Conditional Markov chain and its application in economic time series analysis

Bai, Jushan and Wang, Peng

Columbia University, Hong Kong University of Science and Technology

August 2011

CONDITIONAL MARKOV CHAIN AND ITS APPLICATION IN ECONOMIC TIME SERIES ANALYSIS

JUSHAN BAI[a] AND PENG WANG[b]

[a] *Department of Economics, Columbia University, New York, NY, USA. Email: jb3064@columbia.edu*

[b] *Department of Economics, Hong Kong University of Science and Technology. Email: pwang@ust.hk*

Abstract

Motivated by the great moderation in major U.S. macroeconomic time series, we formulate the regime switching problem through a conditional Markov chain. We model the long-run volatility change as a recurrent structure change, while short-run changes in the mean growth rate as regime switches. Both structure and regime are unobserved. The structure is assumed to be Markovian. Conditioning on the structure, the regime is also Markovian, whose transition matrix is structure-dependent. This formulation imposes interpretable restrictions on the Hamilton Markov switching model. Empirical studies show that this restricted model well identifies both short-run regime switches and long-run structure changes in the U.S. macroeconomic data.

Key Words Markov regime switching, Conditional Markov chain

Second submission: August 2008. Final revision: February 2009.

1 INTRODUCTION

There has been a substantial decline in the volatility of major macroeconomic variables since early 1980s.[1] The evidence of this decline is so striking that economists have named it the "Great Moderation". This feature should be captured in the calibration and estimation

[1]See Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Warnock and Warnock (2000), Blanchard and Simon (2001), Kim *et al.* (2004), etc.

of macroeconomic models that are applied to the entire postwar U.S. data. For example, in a stochastic growth model, when modeling the endowment process, the volatility of the exogenous income should be treated as a random process, as opposed to one parameter to be estimated.[2]

Hamilton's (1989) seminal application of a Markov switching model to U.S. GDP growth data successfully captured its cyclical behavior, but at that time, the changing volatility was not a noteworthy feature of the data. The original Hamilton model (constant variance) only weakly identifies some recessions when using data up to late 1990s, as shown in Figure 3 of McConnell and Perez-Quiros (2000). In this paper, we explicitely model the changes in variance. The proposed model has two mean states and two variance states. Restrictions are imposed on the way in which the two types of states interact with each other. The model can be viewed as a restricted version of the Hamilton model with four states. Section 4 shows that this restricted model provides good estimates for the recession probabilities, and the results are robust with respect to the choice of data range.

A number of previous stuides also allow changing volatility. Kim and Nelson (1999) added an unknown change point to the Markov switching model. For the U.S. postwar GDP growth data, they found not only evidence of a structural change toward stabilization around the first quarter of 1984, but also a narrowing of the gap between growth rates during recessions and booms. Lettau *et al.* (2008) applied an independent Markov switching model, developed by McConnell and Perez-Quiros (2000), to consumption data, and found evidence of a shift to substantially lower volatility regimes at the beginning of 1990s. One of the objectives of this paper is to establish a simple model that captures various key features of the data, such as the narrowing mean growth rate gap, changing volatilities, and time-varying transition probabilities (or equivalently, changing recession durations).

This paper categorizes the state of the economy into two groups, namely, the exogenous

[2]An alternative method is to assume there was a structural break in the variance for the endowment process.

state and the endogenous state. The exogenous state, or structure, is designed to characterize long-run structure changes, while the endogenous state, or regime, is used to describe short-run business cycles. The exogenous state evolves according to a homogeneous Markov chain. Given the exogenous state, the endogenous state also follows a homogeneous Markov chain, whose transition probabilities depend on the exogenous state. The endogenous state thus follows a "conditional Markov chain", where the Markovian property applies only after conditioning on the exogenous state.

This model imposes interpretable restrictions on the conventional Hamilton Markov switching model with properly defined state variables (see Section 3 for details). To convey the main idea, we start with the baseline setup. Let structure A_t take a value in $\{1,2\}$ and regime s_t be either 1 or 2. By assumption, A_t follows a first order stationary Markov chain, with a 2×2 transition matrix P^A. Regime s_t follows a conditional first order Markov chain, where under the structure $A = k$, the regime is driven by the transition matrix P_k, $k = 1, 2$. We assume $P_k(i,j) \equiv \Pr(s_t = i | A_t = k, s_{t-1} = j)$, $k = 1, 2$.[3] The model is characterized by three transition matrices and the joint distribution of initial states (A_0, s_0), namely,

$$P^A = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}, \quad P_1 = \begin{pmatrix} p_1 & 1-q_1 \\ 1-p_1 & q_1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} p_2 & 1-q_2 \\ 1-p_2 & q_2 \end{pmatrix}$$

and $\Pr(A_0, s_0)$, where $A_t \in \{1,2\}$, $s_t \in \{1,2\}$. We assume both states are not observed by the econometrician.

Like the conventional Markov switching model, the econometrician only observes the time series $\{y_t\}_{t=1}^T$, where the data generating process of y_t is given by $y_t = \mu(A_t, s_t) + \sigma(A_t) \cdot e_t$, $e_t \sim N(0,1)$. This implies $y_t \sim N(\mu(A_t, s_t), \sigma^2(A_t))$.

Because both A_t and s_t are hidden, we treat them as missing data and apply the well-known expectation-maximization (EM) algorithm to estimate the model. The estimation

[3] "Pr" means probability of an event, or probability density of a random variable if no value is assigned.

process is fast because no numerical optimization is required. Robustness of the estimation results is checked by trying various initial values. When autoregressive terms are added as in Hamilton (1989), an expectation-conditional-maximization (ECM) algorithm[4] is applied to ensure closed-form solutions throughout. By applying this model to U.S. post-war data on GDP and employment, we find that there is a volatility change at around the first quarter of 1984, consistent with most existing literature, and all NBER recession dates are well identified by looking at smoothed or filtered recession probabilities. The estimated structure transition probabilities also suggest that the volatility change is highly persistent.

2 RELATIONS TO OTHER RESTRICTED MARKOV SWITCHING MODELS

The above model admits rich features while keeps a reasonably parsimonious model structure. It includes the unknown change point Markov switching model (Kim and Nelson, 1999) and the independent Markov switching model (McConnell and Perez-Quiros, 2000, and Lettau *et al*, 2008) as special cases.

Example 1 *(Unknown change point Markov switching model) In the conditional Markov chain model, let* $0 < p < 1$ *and* $q = 1$*, then the low variance structure state is an absorbing state. What remains is to estimate the location of the (deterministic) permanent structural change. By further restricting* $P_1 = P_2$*, the resulting model is equivalent to the unknown change point Markov switching model as in Kim and Nelson (1999)*[5]*. The transition matrices are given by* $P^A = \begin{pmatrix} p & 0 \\ 1-p & 1 \end{pmatrix}$, $P_1 = P_2 = \begin{pmatrix} p_1 & 1-q_1 \\ 1-p_1 & q_1 \end{pmatrix}$. *We will see in the next example that it is also a special case of the independent Markov switching model.*

Example 2 *(Independent Markov switching model) If we let* $p_1 = p_2$ *and* $q_1 = q_2$*, then the resulting model is equivalent to the independent Markov switching model, where the dynamics*

[4] Meng and Rubin (1993) developed a general theory of the ECM algorithm.

[5] They considered a general setup with mean gap difference and variance change, i.e., mean growth rate depends on both A_t and s_t, while variance only depends on A_t.

of regime s_t no longer depends on the structure state A_t. The independent Markov switching model essentially requires the conditional transition matrices of regime s_t to be the same across different structures, i.e., $P_1 = P_2$, with no restrictions imposed on the transition matrix of structure A_t.

