IDENTIFICATION AND NORMALIZATION IN MARKOV SWITCHING MODELS OF "BUSINESS CYCLES"

Penelope A. Smith and Peter M. Summers

October 2004

RWP 04-09

Research Division
Federal Reserve Bank of Kansas City

Penelope Smith is a Research Fellow at the University of Melbourne. Peter Summers is an Associate Professor at the University of Melbourne and a Visiting Scholar at the Federal Reserve Bank of Kansas City. The authors wish to thank Todd Clark, James Hamilton, Don Harding, and Adrian Pagan for helpful comments and Chang-Jin Kim for providing them with his data. Computations in this paper were carried out, in part, using the BACC software developed by John Geweke and Siddhartha Chib (http://www.econ.umn.edu/~bacc/), and James LeSage's Applied Econometrics Toolbox for Matlab (www.spatial-econometrics.com). The authors retain all responsibility for any remaining errors. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Smith email: pasmit@unimelb.edu.au
Summers email: peter.summers@kc.frb.org

Abstract

Recent work by Hamilton, Waggoner and Zha (2004) has demonstrated the importance of identification and normalization in econometric models. In this paper, we use the popular class of two-state Markov switching models to illustrate the consequences of alternative identification schemes for empirical analysis of business cycles. A defining feature of (classical) recessions is that economic activity declines on average. Somewhat surprisingly however, this property has been ignored in most published work that uses Markov switching models to study business cycles. We demonstrate that this matters: inferences from Markov switching models can be dramatically affected by whether or not average growth in the 'low state' is required to be negative, rather than simply below trend. Although such a restriction may not be appropriate in all applications, the difference is crucial if one wants to draw conclusions about 'recessions' based on the estimated model parameters.

JEL classification: E32, E37, C22

Keywords: Business cycles, Recessions, Markov switching, Bayesian inference, Posterior distribution

1 Introduction

What is a business cycle? More specifically, what is a recession? The National Bureau of Economic Research (NBER) defines a recession as "...a significant *decline in economic activity* spread across the economy...."[1] Alternatively, Lucas (1979) describes a business cycle as "...real output [undergoing] serially correlated *movements about trend...*," so that implicitly, a recession is a period of below-trend output growth.

This paper demonstrates that this conceptual difference, while seemingly minor, can have major implications for empirical work. We present examples, using the popular Markov switching class of models introduced by Hamilton (1989), that show how inference about the model's fit can be affected by a fundamental ambiguity regarding the model's parameters. This ambiguity, in turn, is directly related to the difference in the nature of 'recessions' outlined above.

We think there are good reasons to prefer the NBER definition of recessions, not least because in most of the examples we present, use of this definition produces what we would consider to be more 'reasonable' results. Having said this, our goal in this paper is not to argue the superiority of one definition over the other. In fact, use of the second definition can also produce 'reasonable' results. However, one may need to look closely to find them, and should at least be aware of the phenomena that we illustrate.

The key to this paper is what (Hamilton, Waggoner & Zha (2004)) term the "identification principle." Given a model with likelihood function $f(y|\theta)$ and parameter vector $\theta \in \mathbb{R}^k$, suppose that the two values θ_1 and θ_2 are observationally equivalent, so that $f(y|\theta_1) = f(y|\theta_2)$ and hence the model is globally unidentified. Hamilton et al note that if the likelihood surface is continuous in θ, then there are loci along which the model is locally unidentified as well.[2] Along these loci, "...the interpretation of parameters is fundamentally ambiguous and ... the interpretation of parameters necessarily changes," (Hamilton et al. (2004, p. 6)). According to Hamilton et al, the identification principle is to use these loci to delineate a subspace A of $\mathbb{R}^k$, with the property that the model is locally identified at all points $\theta \in A \subset \mathbb{R}^k$.

Waggoner & Zha (2003) provide an illustration of this idea in the context of a simultaneous equations system of money demand and supply. They show that adoption of standard normalization procedures

[1] See www.nber.org/cycles/recessions.html; emphases in both this and the next quote are ours.

[2] That is, at any point $\hat{\theta}$ along such a locus, any open neighborhood of $\hat{\theta}$ also contains points that are observationally equivalent to $\hat{\theta}$.

developed for analysis of recursive models can lead to ambiguous inference regarding jointly determined dynamic responses in simultaneous systems. In particular, an exogenous shift in the money supply curve can lead to bimodal distributions of the response of the equilibrium quantity of money and the interest rate. The ambiguity arises because the imposed normalization admits regions of the likelihood surface that represent economically distinct behavioral responses. In their specific example, the distinct responses are generated by a money supply curve that may slope either upwards or downwards. The consequences of normalization for impulse responses in structural VARs is discussed by Sims & Zha (1999).

In this paper, we illustrate how the identification principle may be violated when modelling business cycles using the Markov switching class of models made popular by Hamilton (1989). These models are globally unidentified, since a re-labelling of the unobserved states and state-dependent parameters results in an unchanged likelihood function. Identification is typically achieved by appealing to an order restriction: growth in 'expansions' is higher than in 'recessions.' Using several examples, we show that this normalization scheme can confound two distinct interpretations of phrases such as 'low-growth regime' or 'recession.' Specifically, we show that in some cases the 'low-growth regime' that one estimates could equally well be termed 'normal growth,' and that some 'recession' periods one might identify are characterized by positive growth in GDP. Both of these interpretations are contrary to what most researchers have in mind when they seek to model 'business cycles.' We show how this ambiguity arises, and how it can be overcome with what we term a 'business cycle identification;' on average, output declines during recessions.

