix to account for the presence of lags.

2.1 Split Sample Estimates

In order to better infer why Hamilton's original specification no longer works, it is useful to estimate the model using separate samples: before and after 1984. First, this would show if we can "independetly" use the model post 1984. Second, we would be able to see if there are any significant changes in the dynamics of the data post 1984 which we may need to account for when using the whole time-series.

I estimate the model specified in (2) for samples covering the following periods: (1) $1952(Q2) : 1984(Q2)$, (2) $1984(Q3) : 2014(Q4)$ and (3) $1952(Q2) : 2014(Q4)$. The estimates are summarized in Table 2.1 below. The dependent variable is $y_t = 100 \times \Delta \text{GNP}_t\%$. For each sample I use standard model selection criteria (AIC, BIC, HIC) and likelihood ratio tests to select the appropriate number of lags.

Column (1) in Table 2.1 gives estimates using data from the first sample, which covers the same

Table 1: MSM(2)–AR(ℓ) model with split samples

	(1)		(2)		(3)	
Period:	$1952(Q2) - 1984(Q2)$		$1984(Q3) - 2014(Q4)$		$1952(Q2) : 2014(Q4)$	
Model:	MSM(2)–AR(4)		MSM(2)–AR(0)		MSM(2)–AR(3)	
μ_0	1.366	(0.009)	0.665	(0.005)	0.901	(0.007)
μ_1	-0.670	(0.018)	-0.894	(0.020)	-0.713	(0.019)
σ_e	0.819	(0.006)	0.522	(0.004)	0.804	(0.003)
$p_{0\|0}$	0.907	(0.003)	0.949	(0.003)	0.952	(0.002)
$p_{1\|1}$	0.795	(0.007)	0.463	(0.017)	0.722	(0.007)
ϕ_1	0.093	(0.008)	–	–	0.213	(0.005)
ϕ_2	0.170	(0.008)	–	–	0.202	(0.005)
ϕ_3	-0.094	(0.008)	–	–	-0.163	(0.005)
ϕ_4	-0.138	(0.007)	–	–	–	–
Duration $S_t = 0$	10.6979		19.764		20.6228	
Duration $S_t = 1$	4.8780		1.8622		3.5971	
Log Likelihood	-184.260		-114.098		-327.953	–
T	124		120		246	

period as Hamilton's original paper. The specification with 4 lags is maintained based on model selection criteria but estimates are slightly different to those of the original paper (Hamilton 1989, Table 1; Kim and Nelson 1999, Chapter 4, Table 4.1). This is due to the new data having been revised and re-based. The probability of staying in the positive growth regime (boom) remains virtually the same at $p_{0|0} = 0.907$ and that of negative growth (recession) rises to $p_{1|1} = 0.795$, still close to the 0.75 estimate using the original data. This value of $p_{1|1}$ implies that the expected duration of a recession is Dur. $= \frac{1}{1-p_{1|1}} = 4.9$ quarters or approximately 15 months. In the data, the average duration of a recession over this period is 3.6 quarters or 11 months, so the model estimates overstates the length of a recession by up to a half a quarter which is a moderately large difference.

The estimates in column (2) are from the second sample; the period following the onset of the Great Moderation. Several observations are worth noting from these estimates. First, the average growth rate of GDP during booms is less than half what what is was in the earlier sample; falling to $\mu_0 = 0.665$ in comparison to $\mu_0 = 1.366$. Second, the volatility is much lower, as has been documented for many macroeconomic time-series since 1984. Third, the model selection and likelihood ratio tests indicate that the there are no longer dynamic effects on growth rate. Finally, estimates of transition probability in a recession $p_{1|1} = 0.463$ is much smaller than estimates from the earlier sample period which implies recessions have become less than half as short in the modern period (compared to the earlier period). However the average length of a recession has not changed, over the second sample period, recessions actually still last 3.6 quarters[6] as in the earlier period. However the average length of booms has doubled, from 13.86 to 29.67 quarters, so the second period model underestimates both the length of recessions and booms (all phis in column (2) are statistically equal to zero.). Finally, estimates in column (3) use the whole sample. For the mean growth rates, the full sample model equally weights the growth rates from the two sample periods as expected. The expected durations are close to the full sample averages: 3.60 and 19.30 (actual) versus 3.59 and 20.62 (estimated) for recessions and booms, respectively. The full sample model however fails to do a good job at identifying recessions as shown in Figure 1 below which plots the smoothed state inference for recession state: Prob$[S_t = 1 | y_{1:T}]$.

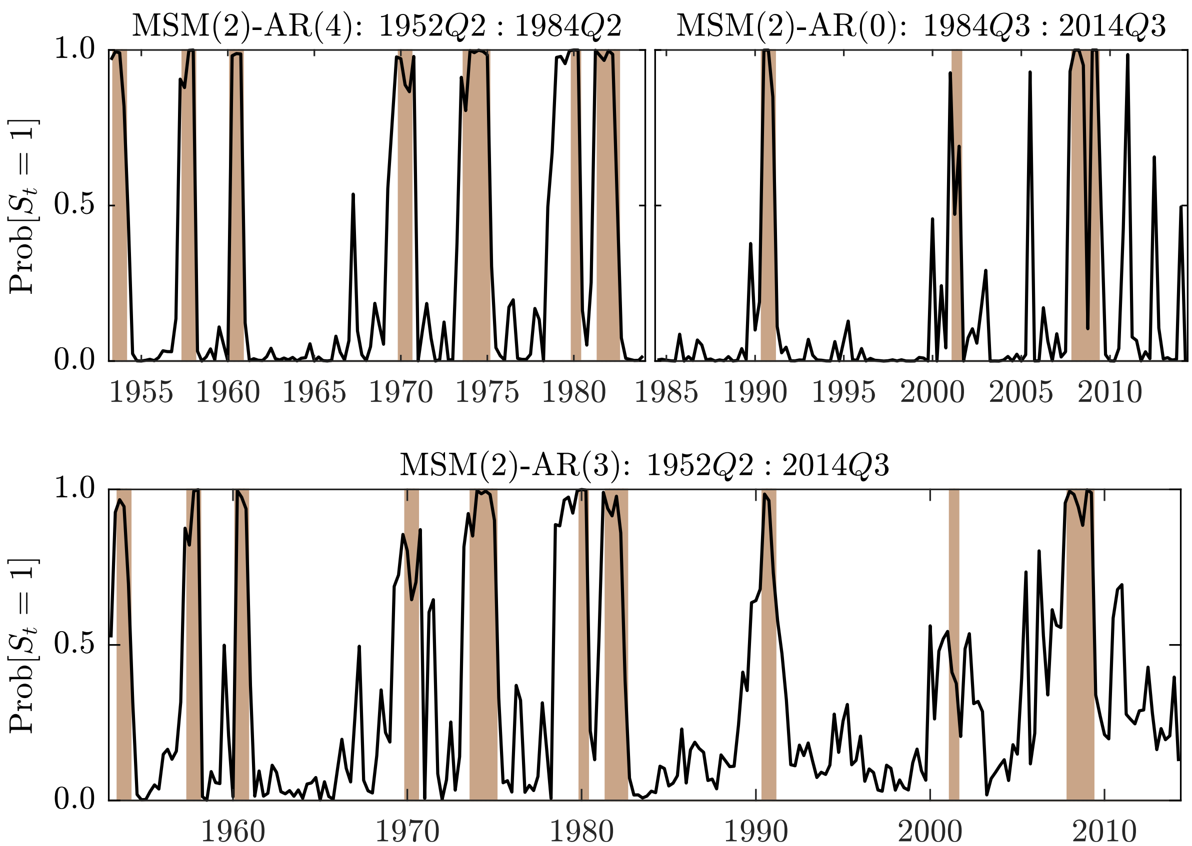

Figure 1: Smoothed Probability: MSM(2)-AR(l)

[6]The sample mean length of a recession, but based on only 3 recession events since the mid 1980s.