Recently, Geweke *et al.* (2007) proposed a Hierarchical Markov Normal Mixture model (HMNM) to study financial asset returns. The conditional Markov chain also includes restrictions implied by the HMNM as special cases.

Example 3 *(HMNM) If we restrict the conditional transition matrices for regimes such that the diagonal terms add up to one and each row contains the same elements, then the conditional Markov chain model becomes an HMNM model. In particular, the resulting transition matrices take the following forms:* $P^A = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$, $P_1 = \begin{pmatrix} p_1 & p_1 \\ 1-p_1 & 1-p_1 \end{pmatrix}$, $P_2 = \begin{pmatrix} p_2 & p_2 \\ 1-p_2 & 1-p_2 \end{pmatrix}$. *The HMNM restricts the conditional transition kernel such that* $\Pr(s_t = i | A_t = k, s_{t-1} = j) = \Pr(s_t = i | A_t = k)$, *for all i, k, j. Conditioning on a specific structure $A_t = k$, the observable y_t follows a mixture of two normal distributions.*

It is worth mentioning that Barro *et al.*'s (2009) formulation for country disasters is also a special case of the conditional Markov chain framework.

In the next section, the proposed model is cast under the conventional Hamilton Markov regime switching framework, with four regime states, $\{(\sigma_H^2, \mu_L^H), (\sigma_H^2, \mu_H^H), (\sigma_L^2, \mu_L^L), (\sigma_L^2, \mu_H^L)\}$. However, the latter setup involves 12 probability parameters concerning the transition matrix of these 4 states, and the number of parameters increases with the same magnitude as the squared number of states, making MLE numerically undesirable. Also in practice, the likelihood function generally has multiple local maxima, and it is difficult to achieve a reasonable local maximum. Instead, by explicitly modeling the long-run and short-run regime changes, the proposed model implies a grand transition matrix with only 6 parameters, which greatly reduces the computational burden. Thus the proposed model can be viewed as a parsimonious way to model the 4-state Hamilton Markov regime switching problem. In the sense of Sims *et al.* (2008), the conditional Markov chain can be treated as a way to provide restrictions on the transition probabilities for the Hamilton Markov regime switching model. Another implication of the conditional Markov chain model is that the "regime state" by itself is generally not Markovian, while a common feature shared by the first two examples is that the marginal process of the "regime state" is Markovian.

3 GENERAL MODEL SETUP

We make the following assumptions throughout.

Assumption 1. $\{A_t\}$ is an exogenous Markov process:

$\Pr(A_{t+1}|A_t, s_t, A_{t-1}, s_{t-1}, ..., A_0, s_0) = \Pr(A_{t+1}|A_t)$.

Assumption 2. $\{s_t\}$ is conditionally Markovian:

$\Pr(s_{t+1}|A_{t+1}, A_t, s_t, A_{t-1}, s_{t-1}, ..., A_0, s_0) = \Pr(s_{t+1}|A_{t+1}, s_t)$.

The first assumption states that A_t forms a sufficient statistic for the entire history of (A, s) for predicting A_{t+1}. The second assumption means conditioning on historical structure states, the regime s is Markovian, whose transition matrix depends on the realization of current structure state.

In general, the model admits M exogenous states and N endogenous states, with $A_t \in \{1, ..., M\}$ and $s_t \in \{1, ..., N\}$. We have an $M \times M$ probability matrix P^A to characterize the evolution of A_t. Accordingly, there are altogether M probability matrices, P_m $(m = 1, ..., M)$, characterizing transition probabilities of s_t conditional on A_t, each being of dimension $N \times N$. Based on assumptions 1 and 2, we may prove that the joint state, $Z_t \equiv (A_t, s_t)$, is first order

Markovian,

$$\Pr(A_{t+1}, s_{t+1}|A_t, s_t, A_{t-1}, s_{t-1}, ..., A_0, s_0) = \Pr(A_{t+1}, s_{t+1}|A_t, s_t).$$

A typical realization of joint state is given by $(A_t = m, s_t = n)$, $m \in \{1, ..., M\}$, $n \in \{1, ..., N\}$, and the number of joint states is MN.

The $MN \times MN$ transition matrix P^Z characterizing the Markov process $\{Z_t\}$ can be constructed as follows,

$$\Pr(A_{t+1}, s_{t+1}|A_t, s_t) = \Pr(s_{t+1}|A_{t+1}, A_t, s_t) \cdot \Pr(A_{t+1}|A_t, s_t) = \Pr(s_{t+1}|A_{t+1}, s_t) \cdot \Pr(A_{t+1}|A_t),$$

where $\Pr(s_{t+1}|A_{t+1}, s_t)$ and $\Pr(A_{t+1}|A_t)$ are given by elements of P_m $(m = 1, ..., M)$ and P^A respectively. Z_t can be viewed as a Hamilton Markov switching process with MN states and with a restricted transition matrix.

Example 4 *Consider the conditional Markov chain setup in Section 1. If we order the joint state $Z_t \equiv (A_t, s_t)$ as $[(1,1),(1,2),(2,1),(2,2)]'$, then the corresponding transition matrix takes the form*

$$P^Z = \begin{pmatrix} p \cdot P_1 & (1-q) \cdot P_1 \\ (1-p) \cdot P_2 & q \cdot P_2 \end{pmatrix}.$$

Clearly, the conditional Markov chain provides certain restrictions on the transition matrix of a conventional Hamilton Markov switching model. Notice that when we assume independent switching where $P_1 = P_2$, the transition matrix for Z_t admits a simple Kronecker tensor product representation $P^Z = P^A \otimes P_1$ as shown in Sims et al. (2008). One may also prove that under the conditional Markov chain restriction, the mapping from (P^A, P_1, P_2) to P^Z is a bijection if the structure state A_t is nondegenerate.

4 APPLICATIONS TO ECONOMIC TIME SERIES DATA

Our data sets consist of the GDP growth and the employment growth data, spanning from the second quarter of 1947 to the fourth quarter of 2006.[6] To see how the model works for the simplest setup, we first abstract from autoregressive components for economic variables, and concentrate on the basic setup as in sections 1 and 2. The growth rate is measured as the log-difference of the data multiplied by 100, i.e., $y_t = 100 \cdot \log(GDP_t/GDP_{t-1})$. We estimate a conditional Markov chain model with two regime states $\{s_H, s_L\}$, and two variance states $\{A_H, , A_L\}$, where "H" means high, and"L" means low. The model is $y_t = \mu(A_t, s_t) + \sigma(A_t)e_t$, $e_t \sim N(0, 1)$ being independent with (A_t, s_t). Let $\mu(A_i, s_j) = \mu^i_j$, and $\sigma(A_i) = \sigma_i$, $i, j = H, L$. The structure state is first order Markovian, with transition matrix P^A. Given that the structure is $A_t = \sigma^2_H$, the regimes are driven by P^H at time t. Accordingly, regimes will be driven by P^L under the low volatility structure. With this specification, the business cycles are characterized by switches between high and low mean growth rate of GDP, while the long-run change of the volatility can be viewed as a transition from σ^2_H to σ^2_L.[7]

4.1 APPLICATIONS TO GDP GROWTH

The estimation procedure features a two-step process. We use the EM algorithm to obtain initial estimates for parameters, and then we directly maximize the likelihood function to refine our estimates and to obtain the standard errors. Maximum likelihood estimation yields a log likelihood of -289.3155, and parameter estimates, with standard errors shown in parenthesis, are given by

[6] Data source: U.S. Department of Commerce, Bureau of Economic Analysis. All data are measured in 2000 chain-weighted dollars.

[7] Long-run change means the probability of staying in the same structure is high, say 0.99.