1.1 Markov switching models of business cycles

Consider the following two-state Markov switching model:

$$\phi(L)(\Delta y_t - \mu(S_t)) = e_t, e_t \sim iidN(0, \sigma^2)$$

where y_t is the logarithm of real GDP and $\mu(S_t)$ is the mean of its growth rate Δy_t conditional on the unobserved state vector S_t. Specifically, $\mu(S_t)$ switches between high and low growth rates as S_t switches between 0 and 1, with these transitions governed by a first-order Markov process:

$$\mu(S_t) = \mu_0 + \mu_1 S_t. \tag{1}$$

That this model is unidentified is easily seen by considering any given binary sequence S_t, say $\hat{S}_t$, and associated estimates $\hat{\mu}_0$ and $\hat{\mu}_1$. Define $\tilde{S}_t = 1 - \hat{S}_t$, $\tilde{\mu}_0 = \hat{\mu}_0 + \hat{\mu}_1$, and $\tilde{\mu}_1 = -\hat{\mu}_1$. Then the likelihood function will take identical values for $\hat{\theta} = (\hat{S}_t, \hat{\mu}_0, \hat{\mu}_1)$ and $\tilde{\theta} = (\tilde{S}_t, \tilde{\mu}_0, \tilde{\mu}_1)$. We therefore need to restrict the parameters of this model in order to identify the event $S_t = 1$ with the low growth state. One way to do this is to require

$$\mu_1 < 0. \tag{2}$$

In using this class of models for analyzing business cycles, one typically wishes to attach economically meaningful labels to the numerical values of S_t, usually 'expansion' ($S_t = 0$) or 'recession' ($S_t = 1$). We argue that in such cases, a more appropriate identification scheme is

$$\mu_0 + \mu_1 < 0. \tag{3}$$

To further appreciate the difference between these restrictions, and to fix notation for the rest of the paper, define $p_H = pr(S_t = 0|S_{t-1} = 0)$ and $p_L = pr(S_t = 1|S_{t-1} = 1)$ to be the probabilities of remaining in the high-growth and low-growth state, respectively. If we correspondingly denote the mean growth rates in the two states as μ_H and μ_L, then equation (1) implies $\mu_H = \mu_0$ and $\mu_L = \mu_0 + \mu_1$. Equation (3) thus exploits what should be uncontroversial prior information about business cycles: output declines (on average) during recessions. Notice that equation (2), on the other hand, places no restrictions on the sign of μ_L. Also note that both models are *identified* in the statistical sense, but are *normalized* in different ways. We illustrate below several ways in which the different normalizations can have striking implications for the estimation and interpretation of these models.

Of course one could also work with the symmetric restriction of (2): $\mu_0 > 0$, and so restrict μ_H rather than μ_L. The important point to note is that all that is being imposed by (2) is the ordering $\mu_L < \mu_H$. Examples of previous work using this identification include Hamilton (1989), Phillips (1991), Albert & Chib (1993), Filardo & Gordon (1998), Kim & Nelson (1999*a*), and Kaufmann (2000). Our argument is that this is a necessary but not sufficient condition for identification of business cycles. This point has been mentioned by Filardo (1994), who tests whether $\mu_L < 0$ and $\mu_H > 0$. The restriction (3) is also used by Clements & Krolzig (1998).

The distinction between restrictions (2) and (3) can be related to the difference between "classical cycles" and "growth cycles" in the business cycle literature.[3] As pointed out by Harding & Pagan (2002), for example, fluctuations in the (log) *level* of economic activity Y_t are properly analyzed using the (logarithmic) growth rate $\Delta y_t = \ln(Y_t/Y_{t-1})$. In this context, restriction (3) clearly identifies (classical) recession periods; the average growth rate of economic activity is negative. Restriction (2) can be thought of as identifying growth cycle slowdowns, in which the growth rate of activity is below trend, but may still be positive. As we demonstrate in section 3.1 below, however, this restriction is not in general sufficient to identify such periods. Specifically, equation (2) does not rule out cases in which $S_t = 0$ corresponds to periods of extremely *fast* growth, while $S_t = 1$ represents periods of "normal," as well as negative, growth.

It turns out that restriction (3) can be used to identify growth cycles as well as classical cycles, if we replace Δy_t with the *demeaned* growth rate of economic activity $\Delta\tilde{y}_t = \Delta y_t - \mu_\tau$, where μ_τ is the average of Δy_t over the full sample. In this case, either (2) or (3)will identify periods of growth cycle slowdowns. However, now the identification of *classical* recessions requires a stronger restriction, namely $\mu_0 + \mu_1 + \mu_\tau < 0$. In section 3.2, we re-examine the results of Kim & Nelson (1999*a*) and show that their analysis confounds classical recessions with growth cycle slowdowns. Of course, one may well be comfortable treating classical recessions and growth cycle slowdowns as the same sort of event, and so use the label $S_t = 1$ to apply to either one. While there may be no problem with this from an economic perspective, we show that this sort of normalization results in a bimodal posterior distribution for μ_L, rendering inference based on posterior means and standard deviations problematic.[4]

One final point of clarification is necessary. We will use the term 'recession' in this paper exclusively in the classical business cycle sense: to refer to a period when the average growth rate of aggregate economic activity (GDP in our examples) is negative. We will also use the terms 'business cycle' and 'classical cycle' synonymously.

[3] We are grateful to Don Harding for suggesting this interpretation.

[4] As Harding & Pagan (2002) point out, it is usually necessary to specify a dating rule that maps the latent Markov states S_t into observed business cycle phases such as the NBER chronology. For the most part we will abstract from this issue, except to note that the examples we present below will also have adverse consequences for any dating rule based on $pr(S_t = 1|y_t)$.

2 Identification and maximum likelihood estimation

Much of our discussion will be based on Bayesian estimates of the basic model's parameters obtained using Gibbs sampling, as described in Kim & Nelson (1999*b*) and Albert & Chib (1993). However, our results apply to estimation by maximum likelihood as well. This section briefly describes the identification problem in this context.

There are several well-documented difficulties associated with the classical estimation of the parameters of Markov switching models, and mixture models more generally. As described in Hamilton (1991), a global maximum of the likelihood function does not exist. Additionally, there are often several bounded local maxima which produce quite similar values for the likelihood function, but rather different parameter estimates. Such characteristics have led several authors such as Goodwin (1993) and Boldin (1996) to stress the importance of experimenting with a variety of starting values in order to find the local optimum associated with the largest value for the likelihood function.[5] It is rare, however, to see explicit discussion in published work of the sensitivity of parameter estimates to starting values.

The preceding point may be illustrated by estimating the model in equations (1)and (1) for the United Kingdom using quarterly GDP data from March 1960 to December 2001. We experimented with several different sets of starting values and found two local maxima, the estimates of which are presented in tables 1 and 2 below.