2.2 Asymptotic Consistency and Finite Sample Bias

Simulation evidence suggests that obtaining estimates close enough to the true data generating process for Markov switching models requires samples longer or larger than what is typically available for macroeconomic time series (Psaradakis and Sola 1998, Ho 2001). In this subsection, I discuss estimates from two papers (Bai and Wang 2011, Chen 2006) using two-state switching models to identify economic growth phases of the US economy. I compute the actual average growth rates over the periods identical to the recession (and boom) phases identified by the models and compare them to the estimates obtained by these authors. This comparison shows large differences between actual values of mean and volatility of GDP to the estimates obtained by the authors. For mean estimates, the source of the bias seems to small sample sizes which may make inference difficult (see footnote 9 below). For the volatility estimates, a proof of the inconsistency of estimators such as those implied by the Markov switching models can found in for example Spanos (2013, see also footnote 11)

The estimates of σ in columns (1) and (2) capture the Great Moderation. The selected number of lags in column (2) estimates show that growth is less persistent in the periods after $1984:Q2$. For the autoregressive model of order $\ell = 1$, specified by (2), the unconditional variance of is[7]:

$$\mathrm{Var}(y_t) = \frac{\pi_0 \pi_1 (\mu_0 - \mu_1)^2}{1 - \phi_1^2}(1 - \phi_1^2 - 2\phi_1(p_{0|0} + p_{1|1} - 1)) + \frac{\sigma_e^2}{1 - \phi_1^2} \tag{4}$$

which collapses to $\mathrm{Var}(y_t) = \pi_0 \pi_1 (\mu_0 - \mu_1)^2 + \sigma_e^2$ when $\phi_1 = 0$. For the second period estimates in Table 2.1, this formula gives an estimate of the sample standard deviation equal to 0.6804 which is close to the true value (the actual standard deviation of the series in 0.6828). The great moderation is observed as change in the unconditional variance of a series as illustrated in Figure 2 for the United States GNP.

The lower volatility of the series post-1984 also coincides with the lack of autoregressive terms as indicated by results in column (2) of Table 2.1. Note that in the estimates in column(3), model selection reduces the significant autoregressive terms to 3 compared to 4 in the earlier period and the whole sample model does a poor job of identifying recessions post 1984. As the discussion in the introduction indicated, Chen (2006), Doornik (2013) and Bai and Wang (2011) amongst others, have moved to specifications that correctly identify recession dates in the modern era by having two independent Markov chains: one for the switching mean growth rates and a second chain for the variance with an "absorbing" low volatility state that occurs around 1984. In all these

[7] See Appendix A for a derivation of this formula. See also Timmermann (2000) and Petričková (2014) for alternative representations of the moments of Markow switching models.

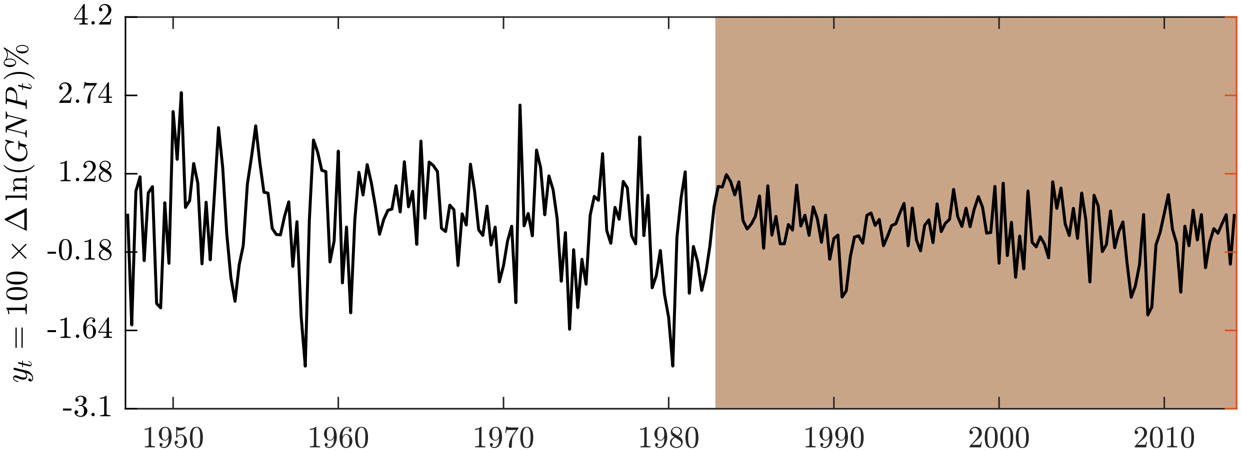

Figure 2: GNP Growth (Quarterly)

specifications, the lag length reduces either one (Chen 2006, Doornik 2013) or zero (Bai and Wang 2011). The failure to properly account for the changing persistence of the data across the volatility phases has large consequences on the estimated unconditional variances.

To see this, first consider Bai and Wang's (2011) estimates using data from 1947Q2:2006Q4. In their four-state regime switching model, the *last* two states are the recession and boom phases of the post-moderation era. Their estimates are as follows: $p_{0|0} = 0.8332$; $p_{1|1} = 0.9630$; $\mu_0 = 0.1716$; $\mu_1 = 0.8913$; $\pi_0 = \frac{(1-p_{1|1})}{(2-p_{1|1}-p_{0|0})} = 0.1816$; $\pi_1 = \frac{(1-p_{0|0})}{(2-p_{1|1}-p_{0|0})} = 0.8184$; $\sigma_e^2 = 0.1590$. These estimates imply that the unconditional variance of post-moderation data (using equation (4) with $\phi_1 = 0$) equals 0.2360. The actual variance for the data over this time period is 0.2961[8] which is a bias of -20.3%. For the mean growth rates, there is also considerable bias: in the post-moderation era, Bai and Wang's (2011) estimates of the mean growth rate during recessions in 0.1716 (which is positive), while actual growth rate during post-moderation recessions in their sample is -0.2616.[9] As another example, consider the estimates of Chen (2006, Table 4, page 98): $p_{0|0} = 0.733; p_{1|1} = 0.926; \mu_0 = -0.265; \mu_1 = 1.258; \pi_0 = \frac{(1-p_{1|1})}{(2-p_{1|1}-p_{0|0})} = 0.3595; \pi_1 = \frac{(1-p_{0|0})}{(2-p_{1|1}-p_{0|0})} = 0.6405; \phi_1 = 0.127; \sigma_e^2 = 0.928$, which gives an estimate of the unconditional variance in the pre-moderation era equal to 1.2650. Again this measure underestimates the actual variance over the pre-moderation period (1.2650 < 1.4229).

While the asymptotic theory on ML estimators of Markov switching model parameters establish their consistency and unbiasedness properties (Krishnamurthy and Ryden 1998, Douc,

[8]This is the unconditional variance over the period $1984Q3 : 2006Q4$. The data are downloaded from the Journal of Applied Econometrics Data Archive: http://qed.econ.queensu.ca/jae/2011-v26.5/bai-wang

[9] This critique may be unfair given the length of their sample: there are only 5 quarters when the US economy is in recession during this period.

Moulines and Rydén 2004), the preceding discussion suggests that estimates of the unconditional variance from Markov switching models are actually biased. This is not surprising given simulation evidence on the finite sample distribution of ML estimates from Markov switching models. These have been studied by Psaradakis and Sola (1998, Table 1a and 1b) and Ho (2001), who both show that the small sample estimates are largely biased and caution against relying on the asymptotic properties MLE estimators for these type of models.[10] Neither are the results from finite sample simulations surprising if one considers theoretical results from the econometric/statistical literature,[11] which have been ignored by researchers implementing the extended Markov switching models with volatility changes.

In the next section, I propose a model specification that accounts for both the change in mean growth rates across business cycle related regimes and the Great Moderation while allowing for potentially different autocorrelation structures across the volatility regimes. While the specification is related to those of Chen (2006), Doornik (2013), Bai and Wang (2011) and Chauvet and Su (2014) it differs in three important dimensions. First, in order to obtain consistent estimates of the volatility within the variance regimes, I "concentrate" the likelihood function in a manner similar to that described by Basu (2011, Sec. 8). Second, the likelihood concentration allows for the mean growth rate to have different lag lengths between the different volatility regimes. While the lag structure selection approach is not new,[12] there is no model in the economic or statistical literature that considers independent variance shifts that determine the AR structure of the underlying time series which is driven by another hidden Markov process. Third, in the implementation of this mixing AR structure model, I account for the change in the duration and mean growth rates during booms (determined by the mean shift chain) by adding a parameter that links the mean shift chain to the variance shift chain.

After specifying this model, I describe a likelihood "profiling" and filtering technique that exploits its natural Bayesian structure to "integrate out" the incidental parameter and facilitate estimation. This form of the likelihood facilitates a two-stage estimation procedure originally due

[10] Simulation results included in Krishnamurthy and Ryden (1998, Section 5.1) do not capture this problem because their models do not have an intercept which would induce the incidental/nuisance parameter problem and bias the MLE estimators.