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2
$\widehat{\theta}_1$	-0.0849	1.4149	0.1716	0.8913	0.8780	0.1590
	(0.2181)	(0.1610)	(0.1492)	(0.0545)	(0.1326)	(0.0258)
θ_2	p_1	q_1	p_2	q_2	p	q
$\widehat{\theta}_2$	0.7572	0.8637	0.8332	0.9630	0.9933	1.0000
	(0.0867)	(0.0629)	(0.1181)	(0.0273)	(0.0066)	(0.0000)

The implied transition matrix for the joint states is

$$P^Z = \begin{pmatrix} 0.7522 & 0.1354 & 0.0000 & 0.0000 \\ 0.2412 & 0.8580 & 0.0000 & 0.0000 \\ 0.0055 & 0.0002 & 0.8332 & 0.0370 \\ 0.0011 & 0.0064 & 0.1668 & 0.9630 \end{pmatrix}$$

The order of the four states is "high variance, low mean", "high variance, high mean", "low variance, low mean" and "low variance, high mean". Probabilities for low growth regimes[8] and high variance structures are shown in Figure 1. The shaded areas are NBER dated recessions. We can see that not only the NBER recessions are very precisely estimated in terms of filtered and smoothed recession probabilities, but the low frequency movement of the variance is well captured[9]. The parameter estimates also suggest that besides a substantial volatility drop, a changing mean growth gap is also an important feature. Along with the assumption that the recession duration depends on the volatility structure, this model

[8] In the model, recession is described as the regime with low mean growth rate. Notice that the NBER recession is defined as "a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales." The intriguing feature of Markov switching model is that its estimates for recession probabilities accord with NBER's recession dates very well, by just looking at a single time series.

[9] The smoothed probabilities for high variance structures around the turning points are $\Pr(1984Q1|I_T) = 0.98$, $\Pr(1984Q2|I_T) = 0.81$, $\Pr(1984Q3|I_T) = 0.29$, and $\Pr(t|I_T) > 0.99$ for $t \leq 1983Q4$.

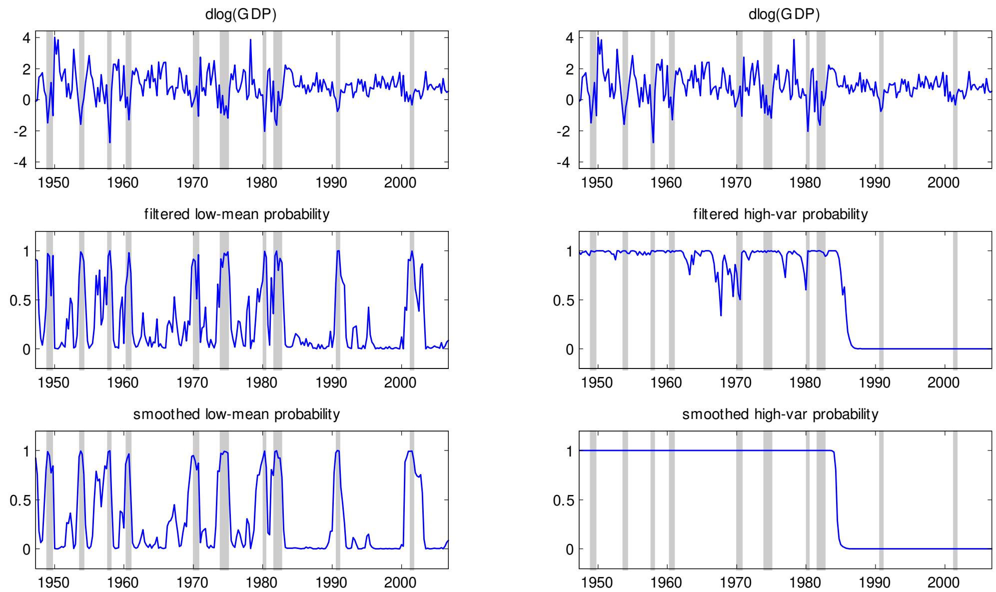

Figure 1: GDP growth and estimated state probabilities, 47Q2 – 06Q4.

is able to provide more precise recession probabilities than the existing literature does. It is natural to ask whether our restrictions on the state transition probabilities are reasonable. We reestimate a Hamilton Markov switching model using the 4-dimensional joint states $Z_t \in \{(A_H, s_L), (A_H, s_H), (A_L, s_L), (A_L, s_H)\}$. No restriction is imposed on the transition matrix for the joint states. Maximum likelihood estimation yields a log likelihood value of -288.8377, with the estimated transition matrix

$$P^Z = \begin{pmatrix} 0.7596 & 0.1418 & 0.0000 & 0.0000 \\ 0.2404 & 0.8472 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.8335 & 0.0377 \\ 0.0000 & 0.0109 & 0.1665 & 0.9623 \end{pmatrix}.$$

We can see that the estimated unrestricted transition matrix is very close to the one implied by the conditional Markov chain restriction. A likelihood ratio test as in Hamilton and Lin

(1998) gives a log likelihood ratio statistic of $-2(-289.3155 + 288.8377) = 0.9556$, which yields a p-value of 0.9872 according to the asymptotic $\chi^2(6)$ distribution. This provides statistical support that the conditional Markov chain is a reasonable restriction on the 4×4 transition matrix of the joint state. A practical advantage of the conditional Markov chain restriction is that it admits fast and robust computation of the MLE, while the estimation of the unrestricted model is numerically challenging. When estimating the unrestricted model, we tried a large variety of initial parameter values to start a Quasi-Newton optimization procedure, and the procedure easily got stuck at a local maximum. The maximum likelihood value is achieved by using the parameter estimates implied by the conditional Markov chain model as starting values. For higher dimensional Markov switching models, the advantage of model parsimony becomes practically crucial.

McConnell and Perez-Quiros (2000) use a Markov switching model with mean and variance having independent switching, which does not capture the reduced recession duration observed in the post-1984 U.S. GDP data. They use the GDP growth data from 1953Q2 to 1999Q2. Although their model identifies the volatility change very well, their original model only weakly identifies the two recessions between 1960 and 1970, as shown in Figure 3 of their paper. We reestimate the independent Markov switching model as in McConnell and Perez-Quiros (2000), where the regime changes of mean growth rate and variance are independent of each other. Their original model is specified as $y_t - \mu(A_t, s_t) = \rho \cdot (y_{t-1} - \mu(A_{t-1}, s_{t-1})) + \sigma(A_t)e_t$. The independent Markov switching model produces the same graph as their Figure 3 if we use the same data as theirs (from 1953Q2 to 1999Q2), while conditional Markov chain model (using $y_t - \mu(A_t, s_t) = \rho \cdot (y_{t-1} - \mu(A_{t-1}, s_{t-1})) + \sigma(A_t)e_t$) generates a much better estimated regime and structure probabilities as shown in our Figure 2. We also notice that, if we use the data span from 1947Q2 to 2006Q4, the independent Markov switching model produce about the same nice graph as the one using our model. Moreover, we find that if we restrict $\rho = 0$ and use the specification $y_t = \mu(A_t, s_t) + \sigma(A_t)e_t$,