Notice that these two sets of parameter estimates yield dramatically different interpretations of the UK 'business cycle.' Recalling that p_H represents the probability of remaining in the high growth state and p_L the probability of remaining in the low growth state, the first set of estimates implies that the UK economy has experienced persistent periods of moderate growth (averaging 0.48 per cent per quarter) with occasional bursts of very high growth (nearly 4 per cent per quarter). The second parametrization implies persistent periods of medium growth ($\mu_H = 0.7$) with occasional recessions ($\mu_L = -0.9$). The parameter estimates in table 1 achieve a higher value of the likelihood function, but do not seem to describe what we usually think of as "business cycles." A researcher confronted with these estimates might be tempted to conclude (wrongly, we would argue) that the Markov switching model does not provide a good description of the UK business cycle.

Which set of estimates should we choose? If our main goal is to model the UK business cycle, then

[5] Goodwin (1993) formalizes a procedure for finding optimal starting values.

as we have argued above, choosing the set with the highest likelihood value is exactly the wrong thing to do. An alternative strategy in this case is to rely on some sort of subjective assumptions regarding what the business cycle should look like. For example, Goodwin (1993), which we view as representative of a large portion of the literature, sees the strength of this particular model of the business cycle as being that it requires "no prior information regarding the dates of the two growth periods or the size of the two growth rates.... In particular, note that the low growth rate need not be negative." [6] He does however regard it as important that both p_H and p_L be significantly different from zero. If this is not the case then "the model is useless for dating turning points in the business cycle because only one state persists through most of the sample period." [7] As he finds that it is sometimes the case that the set of parameter estimates that maximizes the likelihood also produces implausible estimates of p_H and p_L, Goodwin adopts the 'quasi-Bayesian' prior of Hamilton (1991). For seven out of the eight economies that he analyzes, this procedure results in negative estimates of mean growth in the low growth state. Italy is the exception, displaying a positive low-state mean growth rate μ_L, an estimate of p_H very near zero and an estimate of p_L near one.

Hamilton (1991) formalizes the subjective approach by outlining the use of a quasi-Bayesian prior as a means of guiding the parameter estimates to the preferred local maximum. This was chosen over a fully Bayesian approach due to the difficulty of implementing exact Bayesian estimation at the time.[8] We believe that computational power and Bayesian technology have now progressed to such levels that exact Bayesian analysis is now a feasible and attractive alternative.[9] In particular, fully Bayesian analysis of Markov switching models has been developed by Albert & Chib (1993), McCulloch & Tsay (1994), and Kim & Nelson (1999*b*). Also see Frühwirth-Schnatter (2001) and Scott (2002) for more recent developments. However, exact Bayesian methods such as Gibbs sampling are also likelihood-based techniques. In Bayesian analysis, multiple peaks in the likelihood may manifest themselves in the form of multi-modality in the posterior distributions of parameters. This can render measures such as posterior means and standard deviations problematic. Goodwin's anomalous result for Italy is possibly an example of the estimator converging to the "wrong mode." However, in a maximum likelihood setting it is unlikely that this would ever be discovered.

[6] Goodwin (1993, p. 332)

[7] Ibid, p.332.

[8] Hamilton (1991) p. 28.

[9] See Koop & Potter (1999) for a good discussion of the pros and cons of the Bayesian approach to the estimation of this class of non-linear models.

3 Why identification matters

In the following sections we present several examples that illustrate the consequences of using the order restriction or the business cycle restriction for identifying Markov switching models. In all cases, we present features of Bayesian posterior distributions of various functions of interest. Unless stated otherwise, our priors for all of these examples are similar to those used in Albert & Chib (1993). Specifically, the transition probabilities p_L and p_H have beta priors with mean 0.8 and standard deviation 0.16. The mean parameters μ_L and μ_H, and all autoregressive coefficients ϕ have Normal priors centered at zero, and with large standard deviations. The priors for the variances are proper but diffuse.

3.1 Which model is being fit?

In this sub-section, we show that use of the order restriction (2) can distort inference because it leads to a superposition of models. In estimating a Markov switching model for UK GDP in Smith and Summers (2001), we obtained the apparently bizarre result that the posterior mean of the low-growth 'staying probability' was greater than that for the high-growth state. In fact, our posterior simulator output implied that about 38% of the time, the low-growth state is more persistent than the high-growth state (i.e., $p_H \leq p_L$). Given what we know about postwar business cycles, this seems strange: we know that expansions are more persistent than recessions. Why then isn't $p_L < p_H$ all the time?

To see what's happening, figure 1 shows the posterior distribution of the ergodic probability that $S_t = 1$, or $(1 - p_H)/(2 - p_L - p_H)$, across 5,000 Gibbs sampling draws. This represents the average fraction of the sample period that one would expect to be allocated to the low-growth state. According to the business cycle chronology constructed by the Economic Cycle Research Institute (ECRI), the UK economy has actually been in recession about 11% of the time since World War II. This value is not far from the dominant mode in figure 1, while the secondary mode implies that the low-growth state describes over 90% of the UK data.

Figures 2 and 3 show the distributions of the mean parameters μ_L and μ_H, respectively. Each of these figures displays 2 pdf's: one generated under the order restriction (2) and the other under the business cycle restriction (3). The distributions corresponding to the order restriction suggest that there are two distinct 2-state models being fit to the data. In one, GDP growth is usually around 1% per quarter, with rare recessions in which output declines by about 0.5% per quarter; the high-growth state is more persistent than the low-growth state. In the other model, growth is usually around 0.5% per quarter, with rare 'booms' when output grows by about 3% per quarter. Now, the low-growth state is more persistent.[10] Clearly, one's interpretation of the statement "the economy is in the low-growth state" would be quite different in these two models. As shown by Hamilton, Waggoner and Zha, averaging the posterior draws without accounting for this difference involves crossing a locus at which the interpretation of the state-dependent parameters changes. Hence this is a violation of the 'identification principle.'