[11] See Example 2 in Lancaster (2000) or Basu (2011) and Example 3 in Neyman and Scott (1948). These examples describe the inconsistency of ML estimators of the variance or "structural parameter" when the data follow k probability laws. In the context of Markov switching models, we are trying to estimate a common σ from k distributions with $\mu_i \neq \mu_j \forall i, j = 1, \ldots, k$.

[12] In the time-series/statistics/engineering literature, Glasbey (2001), Ailliot et al. (2006), Kalliovirta et al. (2015) use models where the states choose AR models of different lags. Such models are also discussed by Lu and Berliner (1999) and also Douc, Moulines and Stoffer (2014, Chapter 9). See also Psaradakis and Spagnolo (2006) and Ekner (2014) for examples in the STAR/SETAR type models.

to Sir Ronald Fisher (see e.g. Hald 2006, Sec. 19.3). In the first stage, I estimate the parameters of primary interest (the mean growth rates across recessions and booms). In the second stage, I hold fixed the first stage parameters and estimate the volatility related "nuisance" parameters. This approach obtains consistent estimates of variances while identifying the Great Moderation volatility switch.

3 A Double Mixture Autoregressive Model

Let $y_t = 100 \times \Delta\log(\text{GNP}_t)\%,\ t = 1:T$ represent a time series such as that displayed in Figure 2. Let S_t^m and S_t^v represent, respectively, the mean and variance regime indicators. Here $S_t^v = \{0,1\}$ captures volatility changes characterising the Great Moderation while $S_t^m = \{0,1\}$ represents shifts in the growth rate of GNP related to business cycles. The regimes S_t^m and S_t^v are each the outcome an independent first order Markov chain with transition matrices: $\mathbf{P^m} = p^m_{j|i}$ and $\mathbf{P}^v = p^v_{j|i}$, respectively. The two components correspond to a restricted four regime model, with state $S_t = S_t^m \times S_t^v$ and transition matrix:

$$\mathbf{P} = \begin{array}{c|cccc} & \multicolumn{2}{c}{S_t^v = 0} & \multicolumn{2}{c}{S_t^v = 1} \\ & S_t^m = 0 & S_t^m = 1 & S_t^m = 0 & S_t^m = 1 \\ \hline & S_t = 0 & S_t = 1 & S_t = 2 & S_t = 3 \\ \hline S_{t-1} = 0 & & & & \\ S_{t-1} = 1 & \multicolumn{2}{c}{p^v_{0|0}\mathbf{P^m}} & \multicolumn{2}{c}{p^v_{1|0}\mathbf{P^m}} \\ S_{t-1} = 2 & & & & \\ S_{t-1} = 3 & \multicolumn{2}{c}{p^v_{0|1}\mathbf{P^m}} & \multicolumn{2}{c}{p^v_{1|1}\mathbf{P^m}} \end{array} = \mathbf{P^m} \otimes \mathbf{P}^v \tag{5}$$

The state S_t defines a dynamic linear model:

$$y_t = \begin{cases} \mu_{S_t^m} + \sum_{l=1}^{\ell(S_t^v)} \phi_l \left(y_{t-l} - \mu_{S_{t-l}^m}\right) + \sigma_{S_t^v} e_t, & S_t^m,\ S_t^v = \{0,1\} \\ \mu_{S_t} + \sum_{l=1}^{\ell(S_t)} \phi_l \left(y_{t-l} - \mu_{S_{t-l}}\right) + \sigma_{S_t} e_t, & S_t = \{0,1,\dots,K\} \end{cases} \quad e_t \sim \text{iid } N(0,1) \tag{6}$$

with time varying intercepts $\mu_{S_t^m}$ and volatilities $\sigma_{S_t^v}$. The lag length $\ell(S_t^v)$ in (6) is potentially changing across variance regimes. This is the first innovation in the paper. In the original model of Hamilton (1989), there are no volatility changes with the state and σ can be thought of as a nuisance parameter in the sense of Elliott, Müller and Watson (2015) which we are not interested in. In the present context, we are interested in modelling the business cycle related shifts in mean growth rates while treating the change in volatility post 1984 as *incidental* shift parameters in the sense of Neyman and Scott (1948, Example 1).[13] In the next two subsections, I describe an estimation and filtering procedure that implements Basu's intimation "*to fix a prior, compute the posterior, integrate out the nuisance parameter from the posterior, to arrive at the posterior marginal distribution of the parameter of interest, and then let the statistical argument rest on the posterior marginal distribution*"(Basu 2011, paragraph 10, Section 1). I finally describe the two–stage MLE procedure in the third subsection.

3.1 Likelihood function

In the standard ML approach, if S_t were observed, the parameters $\{\mu_{S_t}, \sigma_{S_t}\}_{S_t=0}^{3}$ of equation (6) would be estimated by maximizing the log-likelihood function

$$\ln L = \sum_{t=1}^{T} \ln f(y_t|S_t)$$

where $f(y_t|S_t)$ is the conditional distribution of y_t given the state S_t

$$f(y_t|S_t) = \frac{1}{\sqrt{2\pi}\sigma_{S_t}} \exp\left(-\frac{1}{2}\left\{\frac{y_t - \mu_{S_t} - \sum_{l=1}^{\ell}\phi_l\left(y_{t-l} - \mu_{S_{t-l}}\right)}{\sigma_{S_t}}\right\}^2\right).$$

However since the state S_t is not observed, one would begin by considering the joint density of y_t, S_t and information up to time $t-1$, denoted by $\psi_{t-1} = y_{1:t-1}$. The marginal density $f(y_t|S_t, \psi_{t-1})$ used in the likelihood, is then obtained by integrating out S_t from $f(y_t, S_t|\psi_{t-1})$:

$$f(y_t|\psi_{t-1}) = \sum_{S_t} f(y_t|S_t, \psi_{t-1}) f(S_t|\psi_{t-1}) = \sum_{k=0}^{K} f(y_t|S_t, \psi_{t-1}) P[S_t = k|\psi_{t-1}]$$

where $P[S_t = k|\psi_{t-1}]$, is the *prior* state probability[14] and $f(y_t|\psi_{t-1})$ is the *posterior* likelihood having observed y_t. For the $K = 4$ state case described by (5) without lags the log-likelihood

$$\ln L = \sum_{t=1}^{T} \ln\left(\sum_{k=0}^{4} f(y_t|S_t, \psi_{t-1}) \times P[S_t = k|\psi_{t-1}]\right) \tag{7}$$

would be maximized to obtain the parameters $\theta = \left\{p_{0|0}^{m}, p_{1|1}^{m}, p_{0|0}^{v}, p_{1|1}^{v}, \mu_0, \mu_1, \sigma_0, \sigma_1\right\}$.

[13] Note that we could alternatively treat the μ's as the nuisance/incidental parameters if we were interested in estimating the σ's (see e.g. Spanos 2013)

[14] See Hamilton (1994, p. 692) or Kim and Nelson (1999, p. 63) for textbook illustrations of deriving the updating formula.

To implement Basu's (2011) idea of integrating out the nuisance parameter, start with (7) having fixed a prior $P[S_t = k|\psi_{t-1}]$ and computed the posterior $f(y_t|S_t, \psi_{t-1})$. We now want to integrate out the nuisance parameter ($\sigma_{S_t^v}$) to arrive at the posterior marginal distribution of the parameter of interest ($\mu_{S_t^m}$):

$$f(y_t|S_t^m, \psi_{t-1}) = \sum_{k=0}^{1} f(y_t|S_t^m, S_t^v = k, \psi_{t-1}) \times P[S_t^v = k|S_t^m, \psi_{t-1}] \tag{8}$$

where the predictive density for the mean is:

$$f(y_t|S_t^m, S_t^v = k, \psi_{t-1}) = \sum_{k_m} f(y_t|S_t^m = k_m, S_t^v = k, \psi_{t-1}) \times P[S_t^m = k_m|S_t^v = k, \psi_{t-1}] \tag{9}$$

with k_m indexing the (possibly expanded) number of states of S_t^m in the presence of lags (see Section 3.2 below). The final likelihood is given by:

$$\ln L = \sum_{t=1}^{T} \ln\left(\sum_{k=0}^{1} f(y_t|S_t^m, S_t^v = k, \psi_{t-1}) \times P[S_t^v = k|S_t^m, \psi_{t-1}]\right) \tag{10}$$