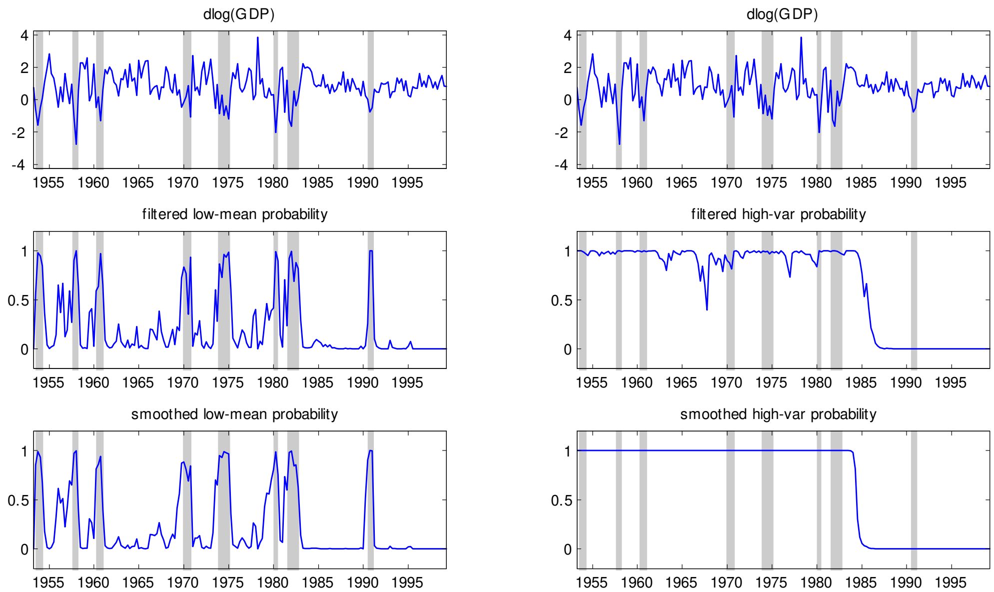

Figure 2: GDP growth and estimated state probabilities, 53Q2 – 99Q2.

both models will produce nice regime and structure probabilities, and this result holds for both the data span from 1953Q2 to 1999Q2 and the data span from 1947Q2 to 2006Q4. The independent switching model reaches a log likelihood value of -290.6113, which is very close to the one (-289.3155) achieved by conditional Markov chain model. The log likelihood ratio statistic is $-2(-290.6113+289.3155) = 2.5916$, which yields a p-value of 0.2737 according to the asymptotic $\chi^2(2)$ distribution. A detailed model comparison between these two models will be discussed in Section 4.3 using Bayes factors.

4.2 APPLICATIONS TO EMPLOYMENT GROWTH

Again, the growth rate is measured as log difference of employment multiplied by 100. We apply the conditional Markov chain model to nonfarm employment data spanned from

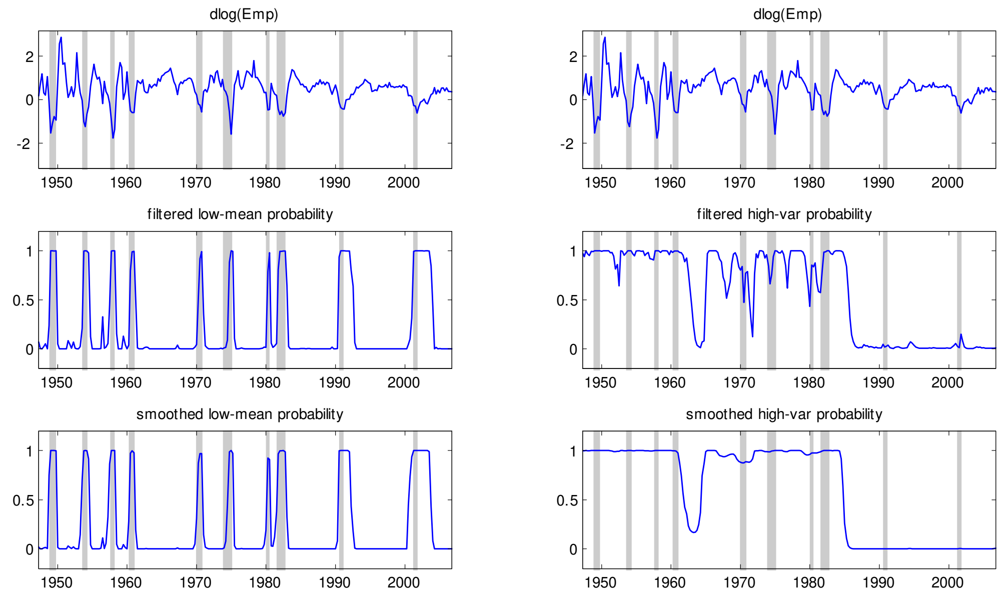

Figure 3: Nonfarm employment growth and estimated state probabilities, 47Q2 – 06Q4.

1947Q2 to 2006Q4. MLE estimates, with standard error in parentheses, are given by

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2
$\widehat{\theta}_1$	-0.6274	0.8550	-0.1354	0.5735	0.2598	0.0364
	(0.1227)	(0.0584)	(0.0587)	(0.0264)	(0.0364)	(0.0061)
θ_2	$p1$	$q1$	$p2$	$q2$	p	q
$\widehat{\theta}_2$	0.7480	0.9252	0.8910	0.9721	0.9847	0.9880
	(0.0823)	(0.0268)	(0.0738)	(0.0195)	(0.0142)	(0.0157)

where log(likelihood) = -138.8607. Regime probabilities are shown in Figure 3. Again, the NBER recession dates are well estimated in terms of smoothed or filtered recession probabilities. Post-1984 periods are identified to be under the low-variance structure. But there are also several pre-1984 years being identified as the low-variance structure, such as early 60s, 70s and 80s. The reason why we get a different result from that of GDP data is

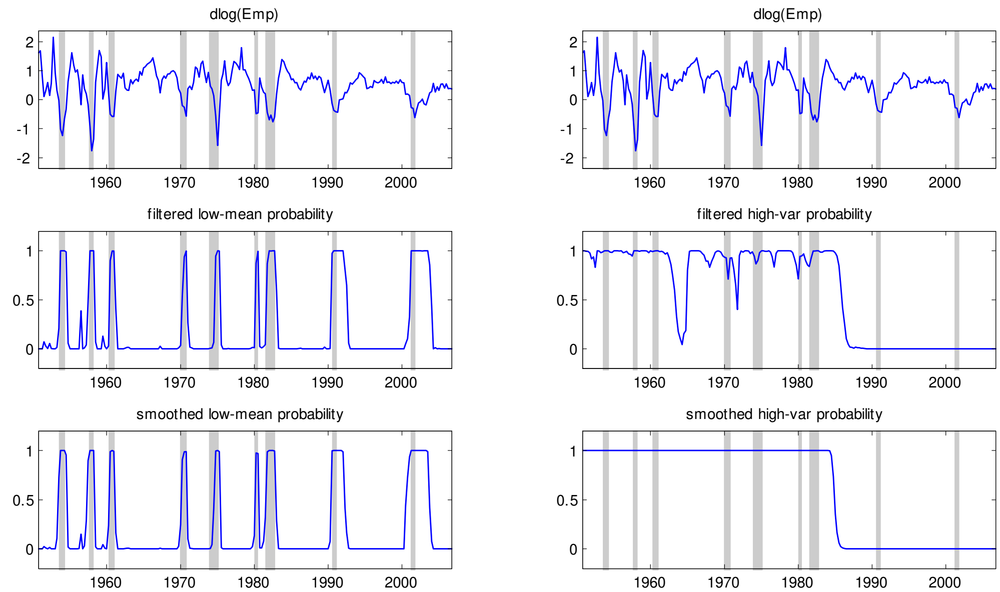

Figure 4: Nonfarm employment growth and estimated state probabilities, 50Q4 – 06Q4.

that the employment growth of early 50s appears to be extremely volatile, compared with what we observe since the 60s. The growth rate shoots up to a record high from a negative growth rate within only several quarters. The data around early 50s tend to bring up our estimates for the high variance to a certain level, such that it is hard for the simplified two-variance structure model to identify the high-variance structure unless the actual variance is high enough. To justify our conjecture, we reestimate the model using data from 1950Q4 to 2006Q4. The resulting recession and high variance probabilities are shown in Figure 4, where all pre-1984 periods are identified as the high-variance structure according to smoothed probabilities. The log(likelihood) = -108.7197, and parameter estimates are given by