On the other hand, the inferences about both μ_L and μ_H are much sharper under (3) than (2). Note that the posterior for the low-state mean growth rate under (3) is truncated at zero in figure 3. In particular, the bimodality that is evident in both mean parameters under (2) is resolved when using (3). This figure illustrates graphically the change in interpretation of the 'low-growth state' when μ_L changes from being negative to being positive: i.e., when the identification principle is violated. From a *statistical* point of view the interpretation is arguably the same; in both cases μ_L is the mean growth rate in the low-growth state. On the other hand, if one's primary goal is to model *business cycles*, then the 'low-growth state' that is identified under (2) does not look very much like what one might expect *a priori*. In fact, the dominant mode of μ_L under the order restriction (0.45% per quarter) is nearly equal to the mean of UK GDP growth over the whole sample (0.58% per quarter).

An examination of the likelihood surface as a function of μ_L and μ_H can help to illustrate these points. Figure 4 shows the contours of the likelihood surface in $\mu_L - \mu_H$ space, with the two posterior modes/local maxima indicated.[11] The order restriction, equation (2), rules out the region to the northwest of the 45-degree line, while the business cycle restriction (3) rules out the region to the north of the horizontal line at $\mu_L = 0$. Both modes satisfy the order restriction, but the point labelled 'ML1' is ruled out under the business cycle restriction.

Our basic argument is that in moving between points 'ML1' and 'ML2,' one crosses a locus that changes the interpretation of μ_L, and that the region defined by our business cycle restriction provides

[10]In fact, the two modes in these figures correspond to the two sets of maximum likelihood estimates in table 1 and 2. The ease with which marginal posterior distributions can be estimated and graphed from the output of posterior simulators makes the diagnosis and interpretation of multiple likelihood peaks much more straightforward than in maximum likelihood estimation. Indeed, we believe that our results provide an additional argument in favor of Bayesian estimation of these models, over and above those listed by Koop & Potter (1999).

[11]To generate this figure, we used a variation of Chib's (1995) method to integrate out the model parameters other than μ_L and μ_H. Specifically, for each of our draws of these two parameters, we generated 500 draws of the other parameters from an auxiliary Gibbs sampler with μ_L and μ_H fixed at the given values, then averaged the 500 likelihood values. The relatively low number of passes in the auxiliary sampler accounts for the roughness of the contours in the figure.

a more intuitive interpretation. This is despite the fact that this restriction does more than statistically identify the model, and so could constrain the maximum value of the likelihood. The figure also illustrates that if one wants to allow for slow but positive growth in 'recessions,' one confronts the question of (literally) where to draw the line. That is, how far above zero should we allow average growth to be and still call it a 'low-growth state'?

One possible response to this argument could be that a low-growth state should be defined relative to trend, as this would be a way of imposing an upper bound on how large μ_L could be. This restriction would entail drawing a horizontal line at the average growth rate in figure 4 and ruling out any likelihood peaks north of that line. Equivalently, one could work with de-trended data and leave the 'business cycle restriction' line at zero. As intuitively plausible as this sounds however, we show in the next sub-section that it does not provide a wholly satisfactory alternative.

3.2 What is a "recession"?

In this sub-section, we re-examine some of the results in Kim & Nelson (1999*a*), in order to demonstrate the importance of considering the mapping between the estimated Markov-switching model parameters and the business cycle features that one is ultimately interested in. Kim and Nelson's paper is primarily concerned with detecting a possible structural break in the mean and/or volatility of U.S. GDP growth. We abstract from the question of structural breaks,[12] and focus on Kim and Nelson's baseline results as reported in table 1 of their paper.

The analysis in Kim & Nelson (1999*a*) is based on the de-meaned growth rate of U.S. GDP from 1953:II to 1997:I, with the pre- and post-1973 periods treated separately. Using data for the same period as Kim and Nelson, and defining μ_τ to be the average growth rate over the full sample, we obtained the following results:

$$\mu_\tau = 0.89 - 0.17D73, \tag{4}$$

where $D73$ is a dummy variable which is zero before 1973:I and one thereafter. The t-statistics on the constant and dummy variable are 7.97 and -1.13, respectively.

Using a Gibbs sampling algorithm similar to Kim and Nelson's, for the same number of iterations (10,000 after a burn-in phase of 2,000) and their prior #1, we obtained the results shown in the left-hand

[12] But see Smith & Summers (2002).

panel of table 3. For comparison, the right-hand panel reproduces the estimates from their table 1.[13] The two sets of estimates differ by less than one-third of one posterior standard deviation (either theirs or ours).

The posterior distribution of μ_L is graphed in figure 5, using both Kim and Nelson's identifying restriction ($\mu_1 > 0$, $S_t = 1$ as high-growth state) and equation (3). Both distributions are clearly bimodal, with peaks at -0.25 and -1.09 per cent per quarter. When combined with the parameter estimates in equation (4) we obtain modal estimates for μ_L of 0.64 and -0.20 in the pre-1973 subsample, and 0.47 and -0.37 in the post-1973 subsample. These estimates thus contain elements of both classical recessions (corresponding to the negative modes of μ_L) and growth cycle slowdowns (the positive modes).

It may well be that one would want to consider both growth cycle slowdowns and recessions as periods of economic weakness, and so model them as the same type of event. In that case, basing inference on the posterior mean and standard deviation (or maximum likelihood estimate and associated standard error) would clearly be unsatisfactory in the presence of bimodality such as that displayed in figure 5. On the other hand, one could argue that economic agents face fundamentally different circumstances in recessions than they do in times when growth in activity is slow, but positive. In any event, given that the NBER peak and trough dates pertain to recessions and not growth cycle slowdowns, it is not clear how to interpret graphs (like figure 1 in Kim & Nelson (1999*a*)) which compare the smoothed probability of being in the low-growth state with the NBER recession dates. This is because the estimated regime probabilities are based on two different interpretations of what a 'low-growth state' is.

The effects of identifying the low-growth state with recessions can be seen in table 4 and figures 5 and 6. The model's parameters now have noticeably smaller posterior standard errors, and the residual variance σ^2 has also decreased. The average rate of contraction μ_L is much steeper than either of the estimates in table 3, while expansions are more subdued.[14] The modal value of μ_L has decreased slightly to -1.12 compared to the lower value reported above. There is no evidence of multimodality (see figure 6), although the distribution remains strongly skewed. Finally, figure 7 demonstrates that the (full-sample) inference about the value of S_t is also much sharper. In particular, note that the pre-

[13] Our Gibbs sampler differs slightly from Kim and Nelson's in that we draw μ_1 from its truncated Normal posterior distribution (conditional on μ_0), rather than employing rejection sampling.