3.2 Filtering

In order to evaluate equations (8) and (9), we need to compute the prior probabilities $P[S_t^m = k_m|S_t^v = k, \psi_{t-1}]$ and $P[S_t^v = k|S_t^m, \psi_{t-1}]$. Given a value of the prior $P[S_t = k|\psi_{t-1}]$ and the likelihood $f(y_t|S_t = l, \psi_{t-1})$, we can compute the posterior, $P(S_t = k|\psi_t)$ using Bayes's law as:

$$P(S_t = k|\psi_t) = \frac{P(S_t = k|\psi_{t-1}) f(y_t|S_t = k, \psi_{t-1})}{f(y_t|\psi_{t-1})} \tag{11}$$

where $\psi_t = \{\psi_{t-1}, y_t\}$. Following Cox and Miller (1965), expand the state space to account for the number of lags to obtain $n_m = 2^{(\ell(S_t^v)+1)}$ states for each $m = S_t^v = \{0, 1\}$ volatility state. Collect the set of possible conditional densities $f(y_t|S_t^m, S_t^v = k)$ in a $(1 \times n_m)$ row vector $\eta_{m,t}$. Likewise, define the $(1 \times n_m)$ vector $\xi_{m,t|t-1}$ whose n_m^{th} element is $P[S_t^m = k_m|S_t^v = k, \psi_{t-1}]$. Taking the inference on $\xi_{m,t-1|t-1}$, the update $\xi_{m,t|t}$ can be calculated after observing y_t. Following Hamilton (1994, p. 692) this is accomplished by using the transition matrix $\mathbf{P_m^m}$ and applying (11) to obtain the predicted and filtered probabilities:

$$\xi_{m,t|t-1} = \xi_{m,t-1|t-1}\mathbf{P}_m^{\mathbf{m}} \tag{12}$$

$$\xi_{m,t|t} = \frac{\xi_{m,t|t-1} \odot \eta'_{\mathbf{m},t}}{\sum \xi_{m,t|t-1} \odot \eta'_{\mathbf{m},t}} \tag{13}$$

where $\mathbf{P}_m^{\mathbf{m}}$ is an expansion similar to (3) of the state matrix $\mathbf{P}^{\mathbf{m}}$ defined by (5), $\odot$ is the Hadamard product and the summation is over n_m states.

To update $P[S_t^v = k|S_t^m, \psi_{t-1}]$, collect these probabilities into the (4×1) vector $\xi_{v,t|t-1}$ and collapse each $\eta_{\mathbf{m},t}$ over their respective lags to obtain the likelihood $f(y_t|S_t^m, \psi_{t-1})$ which are collected in a (4×1) vector $\eta_{v,t}$. Again, using the transition matrix and applying (11), we obtain the predicted and filtered(updated) state probabilities:

$$\xi_{v,t|t-1} = \xi_{v,t-1|t-1}\mathbf{P} \tag{14}$$

$$\xi_{v,t|t} = \frac{\xi_{v,t|t-1} \odot \eta_{v,t}}{\sum \xi_{v,t|t-1} \odot \eta_{v,t}} \tag{15}$$

where $\mathbf{P}$ is defined by (5) and $\odot$ is the Hadamard product. The elements of the transition matrices $\mathbf{P}_m^{\mathbf{m}}$ and $\mathbf{P}^v$are specified by:

$$\begin{aligned}
p_{0,0|0}^m &= \frac{\exp(p_0)}{1+\exp(p_0)}, & p_{0,1|1}^m &= \frac{\exp(p_1)}{1+\exp(p_1)}, & &\text{for} \quad \mathbf{P}_0^{\mathbf{m}} \\
p_{1,0|0}^m &= \frac{\exp(p_0+p_{0d})}{1+\exp(p_0+p_{0d})}, & p_{1,1|1}^m &= \frac{\exp(p_1+p_{1d})}{1+\exp(p_1+p_{1d})}, & &\text{for} \quad \mathbf{P}_1^{\mathbf{m}} \\
p_{0|0}^v &= \frac{\exp(q_0)}{1+\exp(q_0)}, & p_{1|1}^v &= \frac{\exp(q_1)}{1+\exp(q_1)}, & &\text{for} \quad \mathbf{P}^v
\end{aligned} \tag{16}$$

where $p_0, p_1, r_0, r_1, q_0, q_1$ are unconstrained parameters and the transition probabilities are constrained within the $[0,1]$ interval. The filter at $t = 1$, is started by setting $\xi_{v,t-1|t-1}$ equal to the unconditional probability vector:

$$\pi_v = ((\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\begin{pmatrix}\mathbf{0}_{(K\times1)}\\ 1\end{pmatrix}; \quad \text{where } A = \begin{pmatrix}\mathbf{I}_K - \mathbf{P}^v\\ \mathbf{1}_{(1\times K)}\end{pmatrix} \text{ and } K = \text{ length } (\mathbf{P}^v).$$

with a similar procedure for every m of $\xi_{m,t-1|t-1}$. The parameters p_{0d} and p_{1d} in (16) allow for the possibility of different transition probabilities when the volatility state is $S_t^v = 1$. These are similar to a dummy variable coefficient if the volatility state were priorly known. I add a similar "dummy" coefficient for the mean growth rates in the second volatility state: μ_{0d}, μ_{1d}. This completes the characterization of computing the likelihood (10) which is maximized over the parameter vector $\theta = \left\{p_0, p_1, p_{0d}, p_{1d}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}, \sigma_0, \sigma_1, \phi_1^0, \ldots, \phi_{\ell_0}^0, \phi_1^1, \ldots, \phi_{\ell_1}^1\right\}$ and ℓ_0 need not equal ℓ_1.

3.3 Two Stage Estimation

In the first stage, freely estimate all the model parameters: $\theta = \{p_0, p_1, p_{0d}, p_{d1}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}, \sigma_0, \sigma_1, \phi_1^0, \ldots, \phi_{\ell_0}^0, \phi_1^1, \ldots, \phi_{\ell_1}^1\}$ which gives consistent estimates of the parameters: $\theta' = \{p_0, p_1, p_{0d}, p_{d1}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}\}$. In the second stage, hold the consistently estimated parameters θ' fixed and use the integrated out likelihood to estimate second stage parameters $\beta = \left\{\sigma_0, \sigma_1, \phi_1^0, \ldots, \phi_{\ell_0}^0, \phi_1^1, \ldots, \phi_{\ell_1}^1\right\}$.

4 Empirical Results

I estimate the model described in the previous section by using the non-linear programming solver *fminunc* in MATLAB R2016b. After finding a solution, standard errors are obtained from the inverted Hessian matrix using the delta method.[15] I consider two time series at first, real GNP and industrial production. The data are downloaded from the online Federal Reserve Economic Data (FRED).The time series are of quarterly frequency, from 1947:Q2 – 2014:Q2. All series are deflated using the consumer price index (CPI).

4.1 First Stage Estimates: μ_m's

I estimate the Conditional Markov Chain (CMC) model of Bai and Wang (2011) and compare it to the Double Mixture Autoregressive (DM-AR(ℓ_0, ℓ_1)) model proposed here for the quarterly real GNP growth data shown in Figure 2. I then compare the estimates of mean growth rates and unconditional variances to the actual values over periods defined by the Great Moderation and NBER business cycles dates. Table 4.1 gives results of the first stage estimation. I use standard model selection criteria (AIC, BIC, HIC) to select amongst models of different lag lengths.

In column (1) of Table 4.1 are the estimates of the CMC model of Bai and Wang. Columns (2) & (3) are DM–AR(ℓ_0, ℓ_1) models estimates. There is not much to separate the models but AIC and HIC suggest the model in (3) does a better job of representing the data and the null of no autoregressive terms in reject using Wald and likelihood ratio tests. In Bai and Wang's (2011)

[15] See Martin, Hurn and Harris (2012, p. 107) for a discussion.