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2
$\widehat{\theta}_1$	-0.5931	0.8145	-0.1336	0.5666	0.1914	0.0351
	(0.1105)	(0.0444)	(0.0542)	(0.0253)	(0.0251)	(0.0055)

θ_2	$p1$	$q1$	$p2$	$q2$	p	q
$\widehat{\theta}_2$	0.7323	0.9338	0.9008	0.9702	0.9928	1.0000
	(0.0889)	(0.0244)	(0.0673)	(0.0208)	(0.0072)	(0.0001)

4.3 MODEL COMPARISON USING BAYES FACTORS

Because it is easy to evaluate the conditional likelihood for the Markov switching model, we can readily apply Bayes factors to compare various model specifications. To fix idea, we focus on comparing the conditional Markov chain (Model 1) and independent Markov switching specification (Model 2) for the basic setup $y_t = \mu(A_t, s_t) + \sigma(A_t)u_t$ using GDP growth data, where both models provide nice regime probability estimates. The Bayes factor is defined as

$$K = \frac{p(y|M_1)}{p(y|M_2)} = \frac{\int p(y|\theta_1, M_1)p(\theta_1|M_1)d\theta_1}{\int p(y|\theta_2, M_2)p(\theta_2|M_2)d\theta_2},$$

where $p(y|\theta_j, M_j)$ is the conditional likelihood for data y given parameter θ_j under model j, and $p(\theta_j|M_j)$ is the prior density for θ_j under model j. The Bayes factor provides evidence for which model is better supported by the data. Using a wide variety of prior specifications for the parameters, we did not find unanimous support for either model. A general observation is that if we specify prior to be tight around the MLE estimators, the Bayes factors are slightly in favor of model 1 (conditional Markov chain), while if we use less informative priors, Bayes factors indicate that model 2 (independent Markov switching) is slightly preferred.

When specifying priors, we maintain the following identifying restrictions $\mu_L^H < \mu_H^H$, $\mu_L^L < \mu_H^L$, $0 < \sigma_L^2 < \sigma_H^2$. We use a reparametrization which we also applied in the MLE

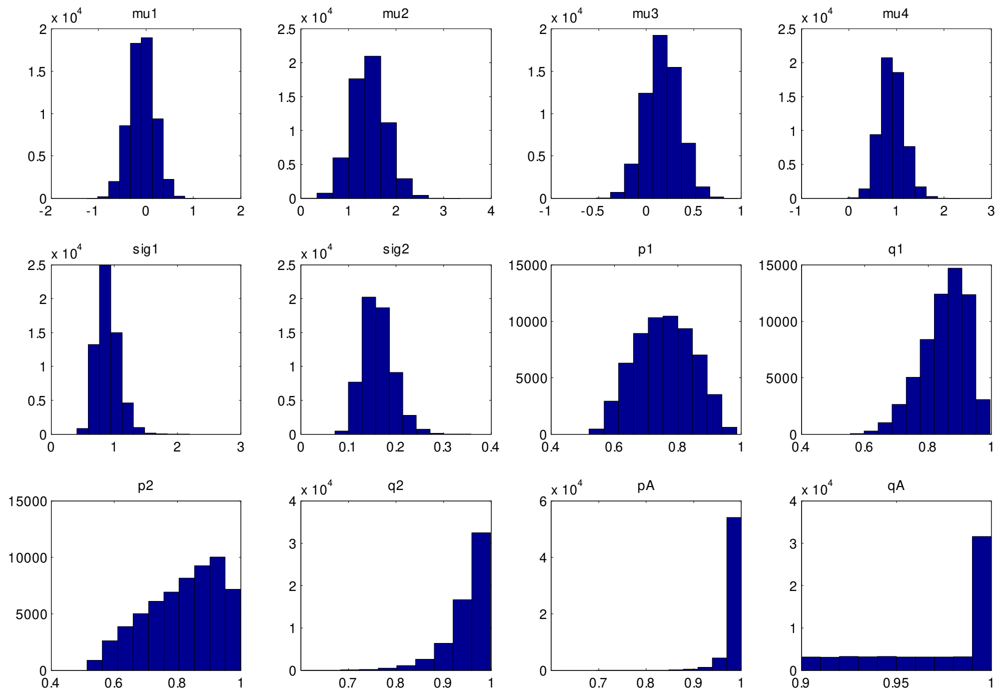

Figure 5: First set of marginal prior densities.

estimation. Let X be a vector of free parameters, and $\mu_L^H = X(1), \mu_H^H = X(1) + e^{X(2)}, \mu_L^L = X(3), \mu_H^L = X(3) + e^{X(4)}, \sigma_L^2 = e^{X(6)}, \sigma_H^2 = e^{X(5)} + e^{X(6)}$. A typical reparametrization for probability parameters is given by $p = \min\{1, (b-a)\frac{e^x}{1+e^x} + a\}$ such that $a < p < 1$.

The first set of priors for X is taken as normal, centered around its MLE estimates, with variance being 1.2 times the estimated asymptotic variance, except that we draw q (the probability of staying in the high variance state) from $\min\{1, (1.09 - 0.9) \cdot u + 0.9\}$ where $u \sim Uniform[0,1]$. The simulated prior densities for the original parameters of the conditional Markov chain model take the shape in Figure 5. The Bayes factor is around 1.41, based on 10 Monte Carlo experiments, each with $6 \cdot 10^4$ draws. When we increase the variance of the prior, the Bayes factor tends to decrease. For example, when the variance for X is taken as 1.6 times the estimated asymptotic variance, and the priors for p and q are

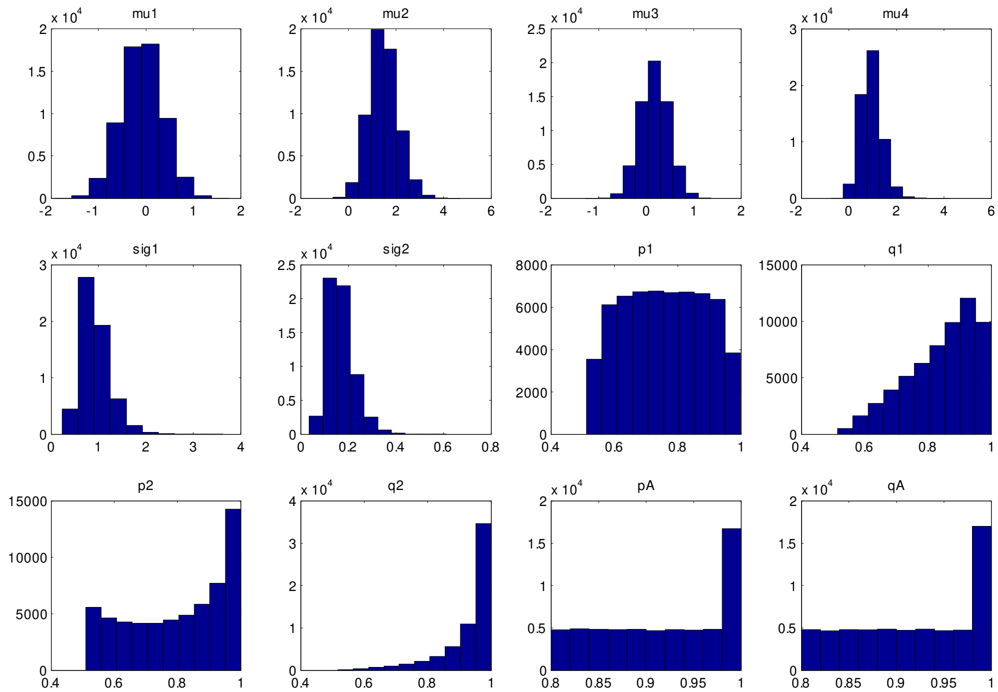

Figure 6: Second set of marginal prior densities.

both $\min\{1, (1.02 - 0.8) \cdot u + 0.8\}$ with $u \sim Uniform[0, 1]$, then the Bayes factor is around 1.0. When we further increase the variance for X to be 2 times the estimated asymptotic variance, the Bayes factor turns out to be around 0.9, where the resulting prior densities for the original parameters of the conditional Markov chain model take the shape in Figure 6. The exercise in this section suggests that there is no significant evidence from the data to choose one specification against the other.