[14] The latter result is due to the fact that we have now included the 'slowdown' observations in the high-growth state. We expand on this point below.

vious identification (which confounds recessions and growth cycle downturns) produces 'background' probabilities of between 10 and 20 per cent. This behavior is also evident in Kim and Nelson's figure 1.

An alternative to this approach would be to treat recessions, growth cycle slowdowns, and 'normal' growth as separate states in a three-state Markov switching model. Because recessions must be preceded by slowdowns, this model would have a restricted transition probability matrix. Such a model could be used to shed additional light on the results of Boldin (1996).

4 Implications for Forecasting

As shown by Hamilton (1994), the forecast of y_{t+1} given data through time t is

$$
\begin{aligned}
E(y_{t+1}|S_t) &= E(y_{t+1}|S_{t+1}=0)\,pr(S_{t+1}=0|S_t) &(5)\\
&+E(y_{t+1}|S_{t+1}=1)\,pr(S_{t+1}=1|S_t) &(6)\\
&=\mu_H pr(S_{t+1}=0|S_t)+\mu_L pr(S_{t+1}=1|S_t),
\end{aligned}
$$

where the transition probabilities of the Markov chain provide estimates of $pr(S_{t+1}=0|S_t)$ for $S_t=0,1$. Uncertainty about the current state S_t can easily be incorporated by expressing each term in (5) as a mixture of the two possible values:

$$
\begin{aligned}
E(y_{t+1}|S_t) &= \mu_H\left[p_H \times pr(S_t=0|y_t)+(1-p_L)\times pr(S_t=1|y_t)\right] &(7)\\
&+\mu_L\left[(1-p_H)\times pr(S_t=0|y_t)+p_L\times pr(S_t=1|y_t)\right].
\end{aligned}
$$

We have demonstrated that the order restriction alone can cause these predictive densities, $E(y_{t+1}|S_{t+1}=0)=\mu_H$ and $E(y_{t+1}|S_{t+1}=1)=\mu_L$, to be bimodal and thus adversely affect the model's forecasting performance. The problem is worse than this however, because this bimodality is a direct consequence of ambiguity regarding which observations are being assigned to $S_t=1$. Our two estimates of the average of $pr(S_t=1|y_t)$ for the UK over our sample period are 0.425 and 0.120 for the order and business cycle restrictions, respectively. As a result, inference about the transition probabilities p_H and p_L is also affected. Figure 8 shows this. Note that while there is still considerable uncertainty in the estimation of p_L with the business cycle normalization, this is indicative of an actual source of uncertainty: we do not have very many observations on recessions, and hence cannot estimate μ_L precisely. In contrast, the uncertainty reflected in the distribution of p_L under the order restriction incorporates the additional ambiguity arising from the two interpretations of what a 'low-growth state' is.

As $k \to \infty$, the long run forecasts of y converge to

$$\begin{aligned} E\left(y_{t+k}|S_t\right) &= E\left(y|S=0\right) pr\left(S=0\right) \\ &\quad + E\left(y|S=1\right) pr\left(S=1\right) \\ &= \mu_H \bar{p}_H + \mu_L \bar{p}_L \end{aligned} \tag{8}$$

where $\bar{p}_H$ and $\bar{p}_L$ are the ergodic probabilities of the Markov process. The long-run forecasts one would make for the UK, based on both identification schemes, are shown in figure 10. With the order restriction, the upward bias due to the second mode in μ_H shows up clearly; the long run forecast of UK GDP growth is 2 per cent *per quarter*. The figure also shows the sample mean of about 0.6 per cent. The long run forecast obtained using our business cycle restriction is virtually indistinguishable from the sample mean.

Figure 9 gives another example, showing the distribution of $\bar{p}_H$ for Canadian GDP growth from 1960:II to 2000:IV. In this case, the posterior of $\bar{p}_H$ is basically flat between values of 0.25 and 0.85. Figure 11 shows the two long-run forecasts and the sample mean growth rate for Canada. Again the implication is clear, though somewhat less striking than in the UK case.

Notice that this information can also be used as a diagnostic tool. Since the Markov switching process estimated under either identifying restriction is stationary (as long as there are no absorbing states), the long run forecasts should converge to the sample mean. The forecasts obtained using the order restriction are clearly unsatisfactory according to this criterion.

5 The effects of the prior

Finally, we briefly discuss the effect of changing the priors on the model parameters under the two alternative identification schemes. As mentioned above, our priors for all model parameters are proper but reasonably diffuse. For example, the priors on the transition probabilities are beta distributions with mean 0.8 and standard deviation 0.16, as in Albert & Chib (1993). We illustrate the effect of changing the priors to ones that are essentially flat, yet still proper.

Specifically, we re-estimated the model of sub-section 3.1 using uniform priors on the transition probabilities and $N(0,10000)$ priors for the mean and autoregressive parameters.[15] We leave the variance priors unchanged as they are already as diffuse as possible while guaranteeing that their posterior moments exist. Using these priors means that the posterior means will essentially correspond to the maximum likelihood estimates.

We present the results of this exercise in figures 12 and 13, in the context of dating recession periods. Figure 12 shows that in the case of the order restriction, the model with a flat prior becomes useless for describing business cycles. Under the flat prior, the model parameters are given by the estimates in table 1, and all but four data points are assigned to the low-growth state. The informative prior is more successful at identifying recession periods, but the signal provided by $pr(S_t = 1|y_T)$ is very noisy and concentrated near 50%.

On the other hand, the model estimated under the business cycle restriction conveys virtually the same information about recession periods using either prior. The signals are less pronounced with the flat prior, as one would expect, but even with the flat prior the 'background' probability of being in the low-growth state is nearly always below 10%.

6 Conclusions

This paper has shown how the phenomena reported in Hamilton et al. (2004) can manifest themselves in the analysis of business cycles using Markov switching models. Basing our identifying restrictions on commonly used terminology in the business cycle literature, we demonstrate that results from both maximum likelihood and Bayesian estimation methods can change in important ways. At the very least, researchers modelling business cycles as fluctuations about trend should be aware of how this characterization can affect inference.