Table 2: First Stage Estimates

	(1)		(2)		(3)	
Model	CMC		DM–AR(0,0)		DM–AR(0,1)	
μ_0	1.3766	(0.0069)	1.3758	(0.0069)	1.3711	(0.0068)
μ_1	-0.5514	(0.0135)	-0.5445	(0.0132)	-0.6067	(0.0131)
μ_{0d}	0.6647	(0.0034)	0.6648	(0.0034)	0.661	(0.0030)
μ_{1d}	-0.8950	(0.0138)	-0.896	(0.0137)	-0.9629	(0.0144)
σ_0	0.9554	(0.0042)	0.958	(0.0042)	0.9489	(0.0041)
σ_1	0.5258	(0.0025)	0.5258	(0.0025)	0.5219	(0.0023)
$p^m_{0,0\|0}$	0.8949	(0.0023)	0.9009	(0.0023)	0.9057	(0.0022)
$p^m_{0,1\|1}$	0.7337	(0.0054)	0.729	(0.0052)	0.7191	(0.0053)
$p^m_{1,0\|0}$	0.95	(0.0017)	0.677	(0.0095)	0.6858	(0.0091)
$p^m_{1,1\|1}$	0.4664	(0.0113)	0.2454	(0.0097)	0.2657	(0.0103)
$p^v_{0\|0}$	1	(0.0000)	0.9933	(0.0004)	0.9933	(0.0004)
$p^v_{1\|1}$	0.9916	(0.0005)	1	(0.0000)	1	(0.0000)
ϕ^1_1	–	–	–	–	-0.1503	(0.0061)
Dur.: $S^m_t = 1\|S^v_t = 0$	3.7555		3.6894		3.5606	
Dur.: $S^m_t = 1\|S^v_t = 1$	1.8742		1.3252		1.3619	
Log Likelihood	-357.89		-357.59		-354.42	
AIC	2.7501		2.7479		2.7419	
BIC	3.1601		3.1578		3.1873	
HIC	2.8145		2.8123		2.8119	
Wald	–		–		614.9011	
p-val$_w$	–		–		0.0000	
LR.	–		–		6.3473	
p-val$_{LR}$	–		–		0.0118	
T	269		269		268	

implementation of the CMC model using real GDP, the models ability to properly identify business cycle turning point broke down when any lags were included, even though they rejected the null of no auto–regression terms. However, for the DM-AR models, I can still successfully identify recessions as shown in Figure 3 using the first stage estimation. Figure 4 shows inference on the onset of the Great Moderation using the DM–AR model, which happens on or about $1984:Q2$.

Table 3 gives an evaluation of the bias of model implied estimates of unconditional moments of the time series. The column True Value contains "true" estimates of the means and variances of the time series over the periods of interest based on our knowledge of published NBER recession dates and the date of the onset of the Great Moderation as identified in the literature. The RMSEs in the last two rows show that as one proceeds to the model that best represents the data, the bias of the model implied unconditional moments increases – which is an undesirable property. The second stage estimation should help resolve this problem.

Table 3: Unconditional Moments: Actual vs. Model Implied (1^{st} Stage)

		Model Estimates		
	"True Value"	CMC	DM–AR(0,0)	DM–AR(0,1)
μ_0	1.1605	1.3766	1.3758	1.3711
μ_1	-0.5781	-0.5514	-0.5445	-0.6067
μ_{0d}	0.7028	0.6647	0.6648	0.661
μ_{1d}	-0.6061	-0.895	-0.896	-0.9629
σ_0^2	1.7140	1.66702	1.64077	1.6365
σ_1^2	0.5106	0.4670	0.7878	0.8115
RMSE_μ		0.1819	0.1823	0.2087
RMSE_{σ^2}		0.0640	0.1442	0.1580

4.2 Second Stage Estimates: σ_v^2's

We now hold fixed the set of parameters relating to the μ_m's and re-estimate the model. The results in Table 4 are the best model estimates based on a combination of the three standard information criteria: AIC, BIC and HIC. In selecting the three specifications shown, I run through combinations of (ℓ_0, ℓ_1) for each $\ell_i = \{0, 1, 2, 3, 4\}, i = \{0, 1\}$, i.e. 25 possible lag structure configurations. The three models shown have the minimum of combinations of the three model selection criteria. The best model is in column (4) with 2 significant lags in the moderation era and zero lags in the pre-moderation: DM–AR(0,2). Based on results in Table 3, we would ordinarily expect the bias in the model implied unconditional moments to increase as we increase the lag structure: but MSEs of

Table 4: Second Stage Estimates

Model	DM-AR(0,0)		DM-AR(1,0)		DM-AR(1,1)		DM-AR(0,2)	
σ_0	0.9579	(0.0040)	0.9549	(0.0041)	0.9541	(0.0041)	0.9495	(0.0040)
σ_1	0.5258	(0.0024)	0.5258	(0.0024)	0.5206	(0.0023)	0.5099	(0.0022)
$p^v_{0\|0}$	0.9933	(0.0004)	0.9933	(0.0004)	0.9933	(0.0004)	0.9932	(0.0004)
$p^v_{1\|1}$	1.0000	(0.0000)	1.0000	(0.0000)	1.0000	(0.0000)	1.0000	(0.0000)
ϕ^1_1	–		0.1124	(0.0066)	0.1124	(0.0066)	–	
ϕ^1_2	–		–		–	(0.0061)	–	
ϕ^2_1	–		–		-0.1452		-0.0767	(0.0061)
ϕ^2_2	–		–		–		0.2540	(0.0068)
Log Like	-357.59		-354.98		-353.961		-349.34	
AIC	2.6884		2.6864		2.6863		2.6617	
BIC	2.8250		2.8577		2.8918		2.8679	
HIC	2.7099		2.7134		2.7186		2.6941	
Wald			287.371		13.398		296.008	
pval$_w$			0		0.0012		0	
LR.			5.2137		7.2578		16.4978	
pval$_{LR}$			0.0224		0.0265		0.0003	
T	269		268		268		267	

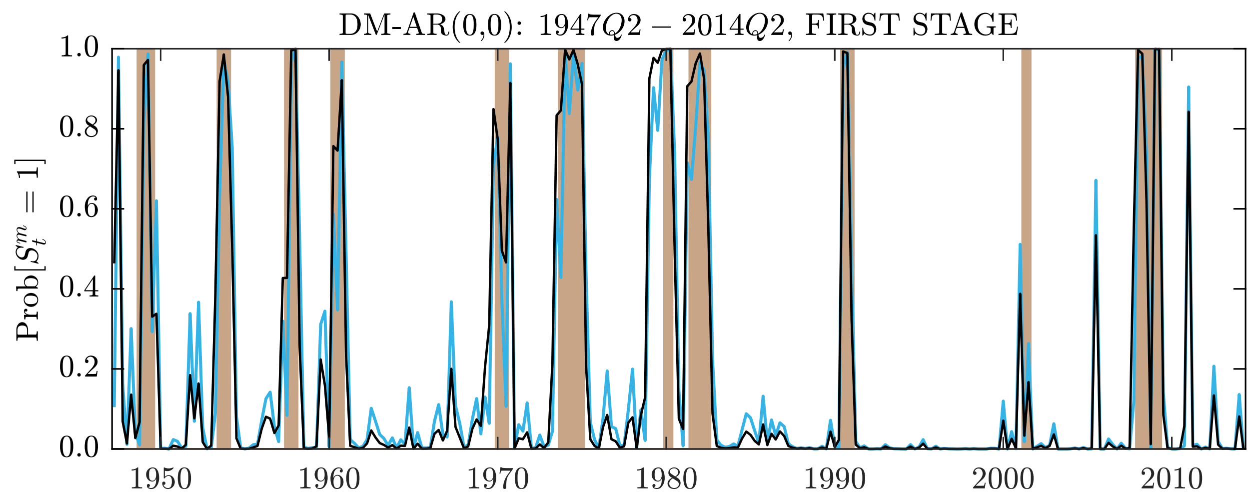

Figure 3: Filtered (light blue) and Smoothed (black) Prob. of Recession

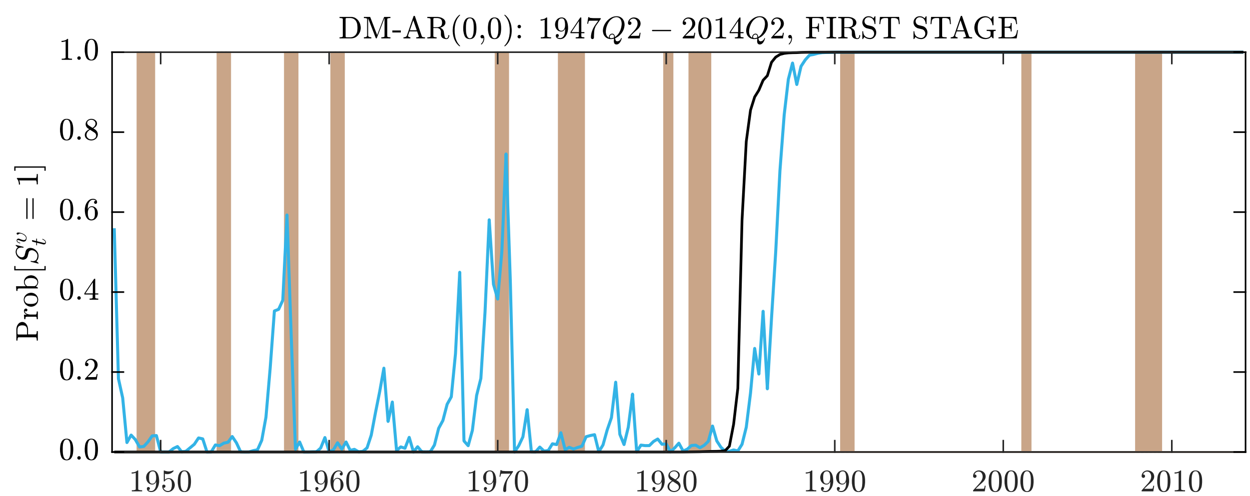

Figure 4: Filtered (light blue) and Smoothed (black) Prob. of Low Volatility Regime

unconditional moments in Table 5 indicate otherwise.[16] As we get to the best model in representing the data, the bias of the unconditional variance estimates actually falls. This is the true innovation of my model and estimation procedure.