5 MULTIPLE-EQUATION REGRESSION

Suppose the vector Z_t of economic variables follows a structural VAR,

$$\Gamma(A_t)(Z_t - \mu(A_t, s_t)) = B(A_t)(Z_{t-1} - \mu(A_{t-1}, s_{t-1})) + C(A_t)u_t,$$

where u_t is iid $N(0, I)$ and has the same dimension as Z_t. The above equation assumes that structural coefficients Γ, B, C only depend on A_t, while the mean vector μ depends on both A_t and s_t. The corresponding reduced form is represented by

$$Z_t - \mu(A_t, s_t) = \Gamma(A_t)^{-1}B(A_t)(Z_{t-1} - \mu(A_{t-1}, s_{t-1})) + \Gamma(A_t)^{-1}C(A_t)u_t, \text{ or}$$
$$Z_t - \mu(A_t, s_t) = \rho(A_t)(Z_{t-1} - \mu(A_{t-1}, s_{t-1})) + \sigma(A_t)u_t.$$

To estimate state probabilities, we may rely on the above reduced form representation instead of the structural one. The estimation approach is a direct extension of the one for single series AR process where time-varying pattern of the parameters are the same, and we may obtain the likelihood function during the filtering step of the EM algorithm. Closed-form solution is available for the ECM algorithm as in the single series case. It is worth mentioning that A_t might represent policy shift, breakthrough technology advance (e.g., better inventory control), which we do not explore here.

Likelihood of the reduced form VAR can be recursively calculated using filtering technique given in the appendix. Numerical optimization of the likelihood is hard because the number of parameters is large. In the applications, we let Z_t be the quarterly growth rate of GDP and nonfarm employment. If we do not impose any restrictions on $\rho(A_t)$ and $\sigma(A_t)$, the quasi-Newton numerical algorithm employed by Matlab 7.0 hardly converges for a wide range of starting parameter values. We also tried a block-updating scheme suggested by Sims, Waggoner and Zha (2008) by dividing the parameters into probability and non-probability parameters. We found that although the first order conditions were met fairly fast in the direction of probability parameters, they are rarely convergent in the direction of non-probability parameters such as μ, ρ and σ. In addition, the probability parameters tend to stop updating at unreasonable regions.

If we restrict $\rho(A_t)$ and $\sigma(A_t)$ as diagonal matrices, the likelihood function is easily maximized by Matlab 7.0. Such restrictions break the causality link between GDP growth and employment growth in the VAR. Their dynamics are linked together only through the states A_t and s_t. Unlike in the single series estimation where we allow different series to have different states (A, s), here in the multiple series setup, we assume that different series share the same states (A, s), i.e., they have common regime switching dynamics. We report parameter estimates, followed by plots of state probabilities.

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2	ρ_H	ρ_L
$\widehat{\theta}_{GDP}$	-0.3159	1.5156	0.0385	0.8685	0.8771	0.1837	0.1871	0.0059
	(0.2062)	(0.1496)	(0.1762)	(0.0500)	(0.1214)	(0.0272)	(0.1091)	(0.0885)
$\widehat{\theta}_{EMP}$	-0.2498	0.8182	0.0929	0.4728	0.2037	0.0214	0.6554	0.8367
	(0.1472)	(0.1077)	(0.1015)	(0.0773)	(0.0311)	(0.0031)	(0.0823)	(0.0459)

θ_2	p_1	q_1	p_2	q_2	p	q
$\widehat{\theta}_2$	0.7393	0.8863	0.7775	0.9527	0.9674	0.9739
	(0.0771)	(0.0418)	(0.1287)	(0.0244)	(0.0216)	(0.0192)

The recession probabilities along with high variance probabilities are shown in Figure 7. Although there are some wrong reports of recessions, the duration of the wrong reports is very short. The post-1984 recessions are better estimated compared with estimates from single series information.

6 IDENTIFYING RECESSION DATES USING PARTIAL DATA

There is a noteworthy lag when NBER's recession dating committee announced the most recent recession. For example, the November 2001 trough was announced July 17, 2003, while the March 1991 trough was announced December 22, 1992. Can we do better in terms of identifying recession dates with a shorter time lag of announcement? By fitting out model

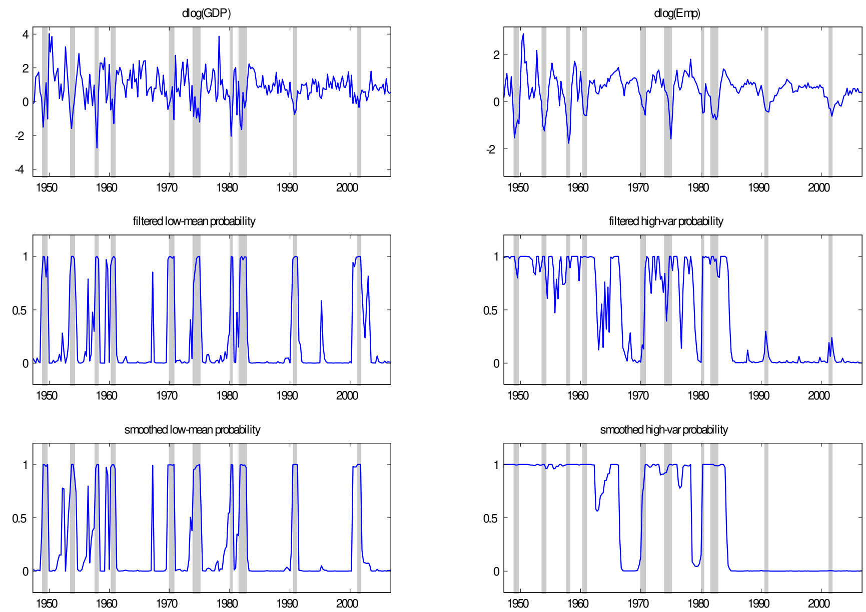

Figure 7: Estimated state probabilities through bivariate VAR, 47Q2 – 06Q4.

to GDP growth data up to 2002Q1, we are already confident to see the 2001 recession by looking at model estimates in Figure 8. The estimated recession probability by 2002Q1 is close to 1. In this exercise, we use observations up to 2002Q1 instead of the whole sample to get parameter estimates. Thus the filtered/smoothed low growth probabilities are different from those obtained in precious full sample exercise. Likewise, using GDP growth data up to 1991Q2, we are able to identify the 1991 March trough. By looking at GDP growth rate alone, and within a very simple Markov switching framework, we can provide a very nice guide to the estimates for recession dates, which precisely accords with NBER's announced recession dates. A noteworthy advantage of our model is that we do not need to wait too long to obtain a reasonable estimate of smoothed or filtered recession probabilities.