Although our discussion has been focused on the analysis of business cycles, we believe our results apply to the use of Markov switching models more generally. Specifically, whenever a researcher has prior beliefs about how one or more unobserved states in the model should correspond to the observable phenomena in which she is ultimately interested, that information should be used to identify the model.

[15] Of course, we are also imposing a further prior on μ_1 corresponding to the relevant normalization.

References

Albert, J. H. & Chib, S. (1993), 'Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *Journal of Business and Economic Statistics* **11**(1), 1–15.

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.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–1321.

Clements, M. P. & Krolzig, H.-M. (1998), 'A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP', *Econometrics Journal* **1**, C47–C75.

Filardo, A. J. (1994), 'Business cycle phases and their transitional dynamics', *Journal of Business and Economic Statistics* **12**(3), 299–308.

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

Frühwirth-Schnatter, S. (2001), 'Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models', *Joural of the American Statistical Association* **96**(453), 194–209.

Goodwin, T. H. (1993), 'Business-cycle analysis with a Markov-switching model', *Journal of Business and Economic Statistics* **11**, 331–339.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of non-stationary time series and the business cycle', *Econometrica* **57**, 357–384.

Hamilton, J. D. (1991), 'A quasi-Bayesian aproach to estimating parameters for mixtures of normal distributions', *Journal of Business and Economic Statistics* **9**(1), 27–39.

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

Hamilton, J., Waggoner, D. F. & Zha, T. (2004), 'Normalization in econometrics', *Working Paper 2004-13, Federal Reserve Bank of Atlanta* .

Harding, D. & Pagan, A. R. (2002), 'Dissecting the cycle: A methodological investigation', *Journal of Monetary Economics* **49**(2), 365–381.

Kaufmann, S. (2000), 'Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods', *Econometrics Journal* **3**, 39–65.

Kim, C.-J. & Nelson, C. R. (1999*a*), 'Has the U.S. economy become more stable? a Bayesian approach based on a Markov-switching model of business cycle', *Review of Economics and Statistics* **81**(4), 608–616.

Kim, C.-J. & Nelson, C. R. (1999*b*), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, Cambridge, MA.

Koop, G. & Potter, S. M. (1999), 'Bayes factors and nonlinearity: Evidence from economic time series', *Journal of Econometrics* **88**, 251–281.

McCulloch, R. E. & Tsay, R. S. (1994), 'Statistical analysis of economic time series via Markov switching models', *Journal of Time Series Analysis* **15**(5), 523–538.

Phillips, K. L. (1991), 'A two-country model of stochastic output with changes in regime', *Journal of International Economics* **31**, 121–142.

Robert E. Lucas, J. (1979), 'An equilibrium model of the business cycle', *Journal of Political Economy* **83**(1), 1113–1144.

Scott, S. L. (2002), 'Bayesian methods for hidden Markov models: Recursive computing in the 21st century', *Journal of the American Statistical Association* **97**(457), 337–351.

Sims, C. & Zha, T. (1999), 'Error bands for impulse responses', *Econometrica* **67**(5), 1113–1156.

Smith, P. A. & Summers, P. M. (2002), 'Regime switches in GDP growth and volatility: Some international evidence and implications for modelling business cycles', *Working paper 21/02, Melbourne Institute, the University of Melbourne* .

Waggoner, D. F. & Zha, T. (2003), 'Likelihood-preserving normalization in multiple equation models', *Journal of Econometrics* **114**(2), 329–47.

Table 1: Maximum likelihood estimates, first local maximum

Variable	Coefficient	Standard Error
p_L	0.974	0.012
p_H	0.139	0.184
ϕ_1	0.091	0.047
ϕ_2	-0.014	0.041
ϕ_3	0.259	0.067
ϕ_4	0.102	0.053
σ^2	0.792	0.030
μ_0	3.910	0.161
μ_1	-3.421	0.087
log-likelihood	-212.5824	

Table 2: Maximum likelihood estimates, second local maximum

Variable	Coefficient	Standard Error
p_L	0.700	0.241
p_H	0.974	0.017
ϕ_1	-0.133	0.125
ϕ_2	0.003	0.159
ϕ_3	0.232	0.102
ϕ_4	0.030	0.091
σ^2	0.868	0.064
μ_0	0.725	0.094
μ_1	-1.628	0.778
log-likelihood	-223.6465	

Table 3: Posterior estimates

	This paper		Kim-Nelson[a]	
	mean	std. dev.	mean	std. dev.
μ_L	-0.894	0.633	-0.817	0.620
μ_H	0.230	0.203	0.297	0.343
ϕ	0.258	0.106	0.258	0.105
σ^2	0.768	0.131	0.776	0.136
p_L	0.704	0.151	0.706	0.161
p_H	0.890	0.128	0.840	0.152

[a] See table 1 in Kim & Nelson (1999*a*).

Table 4: Posterior estimates, recession identification

	mean	std. dev.
μ_L	-1.423	0.467
μ_H	0.194	0.114
ϕ	0.228	0.103
σ^2	0.713	0.115
p_L	0.656	0.138
p_H	0.945	0.035

Figure 1: Ergodic probability of high-growth state, UK GDP data, alternative normalizations

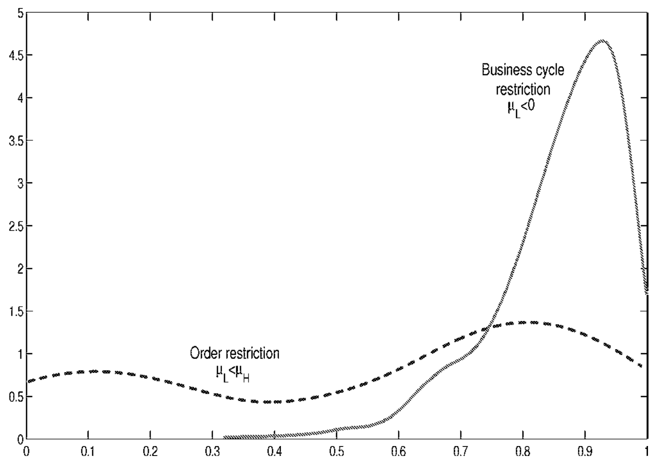

Figure 2: Posterior distributions of low-growth state mean, μ_L, UK data

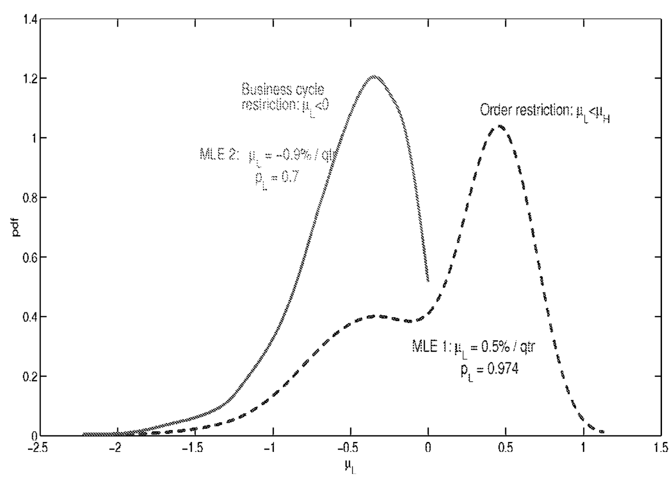

Figure 3: Posterior distributions of high-growth state mean, μ_H, UK data

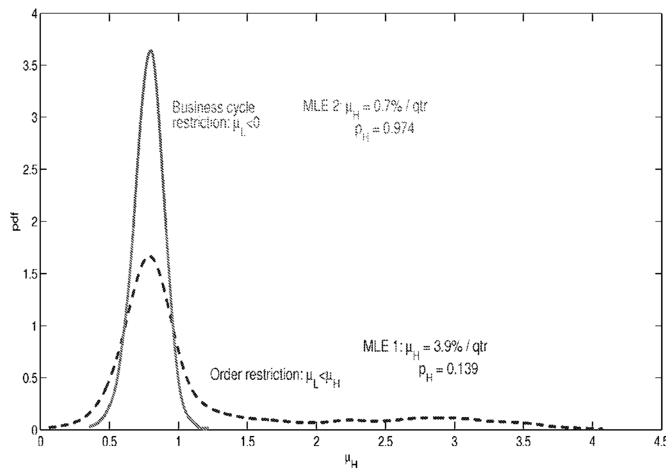

Figure 4: Likelihood surface, UK GDP data

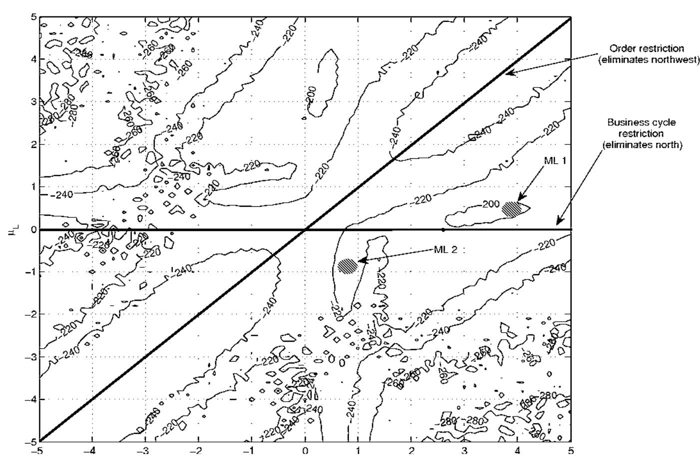

Figure 5: Posterior distribution of μ_L, Kim & Nelson (1999*a*) model

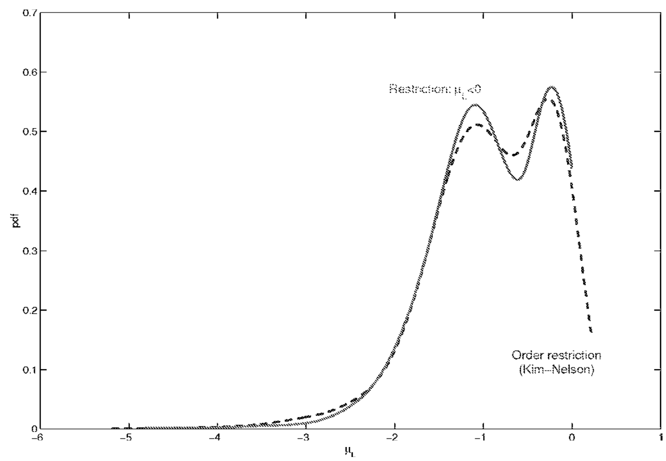

Figure 6: Posterior distribution of μ_L, alternative identifications, Kim & Nelson (1999*a*) model

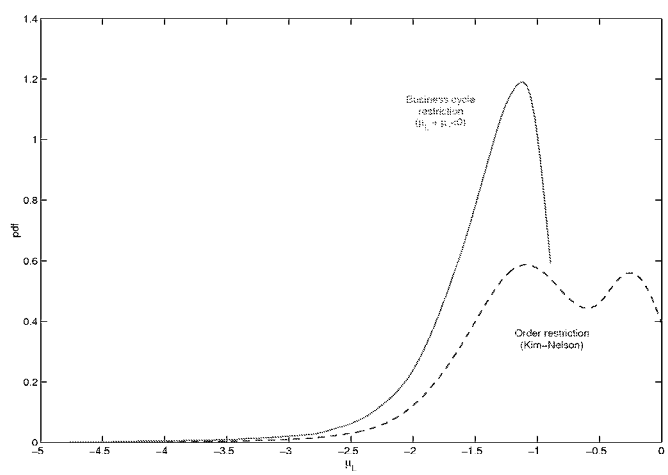

Figure 7: Smoothed probability of low-growth state, alternative identifications, Kim & Nelson (1999*a*) model

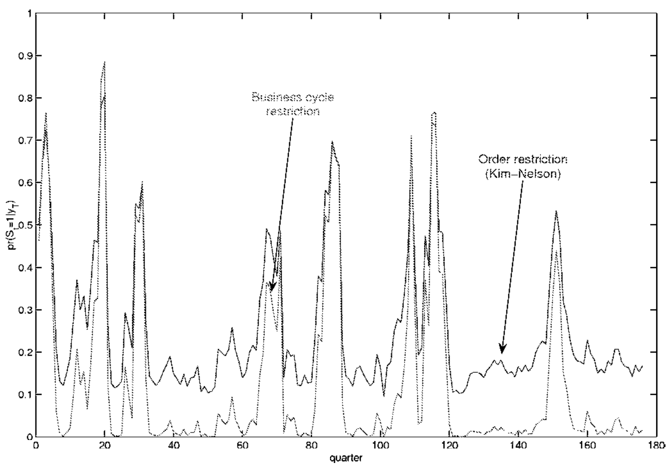

Figure 8: Posterior distribution of p_L, UK GDP growth

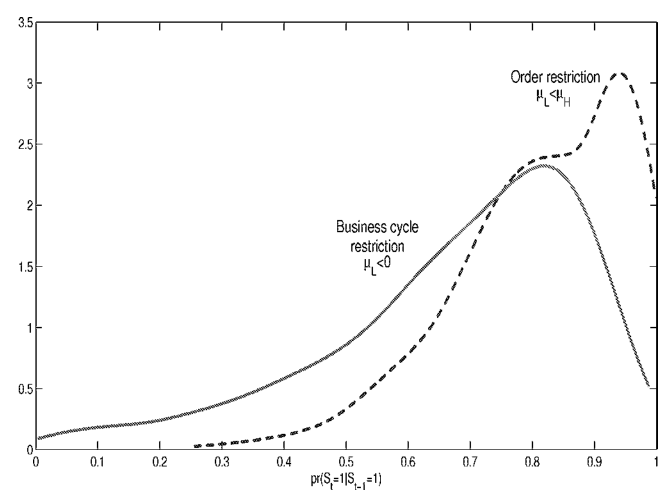

Figure 9: Ergodic probability of high-growth state, Canadian GDP data, alternative normalizations

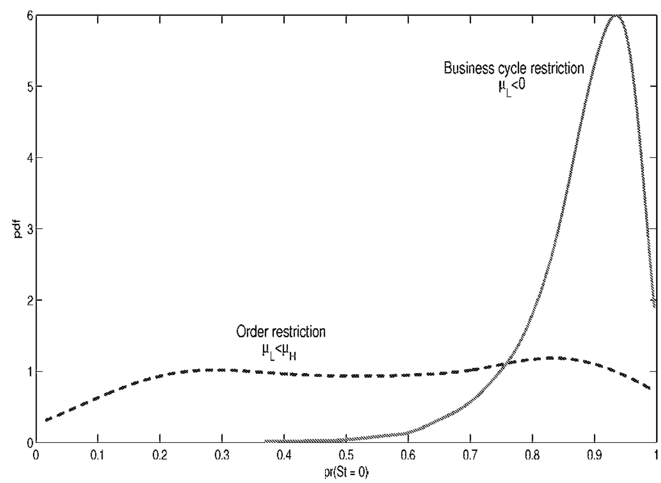

Figure 10: UK GDP growth, sample mean and long run forecasts

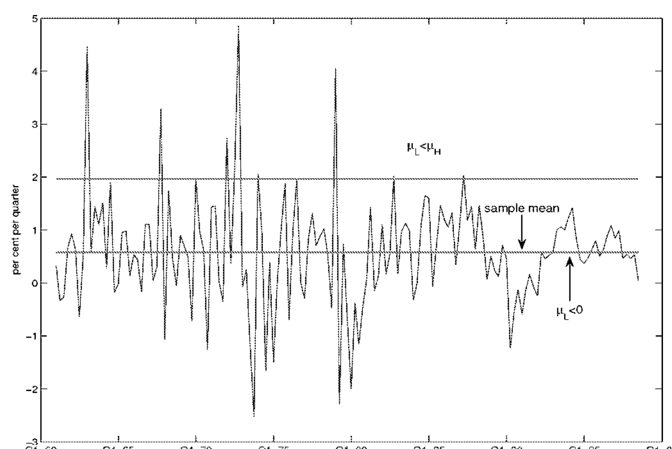

Figure 11: Canadian GDP growth, sample mean and long run forecasts

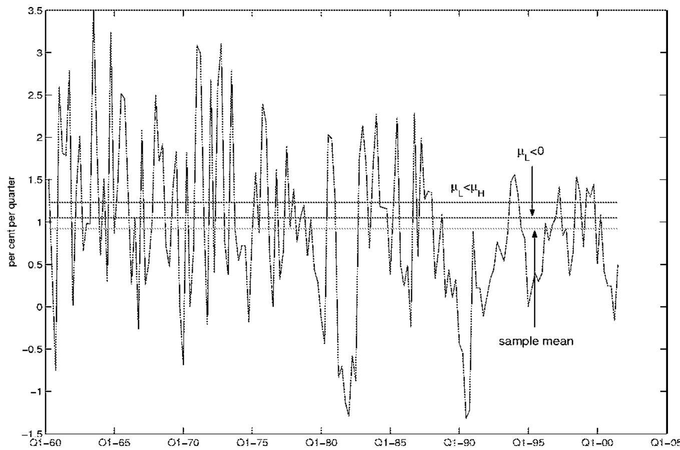

Figure 12: Smoothed probabilities of low-growth state, UK GDP data, alternative priors with order restriction

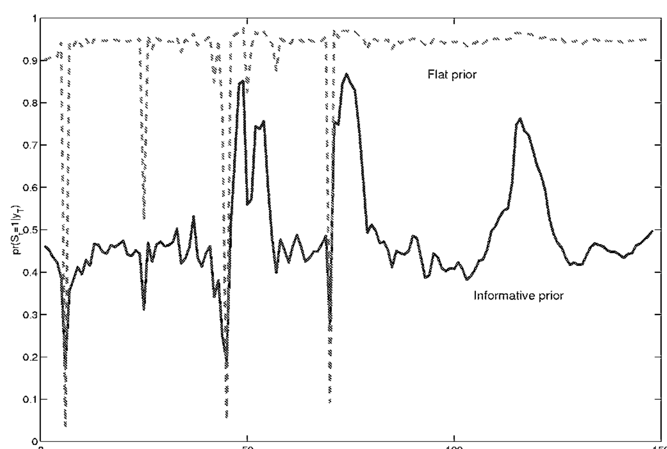

Figure 13: Smoothed probabilities of low-growth state, UK GDP data, alternative priors with business cycle restriction