Finally, in Figure 5, I show inference on the state $S_t^\upsilon = 1$ after the second stage where the parameters q_0, q_1 have been re-estimated and together with the variances. The second stage inference in virtually identical to the first stage which is what we would expect given that the only inconsistent parameter from the first–stage are the σ_υs.

The main achievement of our model specification and estimation procedure is the ability to

[16]Statistically, this should not happen if the higher lag model is the best representation of the data, but we are faced with an inconsistent estimator if we estimate the model in one go.

Table 5: Unconditional Variance: Actual vs. Model Implied(2^{nd} Stage)

		Model Estimates			
	True Value	DM-AR(0,0)	DM-AR(1,0)	DM-AR(1,1)	DM-AR(0,2)
σ_0^2	1.7140	1.6670	1.5351	1.5335	1.6247
σ_1^2	0.5106	0.4670	0.4670	0.4864	0.4849
$\text{Bias}_{\sigma_0^2}$		-2.7422	-10.4391	-10.5293	-5.2136
$\text{Bias}_{\sigma_1^2}$		-8.5293	-8.5390	-4.7332	-5.0238
RMSE		0.0453	0.1302	0.1288	0.0657

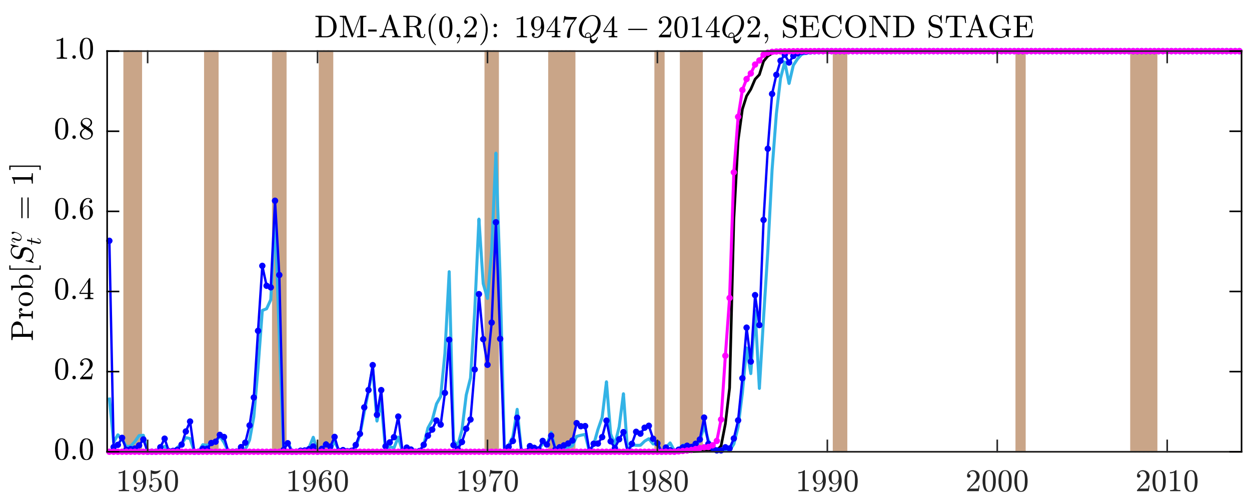

Figure 5: Filtered (1^{st} stage: light blue, 2^{nd} stage: blue dotted) and Smoothed (1^{st} stage: black, 2^{nd} stage: magenta dotted) Probability of Low Volatility Regime

reduce the bias in unconditional moments when proper accounting of the lag structure in our data has been undertaken.

5 Conclusion

In this paper, I have revisited Hamilton's regime switching model for business cycle dating. I began by showing that the unconditional moments implied by finite sample estimates from modern implementations of this model are characterized by large biases. Appealing to almost century old statistical theory, I show that by exploiting the Bayesian nature of the model, we can use a two–stage MLE strategy that maintains the model's usefulness for business cycle dating while resolving the inconsistency problem exacerbated by a proper accounting of the lag structure of the time series of interest.

My results have potential applications in a wide range of econometric applications including the identification by heteroskedasticity literature advanced by Netsunajev (2013) and Lütkepohl and Velinov (2016) amongst others,[17] and the asset pricing and learning literatures of Lettau, Ludvigson and Wachter (2008).[18]

[17] This literature requires estimation of the variance across regimes, so the results here directly apply.

[18] This literature requires the estimation of the variance of say consumption growth, which measures risk and determines the expected return to the aggregate stock market.

References

Ailliot, P., Monbet, V. and Prevosto, M.: 2006, An autoregressive model with time-varying coefficients for wind fields, *Environmetrics* **17**(2), 107–117.
URL: *http://doi.wiley.com/10.1002/env.753* (cited on page 8)

Bai, J. and Wang, P.: 2011, Conditional Markov chain and its application in economic time series analysis, *Journal of Applied Econometrics* **26**(5), 715–734.
URL: *http://doi.wiley.com/10.1002/jae.1140* (cited on page 1, 6, 7, 8, 13)

Barnett, W. A., Chauvet, M. and Leiva-Leon, D.: 2016, Real-time nowcasting of nominal GDP with structural breaks, *Journal of Econometrics* **191**(2), 312–324.
URL: *http://dx.doi.org/10.1016/j.jeconom.2015.12.004* (cited on page 1)

Basu, D.: 2011, On the Elimination of Nuisance Parameters, *Selected Works of Debabrata Basu*, Springer New York, New York, NY, pp. 279–290.
URL: *http://link.springer.com/10.1007/978-1-4419-5825-9_26* (cited on page 2, 8, 10, 11)

Boldin, M. D.: 1996, A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle, *Studies in Nonlinear Dynamics and Econometrics* **1**(1), 35–46.
URL: *http://www.degruyter.com/view/j/snde.1996.1.1/snde.1996.1.1.1010/snde.1996.1.1.1010.xml* (cited on page 1)

Bullard, J. and Singh, A.: 2012, Learning and the great moderation, *International Economic Review* **53**(2), 375–397.
URL: *http://doi.wiley.com/10.1111/j.1468-2354.2012.00685.x* (cited on page 1)

Chang, Y., Choi, Y. and Park, J. Y.: 2017, A new approach to model regime switching, *Journal of Econometrics* **196**(1), 127–143.
URL: *http://dx.doi.org/10.1016/j.jeconom.2016.09.005 http://linkinghub.elsevier.com/retrieve/pii/S030440761*
(cited on page 2)

Chauvet, M. and Su, Y.: 2014, Nonstationarities and Markov Switching Models, *Recent Advances in Estimating Nonlinear Models*, Springer New York, New York, NY, pp. 123–146.
URL: *http://link.springer.com/10.1007/978-1-4614-8060-0_7* (cited on page 1, 8)

Chen, S. W.: 2006, Simultaneously modeling the volatility of the growth rate of real GDP and determining business cycle turning points: Evidence from the U.S., Canada and the UK, *Mathematics and Computers in Simulation* **71**(2), 87–102. (cited on page 1, 6, 7, 8)

Cox, D. R. D. R. and Miller, H. D. H. D.: 1965, *The theory of stochastic processes.*, Methuen. (cited on page 11)

Doornik, J. A.: 2013, A Markov-switching model with component structure for US GNP, *Economics Letters* **118**(2), 265–268.
URL: *http://www.journals.cambridge.org/abstract_S1365100514000455* (cited on page 1, 6, 7, 8)

Douc, R., Moulines, É. and Rydén, T.: 2004, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *Annals of Statistics* **32**(5), 2254–2304. (cited on page 7)