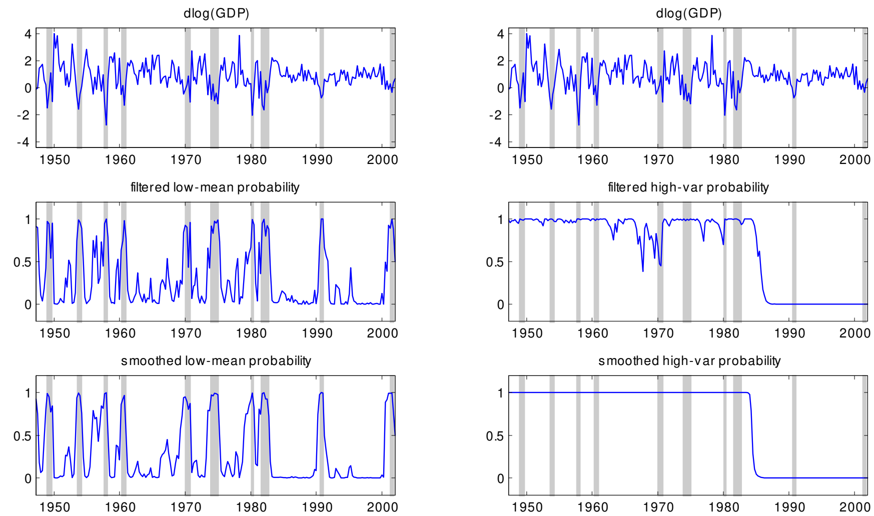

Figure 8: Identification of the 2001 trough using partial data.

7 CONCLUSION

Using a conditional Markov chain restriction on the conventional Hamilton Markov switching model, we are able to incorporate several important features of major aggregate economic time series data. Economic explanations well accord with the model structure. Empirically, the volatility decline since the 1980s is well identified and is highly persistent. Recessions during each volatility structure are also precisely identified in terms of filtered and smoothed probabilities. Compared with the independent Markov switching restriction, the empirical performance of the conditional Markov chain restriction is more robust with respect to the model specification and the choice of data span, although one cannot statistically reject one restriction against the other. The method of this paper can be applied to macro or asset pricing models with Markov switching and learning. For example, it is possible to reformulate the learning mechanism of Lettau *et al.* (2008) for their consumption process. A natural extension includes developing a multivariate model incorporating monthly data to identify and to forecast the state of an economy.

8 APPENDIX: ESTIMATION METHOD

From now on, we use $\widetilde{X}_T = \{X_1, X_2, ..., X_T\}$ to represent the full history of X up to time T. Given the time series of observables, the likelihood function is given by

$$L(\theta; \widetilde{y}_T) = f_y(\widetilde{y}_T|\theta) = \sum_{Z_T} \cdots \sum_{Z_1} f(\widetilde{y}_T, \widetilde{Z}_T|\theta)$$

where $Z_t \equiv (A_t, s_t)$. The model is estimated using the Expectation-Maximization (EM) method. A by-product of the EM procedure is that a recursive representation of the likelihood function is obtained. For a detailed description the EM algorithm, one may refer to Kim and Nelson (1999), McLachlan and Krishnan (1996), or Hamilton (1994). Here we briefly review the algorithm for the particular case in this paper. The EM algorithm will iterate between expectation and maximization steps until some convergence criteria is met.

In the expectation step, we form the following objective function,

$$\begin{aligned} Q(\theta; \widetilde{y}_T, \theta^{k-1}) &= \sum_{\widetilde{Z}_T} \log[f(\widetilde{y}_T, \widetilde{Z}_T|\theta)] \cdot \Pr(\widetilde{Z}_T|\widetilde{y}_T; \theta^{k-1}) \\ &= \sum_{\widetilde{Z}_T} \log[f(\widetilde{y}_T|\widetilde{Z}_T; \theta)] \cdot \Pr(\widetilde{Z}_T|\widetilde{y}_T; \theta^{k-1}) + \sum_{\widetilde{Z}_T} \log[\Pr(\widetilde{Z}_T|\theta)] \cdot \Pr(\widetilde{Z}_T|\widetilde{y}_T; \theta^{k-1}) \end{aligned}$$

where θ^{k-1} is the parameter estimates in step $k-1$, and $\Pr(\widetilde{Z}_T|\widetilde{y}_T; \theta^{k-1})$ can be obtained using a filtering-smoothing procedure as described in Kim and Nelson (1999).

In the maximization step, Q is maximized with respect to θ, resulting in the step-k parameter estimates θ^k.

$$\theta^k = \arg\max_{\theta} Q(\theta; \widetilde{y}_T, \theta^{k-1})$$

One favorable property of EM algorithm is that each iteration increases the likelihood value. With arbitrary initial values of the parameters, θ^0, the above two steps are iterated until θ^k converges to a local maximum of the likelihood function.

The first term of Q can be written as

$$
\begin{aligned}
&\sum\nolimits_{\widetilde{Z}_T} \log[f(\widetilde{y}_T|\widetilde{Z}_T;\theta)] \cdot \Pr(\widetilde{Z}_T|\widetilde{y}_T;\theta^{k-1}) d\widetilde{Z}_T \\
&= \sum_{t=1}^{T} \sum_{\widetilde{Z}_T} \log[f(y_t|Z_t;\theta_1)] \Pr(\widetilde{Z}_T|\widetilde{y}_T;\theta^{k-1}) d\widetilde{Z}_T = \sum_{t=1}^{T} \sum_{Z_t} \log[f(y_t|Z_t;\theta_1)] \Pr(Z_t|\widetilde{y}_T;\theta^{k-1})
\end{aligned}
$$

where for simplicity we assume $\log[f(y_t|Z_t;\theta_1] = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2(A_t)) - \frac{1}{2}\frac{(y_t-\mu(Z_t))^2}{\sigma^2(A_t)}$. Notice that in the $VAR(1)$ specification, we have $\log[f(y_t|Z_t;\theta_1] = -\frac{p}{2}\log(2\pi) - \frac{p}{2}\log\det(\Sigma(A_t)) - \frac{1}{2}[y_t - \mu(Z_t) - \rho(A_t)(y_{t-1} - \mu(Z_{t-1}))]'\Sigma(A_t)^{-1}[y_t - \mu(Z_t) - \rho(A_t)(y_{t-1} - \mu(Z_{t-1}))]$. The same EM algorithm can be applied by defining a new state variable $Z_t^* = (Z_t, Z_{t-1})$.

Similarly, the second term can be written as

$$
\begin{aligned}
&\sum_{\widetilde{Z}_T} \log[\Pr(\widetilde{Z}_T;\theta_2)] \Pr(\widetilde{Z}_T|\widetilde{y}_T;\theta^{k-1}) \\
&= \sum_{t=1}^{T} \sum_{A_t,s_t,s_{t-1}} \log[\Pr(s_t|s_{t-1},A_t;\theta_2)] \Pr(A_t,s_t,s_{t-1}|\widetilde{y}_T;\theta^{k-1}) \\
&+ \sum_{t=1}^{T} \sum_{A_t,A_{t-1}} \log[\Pr(A_t|A_{t-1};\theta_2)] \Pr(A_t,A_{t-1}|\widetilde{y}_T;\theta^{k-1})
\end{aligned}
$$

8.1 CLOSED-FORM SOLUTION

The first order conditions concerning the structure transition imply

$$
p = \frac{\sum_{t=2}^{T} \Pr(A_{t-1}=1, A_t=1|\widetilde{y}_T;\theta^{k-1})}{\sum_{t=2}^{T} \Pr(A_{t-1}=1, A_t=1|\widetilde{y}_T;\theta^{k-1}) + \sum_{t=2}^{T} \Pr(A_{t-1}=1, A_t=2|\widetilde{y}_T;\theta^{k-1})}
$$
$$
q = \frac{\sum_{t=1}^{T} \Pr(A_{t-1}=2, A_t=2|\widetilde{y}_T;\theta^{k-1})}{\sum_{t=1}^{T} \Pr(A_{t-1}=2, A_t=1|\widetilde{y}_T;\theta^{k-1}) + \sum_{t=1}^{T} \Pr(A_{t-1}=2, A_t=2|\widetilde{y}_T;\theta^{k-1})}
$$

where p is the probability of staying in high-volatility structure if $A_t = 1$ means volatility is high.