Douc, R., Moulines, E. and Stoffer, D.: 2014, *Nonlinear Time Series: Theory, Methods and Applications with R Examples*.
URL: *https://www.crcpress.com/Nonlinear-Time-Series-Theory-Methods-and-Applications-with-R-Examples/Douc-Moulines-Stoffer/p/book/9781466502253* (cited on page 8)

Ekner, L. E.: 2014, Cointegration and Regime Switching Dynamics in Macroeconomic Applications. (cited on page 8)

Elliott, G., Müller, U. K. and Watson, M. W.: 2015, Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis, *Econometrica* **83**(2), 771–811.
URL: *https://www.econometricsociety.org/doi/10.3982/ECTA10535* (cited on page 9)

Filardo, A. J. and Gordon, S. F.: 1998, Business cycle durations, *Journal of Econometrics* **85**(1), 99–123.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0304407697000961* (cited on page 2)

Franke, J.: 2012, Markov Switching Time Series Models, *Time Series Analysis: Methods and Applications*, Vol. 30, pp. 99–122.
URL: *http://www.sciencedirect.com/science/article/pii/B9780444538581000053 http://linkinghub.elsevier.com/retrieve/pii/B9780444538581000053* (cited on page 1)

Glasbey, C. A.: 2001, Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*

50(2), 143–154.
URL: *http://doi.wiley.com/10.1111/1467-9876.00225* (cited on page 8)

Hald, A.: 2006, The Revolutionary Paper, 1922, *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713âĂŞ1935*, Springer New York, New York, NY, pp. 175–183.
URL: *http://link.springer.com/10.1007/978-0-387-46409-1_19* (cited on page 9)

Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series, *Econometrica* **57**(2), 357.
URL: *http://www.jstor.org/stable/1912559?origin=crossref* (cited on page 1, 2, 4, 9)

Hamilton, J. D.: 1990, Analysis of time series subject to changes in regime, *Journal of Econometrics* **45**(1-2), 39–70.
URL: *http://linkinghub.elsevier.com/retrieve/pii/0304407690900939* (cited on page 3)

Hamilton, J. D.: 2011, Calling recessions in real time, *International Journal of Forecasting* **27**(4), 1006–1026.
URL: *http://dx.doi.org/10.1016/j.ijforecast.2010.09.001* (cited on page 1)

Hamilton, J. D.: 2016, Macroeconomic Regimes and Regime Shifts, *Handbook of Macroeconomics*, Vol. 2, pp. 163–201.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1574004816000057* (cited on page 1)

Hamilton, J. D. J. D.: 1994, *Time series analysis*, Princeton University Press.
URL: *http://press.princeton.edu/titles/5386.html* (cited on page 10, 11)

Ho, T.-W.: 2001, Finite-sample properties of the bootstrap estimator in a Markov-switching model, *Journal of Applied Statistics* **28**(7), 835–842.
URL: *http://www.tandfonline.com/doi/abs/10.1080/02664760120074924* (cited on page 1, 6, 8)

Justiniano, A. and Primiceri, G. E.: 2008, The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3), 604–641.
URL: *http://pubs.aeaweb.org/doi/10.1257/aer.98.3.604* (cited on page 1)

Kalliovirta, L., Meitz, M. and Saikkonen, P.: 2015, A Gaussian Mixture Autoregressive Model for Univariate Time Series, *Journal of Time Series Analysis* **36**(2), 247–266.
URL: *http://doi.wiley.com/10.1111/jtsa.12108* (cited on page 8)

Kim, C.-J. and Nelson, C. R.: 1999, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press. (cited on page 1, 2, 4, 10)

Krishnamurthy, V. and Ryden, T.: 1998, Consistent Estimation of Linear and Non-linear Autoregressive Models with Markov Regime, *Journal of Time Series Analysis* **19**(3), 291–307.
URL: *http://doi.wiley.com/10.1111/1467-9892.00093* (cited on page 7, 8)

Lancaster, T.: 2000, The incidental parameter problem since 1948, *Journal of Econometrics* **95**(2), 391–413.
URL: *http://www.sciencedirect.com/science/article/pii/S0304407699000445* (cited on page 8)

Lanne, M., Lütkepohl, H. and Maciejowska, K.: 2010, Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**(2), 121–131. (cited on page 1)

Lettau, M., Ludvigson, S. C. and Wachter, J. A.: 2008, The Declining Equity Premium: What Role Does Macroeconomic Risk Play?, *The Review of Financial Studies* **21**(4), 1653–1687. (cited on page 1, 19)

Liu, H. and Miao, J.: 2015, Growth uncertainty, generalized disappointment aversion and production-based asset pricing, *Journal of Monetary Economics* **69**, 70–89. (cited on page 1)

Lu, Z.-Q. and Berliner, L. M.: 1999, Markov switching time series models with application to a daily series time, *Water Resources Research* **35**(2), 523–534. (cited on page 8)

Lütkepohl, H. and Velinov, A.: 2016, Structural vector autoregressions: Checking identifying long-run restrictions via heteroskedasticity, *Journal of Economic Surveys* **30**(2), 377–392. (cited on page 1, 19)

Martin, V. L., Hurn, A. S. and Harris, D.: 2012, Econometric Modelling with Time Series, p. 953. (cited on page 13)

McConnell, M. M. and Perez Quiros, G.: 2000, Output Fluctuations in the United States: What Has Changed Since the Early 1980s?, *American Economic Review* **90**(5), 1464–1476. (cited on page 1)

Netsunajev, A.: 2013, Reaction to technology shocks in Markov-switching structural VARs: Identification via heteroskedasticity, *Journal of Macroeconomics* **36**, 51–62. (cited on page 1, 19)

Neyman, J. and Scott, E. L.: 1948, Consistent Estimates Based on Partially Consistent Observations, *Econometrica* **16**(1), 1–32.
URL: *http://www.jstor.org/stable/1914288* (cited on page 2, 8, 10)

Petričková, A.: 2014, Moments of Markov-Switching Models, *Tatra Mountains Mathematical Publications* **61**(1), 131–140.
URL: *http://www.degruyter.com/view/j/tmmp.2014.61.issue-1/tmmp-2014-0032/tmmp-2014-0032.xml* (cited on page 6, 24)

Psaradakis, Z. and Sola, M.: 1998, Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching, *Journal of Econometrics* **86**(2), 369–386.
URL: *http://www.sciencedirect.com/science/article/pii/S0304407698000104* (cited on page 1, 6, 8)

Psaradakis, Z. and Spagnolo, N.: 2006, Joint determination of the state dimension and autoregressive order for models with Markov regime switching, *Journal of Time Series Analysis* **27**(5), 753–766.
URL: *http://doi.wiley.com/10.1111/j.1467-9892.2006.00487.x* (cited on page 8)

Sims, C. A. and Zha, T.: 2006, Were There Regime Switches in U . S . Monetary Policy ?, *American Economic Association* **96**(1), 54–81. (cited on page 1)

Spanos, A.: 2013, Revisiting the Neyman-Scott model: an Inconsistent MLE or an Ill-defined Model?
URL: *http://arxiv.org/abs/1301.6278* (cited on page 6, 10)

Timmermann, A.: 2000, Moments of Markov switching models, *Journal of Econometrics* **96**(1), 75–111. (cited on page 6, 24, 26)

A Appendix

A.1 Moments of a Two-State Markov Switching Model

Timmermann (2000) and Petričková (2014) have derived moments of a general class of Markov regime–switching models. Here, I derive similar moments but in the case where there are two states and the variance doesn't change with the regime (in contrast to the standard assumption in many applications of regime switching for financial market data). For the model in (2) with $\ell = 0$,

$$\mu = \mathbb{E}(y_t) = \pi_0\mu_0 + \pi_1\mu_1 \tag{A.1}$$

where the π_i is the unconditional probability of being in state $S_t := \{0,1\}$ and satisfies the recursion $\pi P = \pi$ for the transition matrix P . The second central moment is obtained by computing:

$$\mathbb{E}(y_t^2) = \pi_0 \mathbb{E}(y_t^2 | S_t = 0) + \pi_1 \mathbb{E}(y_t^2 | S_t = 1) \tag{A.2}$$

$$= \pi_0 \pi_1 \sigma_e^2 + \pi_0 \mu_0^2 + \pi_1 \mu_1^2 \tag{A.3}$$

$$\mathbb{V}(y_t) = \mathbb{E}(y_t^2) - \mu^2 \tag{A.4}$$

$$= \sigma_e^2 + \pi_0 \pi_1 (\mu_1 - \mu_0)^2 \tag{A.5}$$

So that the unconditional variance is increasing in the persistence of the time series (π_i) and the difference between growth rates. For the general case where $\ell \neq 0$ in (2), I use a state-space representation to recast the model as vector auto-regression of order one (VAR(1)):