The first order conditions regarding the conditional transition matrices imply

$$p_j = \frac{\sum_{t=2}^{T} \Pr(A_t = j, s_t = s_L, s_{t-1} = s_L | \widetilde{y}_T; \theta^{k-1})}{\sum_{t=2}^{T} \Pr(A_t = j, s_t = s_L, s_{t-1} = s_L | \widetilde{y}_T; \theta^{k-1}) + \sum_{t=2}^{T} \Pr(A_t = j, s_t = s_H, s_{t-1} = s_L | \widetilde{y}_T; \theta^{k-1})}$$

$$q_j = \frac{\sum_{t=2}^{T} \Pr(A_t = j, s_t = s_H, s_{t-1} = s_H | \widetilde{y}_T; \theta^{k-1})}{\sum_{t=2}^{T} \Pr(A_t = j, s_t = s_H, s_{t-1} = s_H | \widetilde{y}_T; \theta^{k-1}) + \sum_{t=2}^{T} \Pr(A_t = j, s_t = s_L, s_{t-1} = s_H | \widetilde{y}_T; \theta^{k-1})}$$

where $j = 1, 2$ and p_j is the probability of staying in "low-mean" regime when $A = j$.

Under the normality assumption, we also have closed form solutions for mean and variance. For example the low mean under high variance structure is

$$\mu_L^H = \frac{\sum_{t=1}^{T} y_t \cdot \Pr(A_t = 1, s_t = s_L | \widetilde{y}_T; \theta^{k-1})}{\sum_{t=1}^{T} \Pr(A_t = 1, s_t = s_L | \widetilde{y}_T; \theta^{k-1})}.$$

After obtaining all mean parameters, the high variance is given by

$$\sigma_H^2 = \frac{\sum_{t=1}^{T} \left\{ (y_t - \mu_L^H)^2 \cdot \Pr(A_t = 1, s_t = s_L | \widetilde{y}_T; \theta^{k-1}) + (y_t - \mu_H^H)^2 \cdot \Pr(A_t = 1, s_t = s_H | \widetilde{y}_T; \theta^{k-1}) \right\}}{\sum_{t=1}^{T} \Pr(A_t = 1 | \widetilde{y}_T; \theta^{k-1})}.$$

In order to get the above parameter solutions, we need three types of smoothed probabilities: $\Pr(A_t, s_t | \widetilde{y}_T; \theta^{k-1}), \Pr(A_t, A_{t-1} | \widetilde{y}_T; \theta^{k-1})$ and $\Pr(A_t, s_t, s_{t-1} | \widetilde{y}_T; \theta^{k-1})$, which we briefly described as follows.

8.2 FILTERING AND SMOOTHING PROCEDURE

Let $I_t = \widetilde{y}_t = \{y_0, y_1, ..., y_t\}$ be the information available at time t. It is helpful to work with the Markov joint state $Z = (A, s)$, whose filtered and smoothed probabilities are calculated following the standard formula (see, e.g., Kim and Nelson, 1999).

Step 1. Given $\Pr(Z_{t-1} = i | I_{t-1})$, $i = 1, 2$, at the beginning of time t iteration, the one-step-ahead prediction $\Pr(Z_t = j | I_{t-1})$ is calculated as

$$\Pr(Z_t | I_{t-1}) = \sum_{Z_{t-1}} \Pr(Z_t, Z_{t-1} | I_{t-1}) = \sum_{Z_{t-1}} \Pr(Z_t | Z_{t-1}) \cdot \Pr(Z_{t-1} | I_{t-1}).$$

Step 2. Use Bayes rule to obtain the filtered probabilities

$$\Pr(Z_t|I_t) = \Pr(Z_t|I_{t-1}, y_t) = \frac{\Pr(Z_t, y_t|I_{t-1})}{f(y_t|I_{t-1})} = \frac{f(y_t|Z_t) \cdot \Pr(Z_t|I_{t-1})}{\sum_{Z_t} f(y_t|Z_t) \cdot \Pr(Z_t|I_{t-1})},$$

where by definition $I_t = \{I_{t-1}, y_t\}$.

To start the above iteration, we need an initial guess for $\Pr(Z_0|I_0)$. A good candidate is the invariant distribution computed from the last step parameter estimates θ^{k-1}.

As a by-product, we also obtain the likelihood function as $Likelihood = \sum_t \sum_{Z_t} f(y_t|Z_t) \cdot \Pr(Z_t|I_{t-1})$. For any model setup where the conditional density $f(y_t|Z_t)$ has closed-form representation, one may directly maximize the likelihood function using numerical optimization methods.

Based on the complete information, the smoothed probabilities can be obtained as follows,

$$\begin{aligned} \Pr(Z_t &= j, Z_{t+1} = k|I_T) = \frac{\Pr(Z_{t+1} = k|I_T) \cdot \Pr(Z_{t+1} = k|Z_t = j) \cdot \Pr(Z_t = j|I_t)}{\Pr(Z_{t+1} = k|I_t)}, \\ \Pr(Z_t &= j|I_T) = \sum_{k=1}^{4} \Pr(Z_t = j, Z_{t+1} = k|I_T). \end{aligned}$$

Thus given $\Pr(Z_T|I_T)$ from the filtering procedure, one may work backwards to get the smoothed probabilities for $t = 1, 2, ..., T$. Notice that $\Pr(Z_t, Z_{t-1}|I_T) = \Pr(A_t, s_t, A_{t-1}, s_{t-1}|I_T)$. Integrating out the effect of A_{t-1} will give us $\Pr(A_t, s_t, s_{t-1}|I_T)$. And $\Pr(A_t, A_{t-1}|I_T)$ is obtained by integrating out s_t and s_{t-1}.

REFERENCES

Barro R, Nakamura E, Steinsson J, Ursua J. 2009. Crises and recoveries in an empirical model of consumption disasters. Working paper.

Blanchard O, Simon J. 2001. The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity*, 1: 135-64.

Dempster A, Laird N, Rubin D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, 39(1):1-38.

Diebold FX, Lee H, Weinbach G. 1994. Regime switching with time-varying transition probabilities. C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration.* (Advanced Texts in Econometrics, C.W.J. Granger and G. Mizon, eds.), 283-302, Oxford: Oxford University Press.

Geweke J, Amisano G. 2007. Hierarchical Markov Normal mixture models with applications to financial asset returns. Working paper.

Hamilton J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57: 357-384.

Hamilton J. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton J, Lin G. 1998. Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11-5: 573-593.

Kim CJ, Nelson C. 1999. *State Space Models With Regime Switching: Classical and Gibbs Sampling Approaches With Applications*. Massachusetts: MIT Press.

Kim CJ, Nelson C. 1999. Has the U.S. economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics*, 81: 608-616.

Kim CJ, Nelson C, Piger JM. 2004. The less volatile U.S. economy: a Bayesian investigation of timing, breadth, and potential explanations. *Journal of Business and Economic Statistics*, 22: 80-93.

Lettau M, Ludvigson S, Wachter JA. 2008. The declining equity premium: what role does macroeconomic risk play? *The Review of Financial Studies* 21(4): 1653-1687.

McLachlan GJ, Krishnan T. 1996. *The EM Algorithm and Extensions*, Wiley-Interscience.

McConnell M, Perez-Quiros G. 2000. Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review*, 90: 1464-76.

Meng XL, Rubin DB. 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80 (2): 267-278.

Sims CA, Waggoner DF, Zha T. 2008. Methods for inference in large multiple-equation Markov-switching models. Forthcoming in *Journal of Econometrics.*

Warnock MVC, Warnock FE. 2000. The declining volatility of U.S. employment: was Arthur Burns right? Working paper.