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-\ell} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_\ell & 0 \\ & & & & 0 \\ & & \mathbf{I}_{\ell\times\ell} & & \vdots \\ & & & & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-\ell-1} \end{bmatrix} + \begin{bmatrix} 1 & -\phi_1 & \dots & -\phi_\ell \\ 0 & \dots & & 0 \\ \vdots & \ddots & & \\ 0 & \dots & & 0 \end{bmatrix} \begin{bmatrix} \mu_{S_t} \\ \mu_{S_{t-1}} \\ \vdots \\ \mu_{S_{t-\ell}} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} e_t \tag{A.6}$$

or in matrix form,

$$\mathbf{Y_t} = \Phi_1 \mathbf{Y_{t-1}} + \Phi_2 \mathbf{M_t} + \mathbf{C} e_t \tag{A.7}$$

The moments are then defined as follows:

$$\mathbf{M} = \mathbb{E}(\mathbf{M_t}) = \mathbf{1}_{\ell+1} \mu \tag{A.8}$$

where $\mathbf{1}_{\ell+1}$ is a vector of ones and μ is defined by (A.1). Now define the variance-covariance matrix of the time varying mean vector $\mathbf{M_t}$ by:

$$\Sigma_{\mathbf{M}} = \mathbb{E}(\mathbf{M_t} - \mathbf{M})(\mathbf{M_t} - \mathbf{M})' \tag{A.9}$$

$$= \begin{bmatrix} \mathbb{E}(\mu_{S_t} - \mu)^2 & \mathbb{E}(\mu_{S_t} - \mu)(\mu_{S_{t-1}} - \mu) & \dots & \mathbb{E}(\mu_{S_t} - \mu)(\mu_{S_{t-\ell}} - \mu) \\ \mathbb{E}(\mu_{S_t} - \mu)(\mu_{S_{t-1}} - \mu) & \mathbb{E}(\mu_{S_{t-1}} - \mu)^2 & \dots & \\ \vdots & & \ddots & \\ & \dots & & \mathbb{E}(\mu_{S_{t-\ell}} - \mu)^2 \end{bmatrix} \tag{A.10}$$

where the main diagonal entries are give by $\mathbb{E}(\mu_{S_t} - \mu)^2 = \vec{\pi}' \left((\vec{\mu}_s - \vec{\mu}) \odot (\vec{\mu}_s - \vec{\mu}) \right)$ with the vectors $\vec{\pi} = (\pi_0, \pi_1)'$, $\vec{\mu}_s = (\mu_0, \mu_1)'$, $\vec{\mu} = (\mu, \mu)'$ and $\odot$ is the element by element multiplication operator.

The off diagonal entries are given by $\mathbb{E}(\mu_{S_t}-\mu)(\mu_{S_{t-n}}-\mu)=\vec{\pi}'\left(\left(P^n(\vec{\mu}_s-\vec{\mu})\right)\odot(\vec{\mu}_s-\vec{\mu})\right)$ where the 2×2 matrix P is the matrix of transition probabilities.[19]

Taking expectations over (A.7), we have:

$$\mathbb{E}(\mathbf{Y_t})=\Phi_\mathbf{1}\mathbb{E}(\mathbf{Y_{t-1}})+\Phi_\mathbf{2}\mathbb{E}(\mathbf{M_t}) \tag{A.11}$$

where we have used the zero mean property of e_t. The covariance-stationary property of y_t then implies that

$$\mathbb{E}(\mathbf{Y_t})=(\mathbf{I}_{\ell+1\times\ell+1}-\Phi_\mathbf{1})^{-1}\Phi_\mathbf{2}\mathbf{1}_{\ell+1}\mu=\mathbf{1}_{\ell+1}\mu$$

To determine the unconditional variance, first subtract (A.11) from (A.7) to obtain

$$\mathbf{Y_t}-\mathbb{E}\mathbf{Y_t}=\Phi_\mathbf{1}(\mathbf{Y_{t-1}}-\mathbb{E}\mathbf{Y_{t-1}})+\Phi_\mathbf{2}(\mathbf{M_t}-\mathbf{M})+\mathbf{C}e_t$$

The orthogonality of e_t to $(\mathbf{Y_{t-1}}-\mathbb{E}\mathbf{Y_{t-1}})$ and $(\mathbf{M_t}-\mathbf{M})$ then implies:

$$\mathbb{E}(\mathbf{Y_t}-\mathbb{E}\mathbf{Y_t})(\mathbf{Y_t}-\mathbb{E}\mathbf{Y_t})'=\Phi_\mathbf{1}\mathbb{E}(\mathbf{Y_{t-1}}-\mathbb{E}\mathbf{Y_{t-1}})(\mathbf{Y_{t-1}}-\mathbb{E}\mathbf{Y_{t-1}})'\Phi_\mathbf{1}'+\Phi_\mathbf{2}\mathbb{E}(\mathbf{M_t}-\mathbf{M})(\mathbf{M_t}-\mathbf{M})'\Phi_\mathbf{2}'+\mathbf{CC}'\sigma_e^2 \tag{A.12}$$

which can be written compactly as:

$$\Sigma_\mathbf{t}=\Phi_\mathbf{1}\Sigma_\mathbf{t-1}\Phi_\mathbf{1}'+\Phi_\mathbf{2}\Sigma_\mathbf{M}\Phi_\mathbf{2}'+\mathbf{CC}'\sigma_e^2 \tag{A.13}$$

where $\Sigma_\mathbf{M}$ is defined by (A.9). Again, stationary y_t implies $\Sigma_\mathbf{t}=\Sigma_\mathbf{t-1}=\Sigma$ so the unconditional variance is defined by:

$$\operatorname{vec}(\Sigma)=\left(\mathbf{I}_{(\ell+1)^2}-\Phi_\mathbf{1}\otimes\Phi_\mathbf{1}\right)^{-1}\operatorname{vec}\left(\Phi_\mathbf{2}\Sigma_\mathbf{M}\Phi_\mathbf{2}'+\mathbf{CC}'\sigma_e^2\right) \tag{A.14}$$

$$=\left(\mathbf{I}_{(\ell+1)^2}-\Phi_\mathbf{1}\otimes\Phi_\mathbf{1}\right)^{-1}\left[(\Phi_\mathbf{2}\otimes\Phi_\mathbf{2})\operatorname{vec}(\Sigma_\mathbf{M})+\operatorname{vec}(\mathbf{CC}')\sigma_e^2\right] \tag{A.15}$$

Applying this result to the AR(1) case we would have: $\Phi_\mathbf{1}=\left[\phi_1,0;1,0\right]$ and $\Phi_\mathbf{2}=\left[1,-\phi_1;0,0\right]$; $\mathbf{C}=(1,0)'$. After some tedious algebra, it can be shown that the variance-covariance matrix Σ_M has entries in the main diagonal $\mathbb{E}(\mu_{S_t}-\mu)^2=\pi_0\pi_1(\mu_0-\mu_1)^2$ and off-diagonal elements $\mathbb{E}(\mu_{S_t}-$

[19] This assumes that the process generated by P is time reversible, which holds in general for 2×2 chains. See Footnote 5 of Timmermann (2000) for discussion of this issue and a definition of the "backward" probabilities required in higher order chains

$\mu)(\mu_{S_{t-1}} - \mu) = \pi_0\pi_1(\mu_0 - \mu_1)^2(p_{0|0} + p_{1|1} - 1)$. This gives the variance expression

$$\Sigma_{\mathbf{11}} = \mathrm{Var}(y_t) = \frac{\pi_0\pi_1(\mu_0 - \mu_1)^2}{1 - \phi_1^2}\left[1 - \phi_1^2 - 2\phi_1(p_{0|0} + p_{1|1} - 1)\right] + \frac{\sigma_e^2}{1 - \phi_1^2} \tag{A.16}$$

and covariance

$$\Sigma_{\mathbf{12}} = \mathrm{Cov}(y_t, y_{t-1}) = \frac{\pi_0\pi_1(\mu_0 - \mu_1)^2}{1 - \phi_1^2}\left[\phi_1(1 + \phi_1^2) - 2\phi_1^2(p_{0|0} + p_{1|1} - 1)\right] + \frac{\phi_1\sigma_e^2}{1 - \phi_1^2} \tag{A.17}$$

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl