Conditional Markov Chain And Its Application In Economic Time Series Analysis

Jushan Bai* and Peng Wang†

New York University, New York, NY

March 16, 2008

Abstract

Motivated by the great moderation in major U.S. macroeconomic time series, we propose a new method to study the nonstationarity of time series data. We take the long-run volatility change as a recurrent structure change, while the short-run/medium-run mean growth rate change is viewed as switching of regimes. The state of the economy is thus characterized by its structure and regime, both of which are unobserved and must be inferred from the data. We assume that structure follows an exogenous first order Markov chain. Conditioning on the structure, regimes also follow a first order Markov chain, whose transition matrix is structure-dependent. Empirical studies show that this model well identifies both medium-run regime switching and long-run structure change in U.S. macroeconomic data.

Key Words Structure, Regime switching, Conditional Markovian

*Department of Economics, New York University, New York, NY 10012. Email: jushan.bai@nyu.edu
†Department of Economics, New York University, New York, NY 10012. Email: pw513@nyu.edu

1 Introduction

The volatility of major macroeconomic variables has experienced a substantial decline since early 1980s.[1] The evidence to this decline is so striking that economists have named it the "Great Moderation". This feature should be captured in the calibration and estimation 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 exogenous income should be treated as a random process, not just one parameter to be estimated.[2]

Hamilton (1989)'s 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. In fact, the original version of the Hamilton model (constant variance) fails to simultaneously explain pre-1984 and post-1984 periods.

To allow for changing volatility, Kim and Nelson (1999) added an unknown change point to the Markov switching model. They found that, for U.S. postwar GDP growth data, not only is there a structure change toward stabilization around the first quarter of 1984, but the gap between growth rates during recessions and booms has narrowed. Lettau, Ludvigson and Wachter (2006) applied an independent Markov switching framework, as in McConnell and Perez-Quiros (2000), for consumption data, and found evidence of a shift to substantially lower volatility regimes at the beginning of 1990s. By dividing the whole sample of GDP growth time series into two subsamples according to these estimated structure breaks, we find evidence to support Kim and Nelson (1999)'s conclusion on the narrowing gap between mean growth rates. We also find evidence for a notable difference between pre-break and post-break transition probabilities for booms and recessions. Recession durations also vary across the subsamples.

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

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

One of the objectives of this paper is to establish a simple model that captures these key features of the data, such as the narrowing mean growth rate gap and changing transition probabilities[3]. A good model should be able to identify all these features endogenously, using the whole sequence of the observed data.

In order to do so, this paper explicitly models both long-run and short-run regime switches. The state of the economy is categorized into two groups: the exogenous state and the endogenous state. The exogenous state is named structure. This is designed to characterize long-run structure change. The endogenous state is called regime, and will be used to describe the short-run/medium-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 are determined by the exogenous state. The endogenous state thus follows a "conditional Markov chain", where the Markov property applies only after conditioning on the exogenous state.

The model is closely related to the conventional Markov regime switching model. It can be viewed as an extension of the Markov switching model to allow for time varying transition probabilities that are driven by some (persistent) hidden states. For simplicity, we will call our model the conditional Markov chain model. Later we will see that a version of our model can also be viewed as a way to impose structural restrictions on a Markov switching model, with only part of the states directly affecting the observed data. However, the conditional Markov chain framework admits much richer forms to model cyclical behavior. For example, we can allow mean gap and the duration for regimes to vary across different structure states. Another benefit of our model is that it keeps the number of parameters reasonably small.

Consider the following example. Let y_t be the log difference of GDP time series data,

[3] The later also implies changing recession durations.

i.e., $y_t = \log(GDP_t) - \log(GDP_{t-1})$.

$$y_t = \mu_t + e_t$$

where $e_t \sim N(0, \sigma_t^2)$. Let there be two structure states: high or low volatility. Thus the structure state belongs to $\{A_H \equiv \sigma_H^2, A_L \equiv \sigma_L^2\}$.[4] The structure state evolves over time according to an exogenous first order Markov chain with transition matrix P^A. Let there be two regime states: high/low mean $\{s_L, s_H\}$. Within high volatility structure, the pair of means is given by $\{\mu_L^H, \mu_H^H\}$; while within low volatility structure, the mean takes value in $\{\mu_L^L, \mu_H^L\}$.[5] Given that the structure is σ_H^2 for time t, the transition between regimes s_L and s_H is given by P^H; P^L is defined as transition matrix for regimes under low volatility structure. With this specification, the medium-run business cycles are characterized by switches between high and low mean growth rate of GDP, while the long-run change of volatility can be viewed as a transition from σ_H^2 to σ_L^2.[6]

To convey the main idea, we start with a simplified notation. 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 described as follows:

Under structure $A = k$, regime switching is dominated by the transition matrix P_k, where $k = 1, 2$. For example, we assume $P_1(i, j) \equiv \Pr(s_t = i | A_t = 1, s_{t-1} = j)$[7]. By assumption, the model is characterized by three transition matrices and joint distribution of initial states

[4] The subscript "H" denotes high, while "L" denotes low.

[5] We can think of μ as a function of the joint state (A_t, s_t), for example $\mu(A_H, s_L) = \mu_L^H$; and σ^2 as a function of structure A_t alone. Then we can also write the model as $y_t = \mu(A_t, s_t) + e_t$, $e_t \sim N(0, \sigma^2(A_t))$.

[6] It is a long-run change in the sense that the probability of staying in the same structure will be highly persistent, say 0.99.

[7] "Pr" means probability of an event.

(A_0, s_0), that is

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

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

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

Because both A_t and s_t are not observed, we treat them as missing data and apply an expectation-maximization algorithm to estimate the entire model. By applying this model to US post-war data on GDP, 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 precisely 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 Markov Switching Models

The above model is designed such that it can identify both short-run/medium-run regime switching and long-run structure change.[8] It admits rich model features while keeps a rea-

[8]Notice that we can also model this long-run and short-run transitions within the conventional Markov regime switching framework, with four regime states, i.e., $s_t \in \{(\sigma_H^2, \mu_L^H), (\sigma_H^2, \mu_H^H), (\sigma_L^2, \mu_L^L), (\sigma_L^2, \mu_H^L)\}$. The problem with this setup is that there are 12 probability parameters concerning the transition matrix of these 4 states, whose MLE estimates are hard to find. Also there might be multiple local maximum for the likelihood function, and it is hard to achieve a reasonable local maximum. Instead, by explicitly modelling the long-run and short-run regime changes, our model only involves 6 parameters concerning the transition matrix, which greatly reduces the computation burden. Thus our model can also be viewed as a parsimonious way to model the above 4-state Markov regime switching problem.

sonably parsimonious model structure. For example, 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, 2006) 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)*[9]*. 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 restrict our model such that* $p_1 = p_2$ *and* $q_1 = q_2$*, then the resulting model is equivalent to the independent Markov switching model, where the switching of regime* s_t *no longer depends on the structure state* A_t*. We can see that the independent switching model essentially requires the conditional transition matrix for regime* s_t *to be the same across different structures, i.e.,* $P_1 = P_2 = \begin{pmatrix} p_1 & 1-q_1 \\ 1-p_1 & q_1 \end{pmatrix}$, *with no restrictions imposed on transition matrix for structure* A_t*.*

Recently, Geweke et al. (2007) proposed a Hierarchical Markov Normal Mixture model (HMNM) to study financial asset returns. Our model also includes HMNM as a special case by imposing some restrictions on the transition matrices.

Example 3 *(HMNM) If we restrict the conditional transition matrix for regimes such that the diagonal terms add up to one and each row contains the same elements, then our model becomes a HMNM model. In particular, the resulting transition matrices take the following*

[9] They considered a general setup to allow 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.

forms: $P^A = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$, $P_1 = \begin{pmatrix} p_1 & p_1 \\ 1-p_1 & 1-p_1 \end{pmatrix}$, $P_2 = \begin{pmatrix} p_2 & p_2 \\ 1-p_2 & 1-p_2 \end{pmatrix}$. *In particular, HMNM restricts the conditional transition kernel such that it degenerates:* $\Pr(s_t = i|A_t = k, s_{t-1} = j) = \Pr(s_t = i|A_t = k)$, *for all* i, k, j. *Conditioning on a specific structure* $A_t = k$, *the regime* s_t *follows a mixture of Normals distribution.*

3 General Model Setup and Properties

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 characterizing transitions of s_t conditional on A_t, each being of dimension $N \times N$.

We can define a joint state $Z_t \equiv (A_t, s_t)$, which is proved to be first order Markovian (see the Lemma). A typical realization of joint state is given by $(A_t = m, s_t = n)$, $m \in \{1, ..., M\}$, $n \in \{1, ..., N\}$. The number of joint states is given by MN.

Lemma 4 *The joint state* Z_t *is first order Markovian, i.e.,*

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

Proof: Write out the probability using Bayes rule:

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

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

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

The second equality comes from two model assumptions. Firstly, conditional on the structure state A_t, s_t is first order Markovian. This implies that conditioning on A_t, s_{t-1} forms a sufficient statistic for all past history for predicting s_t, or $\Pr(s_{t+1}|A_{t+1}, A_t, s_t, A_{t-1}, s_{t-1}, ..., A_0, s_0) =$

$\Pr(s_{t+1}|A_{t+1}, s_t)$. Secondly, A_t is assumed to be exogenously first order Markovian, which means A_{t-1} forms a sufficient statistic for predicting A_t, or $\Pr(A_{t+1}|A_t, s_t, A_{t-1}, s_{t-1}, ..., A_0, s_0) = \Pr(A_{t+1}|A_t)$. Q.E.D.

Then the $MN \times MN$ transition matrix P^Z characterizing the Markov process $\{Z_t\}$ can be constructed as follows[10]

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

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.

Example 5 *Consider a simple case with only two structures ($A_t = 1$ or 2) and two regimes ($s_t = 1$ or 2), where* $P^A = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$, $P_1 = \begin{pmatrix} p_1 & 1-q_1 \\ 1-p_1 & q_1 \end{pmatrix}$, $P_2 = \begin{pmatrix} p_2 & 1-q_2 \\ 1-p_2 & q_2 \end{pmatrix}$. *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*

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

[10] We can also let transition from s_{t-1} to s_t to be 1st-order stationary conditional on past structure A_{t-1}. For example, we can let $P_1(i,j) \equiv \Pr(s_t = i|A_{t-1} = 1, s_{t-1} = j)$. Then we can carry the same analysis throughout with the new joint state $\hat{Z}_t = (A_{t-1}, s_t)$— we will have $\Pr(A_t, s_{t+1}|A_{t-1}, s_t) = \Pr(s_{t+1}|A_t, s_t) \cdot \Pr(A_t|A_{t-1})$. Also notice that the structure is unobservable, so all the properties developed as follows do no depend on whether A_{t-1} or A_t contribute to the evolution from s_{t-1} to s_t.

Denote the time t unconditional distribution of A_t and Z_t by

$$\pi_{A,t} \equiv [\Pr(A_t = 1), ..., \Pr(A_t = M)]'$$
$$\pi_{Z,t} \equiv [\Pr(Z_t = (1,1)), \Pr(Z_t = (1,2)), ..., \Pr(Z_t = (M, N-1)), \Pr(Z_t = (M,N))]'$$

then from the Markovian property,

$$\pi_{A,t+1} = P^A \cdot \pi_{A,t}$$
$$\pi_{Z,t+1} = P^Z \cdot \pi_{Z,t}$$

It is worth mentioning that the marginal process $\{s_t\}$ by itself is not Markovian, i.e. $\Pr(s_{t+1}|s_t, ..., s_0) \neq \Pr(s_{t+1}|s_t)$. In the business cycle example, regime represents low or high growth rate, while structure represents high or low variance state. Under Hamilton's (1989) original setup, low (high) growth rate means recession (boom), and regimes are assumed to be an exogenous first order Markov process. Suppose all historical regimes are observed. Then Hamilton's model assumption implies that we can drop all except the most recent observation of regimes, and still will not lose any information in predicting the next period regime. While under our conditional Markovian assumption, since all information on historical regimes is useful for making inference on the variance state, dropping any historical information would result in a less efficient prediction for future regimes. The non-Markovian property for $\{s_t\}$ is summarized in the following proposition.

Proposition 6 *In the conditional Markov chain model, the marginal process of regime* $\{s_t\}$ *is not Markovian*[11]*, i.e.,*$\Pr(s_{t+1}|I_t^S) \neq \Pr(s_{t+1}|s_t)$ *for* $t \geq 1$*, where* $I_t^S = \{s_0, s_1, ..., s_t\}$

[11] Throughout the paper, we use Markovian to denote first order Markov property.

Proof:

$$\begin{aligned}
\Pr(s_{t+1}|I_t^S) &= \sum_{A_{t+1}}\sum_{A_t} \Pr(A_{t+1}, s_{t+1}, A_t|I_t^S) \\
&= \sum_{A_{t+1}}\sum_{A_t} \Pr(A_{t+1}, s_{t+1}|A_t, I_t^S) \cdot \Pr(A_t|I_t^S) \\
&= \sum_{A_{t+1}}\sum_{A_t} \Pr(A_{t+1}, s_{t+1}|A_t, s_t) \cdot \Pr(A_t|I_t^S)
\end{aligned}$$

Here $\Pr(A_{t+1}, s_{t+1}|A_t, I_t^S) = \Pr(A_{t+1}, s_{t+1}|A_t, s_t)$ because the joint state (A, s) is 1^{st}-order Markovian. But $\Pr(A_t|I_t^S)$ generally depends on the whole history I_t^S. Q.E.D.

The intuition is as stated in the business cycle example. Under conditional Markovian assumption, all information on historical regimes is useful for inferring A_{t+1}. Thus by using today's regime information alone would result in a less accurate prediction for future regimes because of a less efficient estimation of A_{t+1}. Precisely estimating A_{t+1} is crucial because it determines the pattern of transition probabilities from s_t to s_{t+1} in our model.

Example 7 *To see how* $\Pr(A_t|I_t^S)$ *depends on the whole history* I_t^S, *the simplest example would be* $\Pr(A_1|s_1, s_0) \neq \Pr(A_1|s_1)$. *Suppose the initial distribution* $\pi_{Z,0}$ *is given, from which we can compute* $\Pr(A_0|s_0) = \frac{\Pr(A_0,s_0)}{\sum_{A_0}\Pr(A_0,s_0)}$. *Then we can update* $\Pr(A_1|s_0) = \sum_{A_0}\Pr(A_1, A_0|s_0) = \sum_{A_0}\Pr(A_1|A_0)\Pr(A_0|s_0)$, *which depends on* s_0. *We can further update* $\Pr(A_1|s_1, s_0) = \frac{\Pr(A_1,s_1|s_0)}{\Pr(s_1|s_0)} = \frac{\Pr(s_1|A_1,s_0)\Pr(A_1|s_0)}{\sum_{A_1}\Pr(s_1|A_1,s_0)\Pr(A_1|s_0)}$ *which depends on both* s_0 *and* s_1. *The following is a numerical example. Let* $P_{s0} = [\ 1/3 \quad 1/5; \quad 2/3 \quad 4/5\ ]$, *where* $P_{s0}(i,j) = \Pr(A_0 = i|s_0 = j)$. *Assume the conditional transition matrix and structure transition matrix are given by*

$$P^A = \begin{pmatrix} 3/5 & 1/5 \\ 2/5 & 4/5 \end{pmatrix}, P_1 = \begin{pmatrix} 3/4 & 1/3 \\ 1/4 & 2/3 \end{pmatrix}, P_2 = \begin{pmatrix} 11/12 & 2/3 \\ 1/12 & 1/3 \end{pmatrix}$$

Using the formula given at the beginning, we obtain two matrices

$$P_{A1} = \begin{pmatrix} 0.2903 & 0.1628 \\ 0.6000 & 0.4375 \end{pmatrix}, P_{A2} = \begin{pmatrix} 0.7097 & 0.8372 \\ 0.4000 & 0.5625 \end{pmatrix}$$

where $P_{Ak}(i,j) = \Pr(A_1 = k|s_1 = i, s_0 = j)$, $k,i,j = 1,2$. *We can see that all elements in* P_{A1} *and* P_{A2} *are different, meaning that all the conditioning information of "s" is useful in predicting* A_1. *For example* $\Pr(A_1 = 1|s_1 = 1, s_0 = 1) \neq \Pr(A_1 = 1|s_1 = 1, s_0 = 2)$ *such that* s_0 *is relevant for inferring* A_1.

In the long run, the whole dynamic system is stationary, and detailed properties are analyzed in Appendix A.

3.1 A state-space representation

Suppose in the general setup, $y_t = \mu(A_t, s_t) + e_t$, $e_t \sim N(0, \sigma^2(A_t))$. The process for joint state Z_t admits an AR(1) representation with a specially defined system error

$$X_{t+1} = P^Z X_t + V_{t+1}$$

where $X_t \in \{e_1 = [1,0,...,0]', e_2 = [0,1,0,...,0]', ..., e_N = [0,...,0,1]'\}$, with e_j representing j^{th} joint state, $j = 1, ..., MN$. P^Z is the transition matrix for (A_t, s_t), and $V_{t+1} \equiv X_{t+1} - P^Z X_t$ satisfies $E(V_{t+1}|X_t) = 0$[12].

The process for observed y_t is given by $y_t = \mu(A_t, s_t) + \sigma(A_t) \cdot \varepsilon_t$, $\varepsilon_t \sim N(0,1)$. If we define two $MN \times 1$ constant vectors $\bar{\mu}$ and $\bar{\sigma}$, whose j^{th} elements correspond to $\mu(Z_t = e_j)$

[12] I.e., V_{t+1} is a martingale process adapted to X_t.

and $\sigma(Z_t = e_j)$, then the measurement equation takes a nonlinear form

$$y_t = \bar{\mu}' \cdot X_t + \bar{\sigma}' \cdot X_t \cdot \varepsilon_t, \text{ with } \varepsilon_t \sim N(0,1)$$

4 Applications to economic time series data

The GDP growth data we considered span from the second quarter of 1947 to the fourth quarter of 2006.[13] To see how the model works for the simplest setup, we abstract from autoregressive components for economic variable for the moment, and concentrate on the basic mixture-of-normals setup as in section 1 and 2.

The employment growth data span from the second quarter of 1947 to the fourth quarter of 2006.

4.1 GDP

The growth rate is measured as the difference of log-valued GDP multiplied by 100. We estimate a conditional Markov chain model with two pairs of means corresponding to the regime state, and two states of variance (i.e., with mean gap). The estimation procedure features a two-step process. We use EM algorithm as the first step to obtain initial estimates for parameters, while in the second we directly maximize the likelihood function to refine our estimates and obtain the standard errors.

Maximum likelihood parameter estimates are given by

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

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)

where standard errors are shown in parenthesis.

log(likelihood) = -188.6801

Implied transition probabilities for the joint states[14] are given by

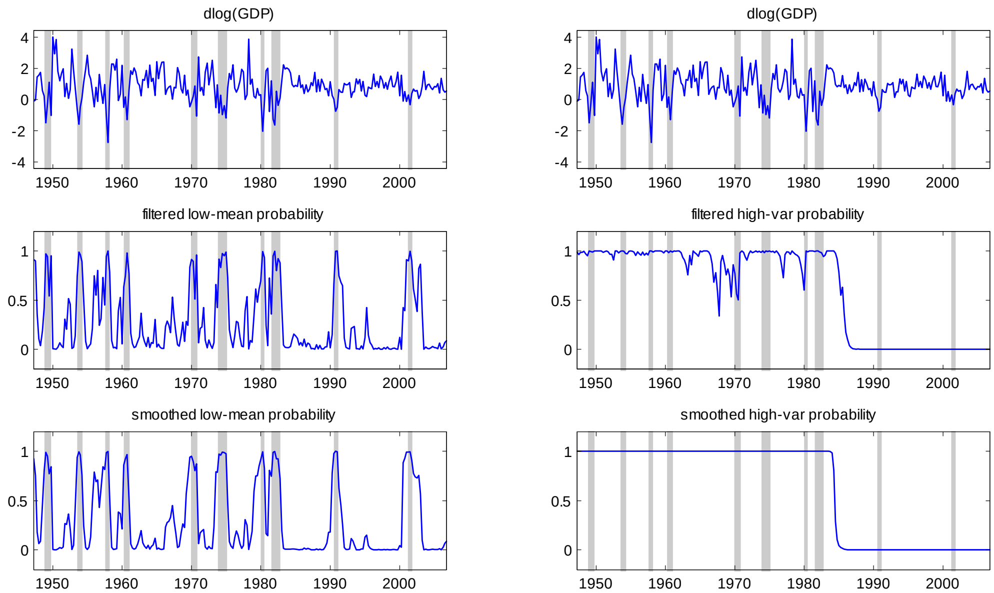
$$P^Z = \begin{pmatrix} 0.7522 & 0.1354 & 0.0000 & 0.0000 \\ 0.2412 & 0.8580 & 0.0000 & 0.0000 \\ 0.0055 & 0.0002 & 0.8332 & 0.0370 \\ 0.0011 & 0.0064 & 0.1668 & 0.9630 \end{pmatrix}$$

Probabilities for low growth regimes[15] and high variance structures are shown in the following figure. The shaded areas are NBER dated recessions. Reference period: 1947Q2 –

[14] The order of the four states are "high variance, low mean", "high variance, high mean", "low variance, low mean" and "low variance, high mean".

[15] 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 amazingly well, by just looking at a single time series.

2006Q4.

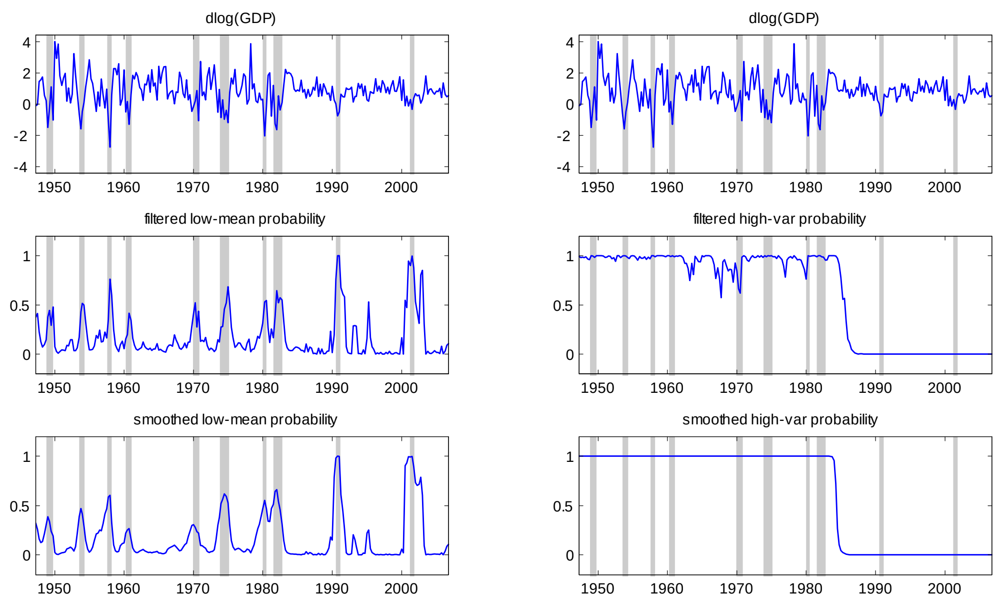


We can see that not only the NBER recessions are very precisely estimated[16], but the low frequency movement of variance is well captured[17]. The parameter estimates also suggest that besides a substantial volatility drop, a changing mean growth gap is also an important feature. Along with the added assumption that recession duration depends on volatility structure, this model is able to provide much more precise recession probabilities than existing literature. 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. Although their model identifies the volatility change very well, the recessions are estimated with much less precision as we will see from the following exercise.

[16] Here in terms of filtered and smoothed recession probabilities.

[17] The smoothed probabilities for high variance structures around the turning point are given by $\Pr(1984Q1|I_T) = 0.9821$, $\Pr(1984Q2|I_T) = 0.8075$, $\Pr(1984Q3|I_T) = 0.2880$, and $\Pr(t|I_T) > 0.99$ for $t \leq 1983Q4$.

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. Independent Markov switching model can be seen as a special case of our conditional chain model by forcing the transition matrix for mean growth rate to be the same across different structures. The graph shows that although the persistent volatility change is well identified, the recessions before 1984 are only weekly identified in terms of lower filtered or smoothed recession probabilities.



4.2 Employment

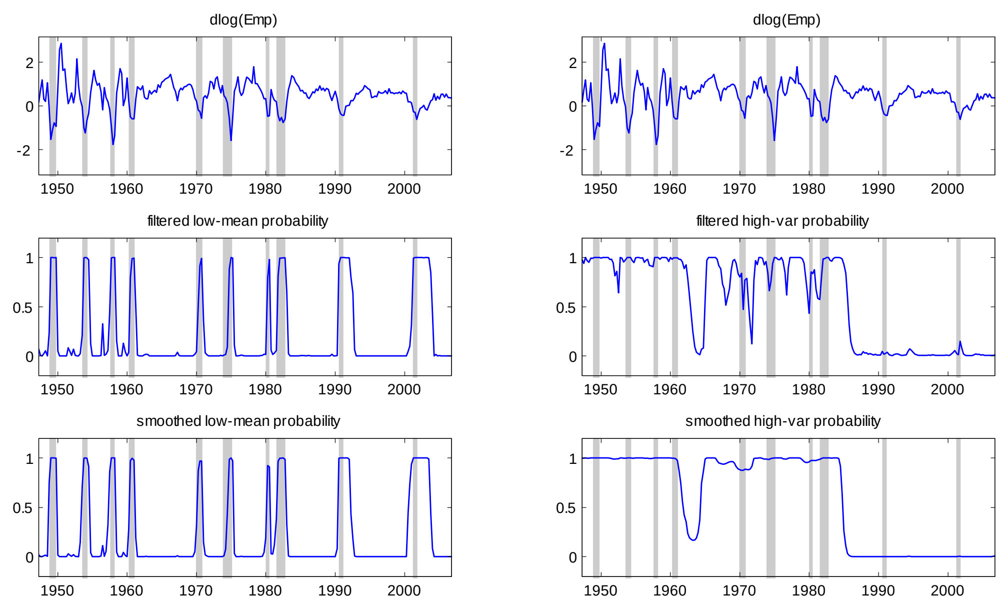
Again, the growth rate is measured as log difference of employment multiplied by 100. We apply the model to nonfarm employment data spanned from 1947Q2 to 2006Q4.

Two-step parameter estimates, with standard error in parentheses, are given by

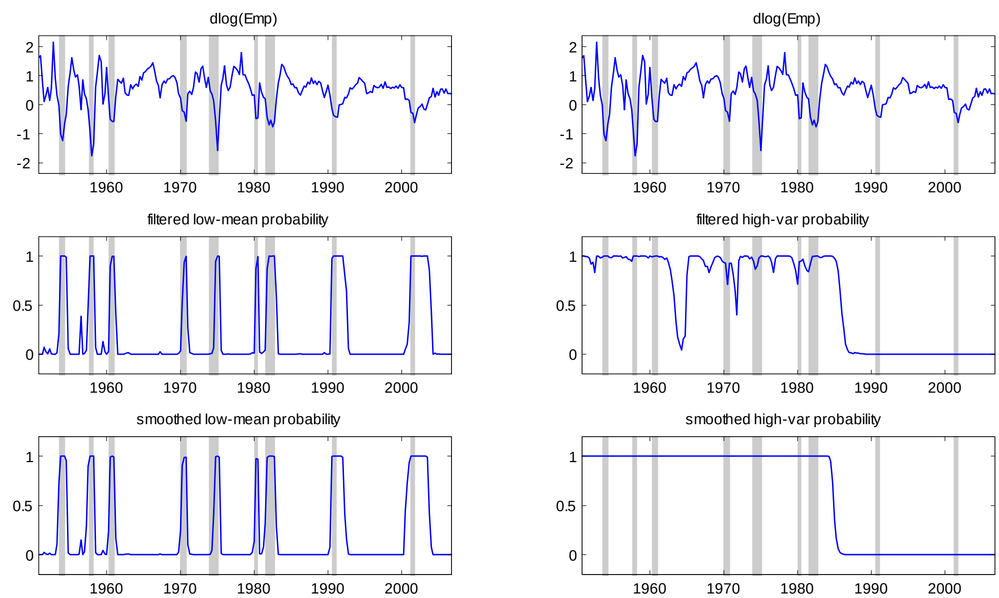
θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2
$\widehat{\theta}_1$	-0.6274	0.8550	-0.1354	0.5735	0.2598	0.0364
	(0.1227)	(0.0584)	(0.0587)	(0.0264)	(0.0364)	(0.0061)
θ_2	$p1$	$q1$	$p2$	$q2$	p	q
$\widehat{\theta}_2$	0.7480	0.9252	0.8910	0.9721	0.9847	0.9880
	(0.0823)	(0.0268)	(0.0738)	(0.0195)	(0.0142)	(0.0157)

log(likelihood) = -138.8607

Recession and high variance probabilities are shown in the following graph.



Again, the NBER recession dates are very precisely estimated in terms of smoothed or filtered recession probabilities. Post-1984 periods are identified to be under low-variance structure. But there are also several pre-1984 years to be identified as low variance structure, such as early 60s, 70s and 80s. The reason why we get different result from that of GDP data is 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 high variance to a certain level, such that it is hard for the simplified two-variance structure model to identify 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 as follows



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)

log(likelihood) = -108.7197

4.3 Adding an autoregressive component

We modify the above model by adding an autoregressive (AR) component. The estimated AR coefficient for GDP data is rather small. But for employment data, it is very large. This finding suggests that the model in previous sections is a good approximation for modeling GDP time series, but an AR component is needed for a better characterization of the employment data.

4.3.1 GDP with AR component

The model is specified as

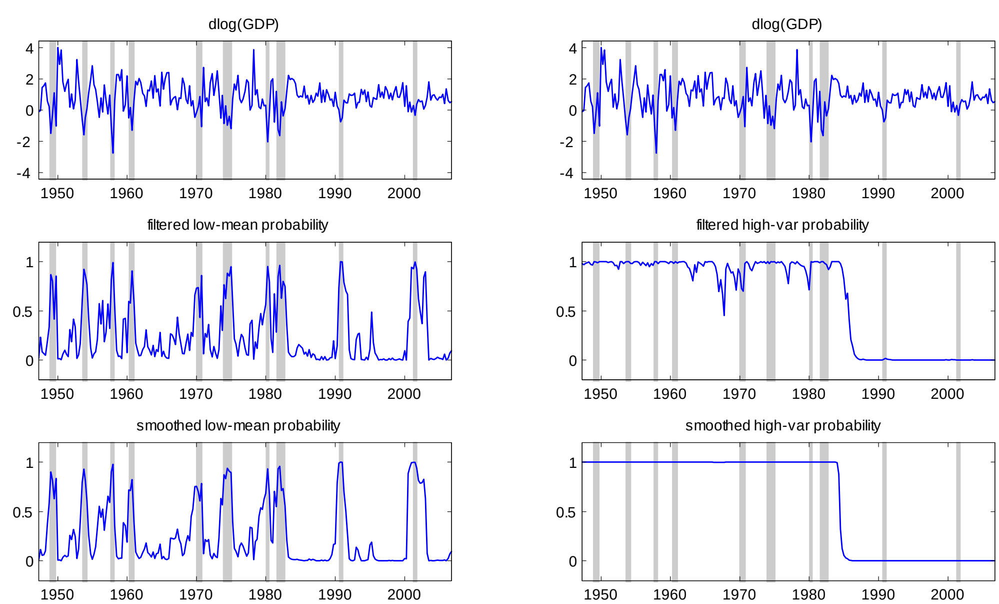
$$y_t - \mu(A_t, s_t) = \rho(A_t) \cdot (y_{t-1} - \mu(A_{t-1}, s_{t-1})) + e_t, \quad e_t \sim N(0, \sigma^2(A_t)).$$

Parameter estimates are about the same as in previous sections, with very small AR coefficients,

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2	ρ_H	ρ_L
$\widehat{\theta}_1$	-0.0583	1.3371	0.1752	0.8953	0.9369	0.1543	0.2336	-0.0818
	(0.3745)	(0.2392)	(0.1308)	(0.0505)	(0.1738)	(0.0253)	(0.1312)	(0.1213)
θ_2	p_1	q_1	p_2	q_2	p	q		
$\widehat{\theta}_2$	0.7397	0.8703	0.8328	0.9627	0.9933	0.9997		
	(0.1405)	(0.0764)	(0.1157)	(0.0283)	(0.0067)	(0.0020)		

log(likelihood) = -285.2330

Recession probabilities are also similar to previous results



4.3.2 Employment with AR component

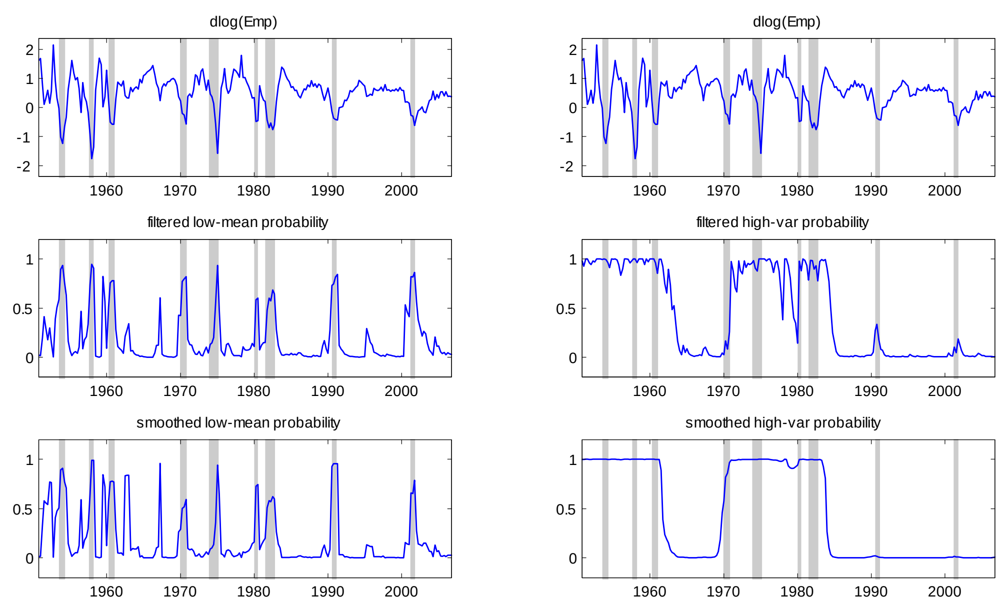
Data range is from 1950Q4 to 2006Q4. The model is specified as $y_t - \mu(A_t, s_t) = \rho(A_t) \cdot (y_{t-1} - \mu(A_{t-1}, s_{t-1})) + e_t$, $e_t \sim N(0, \sigma^2(A_t))$. Parameter estimates are given by

θ_1	μ_L^H	μ_H^H	μ_L^L	μ_H^L	σ_H^2	σ_L^2	ρ_H	ρ_L
$\widehat{\theta}_1$	-0.2541	0.6796	0.1102	0.4989	0.1986	0.0223	0.7132	0.8909
	(0.2607)	(0.2109)	(0.1530)	(0.1343)	(0.0622)	(0.0037)	(0.0944)	(0.0401)

θ_2	$p1$	$q1$	$p2$	$q2$	p	q
$\widehat{\theta}_2$	0.7297	0.8917	0.6916	0.9599	0.9778	0.9899
	(0.1323)	(0.0680)	(0.1846)	(0.0248)	(0.0158)	(0.0101)

log(likelihood) = -52.9063

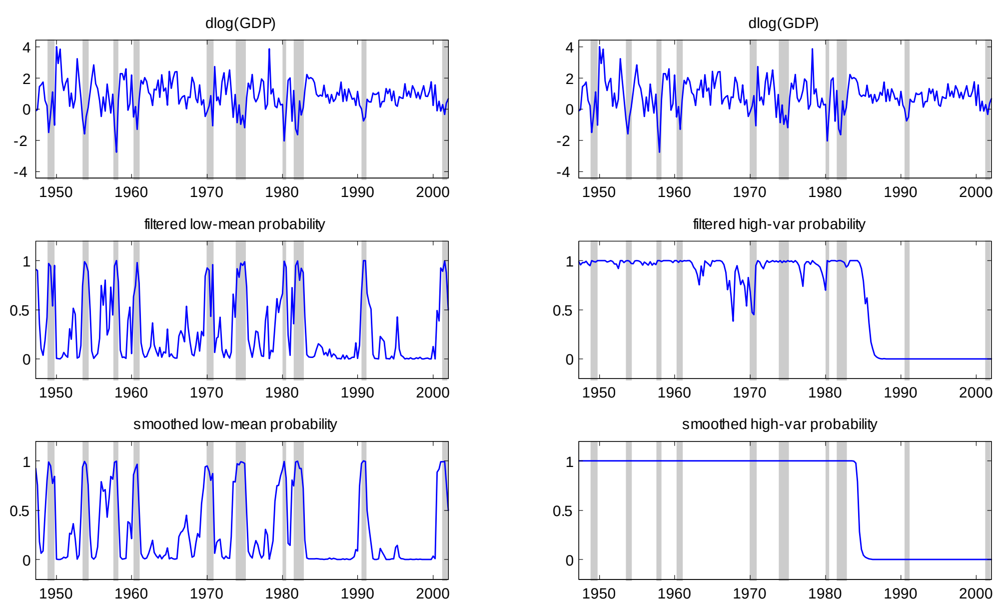
Estimates for recession probabilities



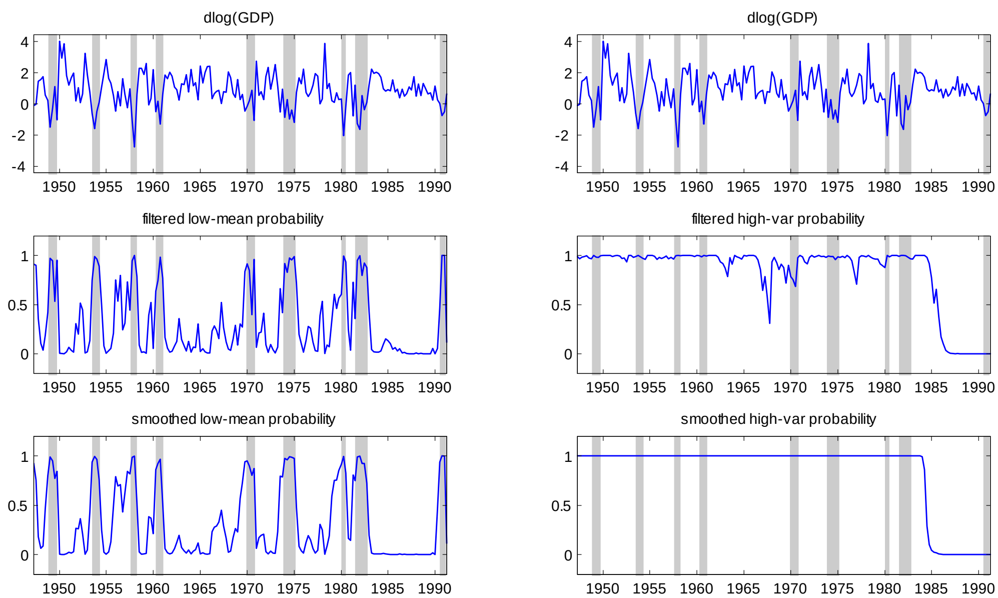
The above figure shows that estimated duration of recessions fit the data better than the case without AR component, especially for post-1984 periods. A large part of 1960s is identified to be under the low-variance structure.

5 Identifying recessions dates using partial data

There is a noteworthy lag when NBER 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 to GDP growth data up to 2002Q1, we are already confident to see the 2001 recession by looking at model estimates. 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 section 4.1.



Using GDP growth data up to 1991Q2, we are also able to identify the 1991 March trough; the estimated recession probability is close to 1, as is shown below.



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.

6 Conclusion

Using a conditional Markov chain model, we are able to incorporate several important features for major aggregate economic time series data. Economic explanations well accord with the model structure. Empirically, the volatility decline since the 80s is well identified and is highly persistent. Recessions during each volatility structure are also precisely identified in terms of filtered and smoothed probabilities. 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. (2006) for their consumption process. Our next step includes developing a multivariate model incorporating monthly data to identify and forecast state of the economy.

References

[1] Blanchard, Olivier, and John Simon (2001): "The Long and Large Decline in U.S. Output Volatility," *Brookings Papers on Economic Activity*, 1, 135-64.

[2] Dempster, Arthur, Nan Laird, and Donald Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, Series B, 39(1):1-38.

[3] Diebold, Francis X., .-H. Lee and G. Weinbach (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.

[4] Geweke, John, and Gianni Amisano (2007): "Hierarchical Markov Normal Mixture Models With Applications To Financial Asset Returns," Working Paper.

[5] Hamilton, James. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

[6] — (1994): *Time Series Analysis*, Princeton: Princeton University Press.

[7] Kim, Chang-Jin, and Charles Nelson (1999), *State Space Models With Regime Switching: Classical and Gibbs Sampling Approaches With Applications*, Massachusetts: MIT Press.

[8] — (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.

[9] Kim, Chang-Jin, Charles Nelson, and Jeremy M. Piger (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.

[10] Lettau, Martin, Sydney Ludvigson, and Jessica A. Wachter (2006): "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?" Forthcoming in *The Review of Financial Studies.*

[11] McLachlan, Geoffrey J. and Thriyambakam Krishnan (1996), *The EM Algorithm and Extensions*, Wiley-Interscience

[12] McConnell, Margaret, and Gabriel Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980s?" *American Economic Review*, 90, 1464-76.

[13] Warnock, M.V. Cacdac, and Francis Warnock (2000): "The Declining Volatility of U.S. Employment: Was Arthur Burns Right?," Working Paper.

7 Appendix A: Properties of the regime process

Because the number of joint states is finite, there exist invariant distributions[18] $\pi_{Z,t} \to \pi_Z$ and $\pi_{A,t} \to \pi_A$, as $t \to \infty$. These two are mutually consistent in that they are related by the equality $\sum_{n=1}^{N} \pi_Z(A_t = m, s_t = n) = \pi_A(A_t = m)$.[19] Also the marginal stationary distribution of s is obtained via $\sum_{m=1}^{M} \pi_Z(A_t = m, s_t = n) = \pi_s(s_t = n)$.

Lemma 8 *The conditional probability* $\Pr(s_{t+1}|s_t)$ *is asymptotically time invariant.*

To see this, let's assume joint stationarity of (A_t, s_t) has been achieved. Then the joint distribution of $Z_t = (A_t, s_t)$ does not change over time, so does the marginal distribution for s_t. Moreover the conditional probability $\Pr(A_t|s_t)$ is time invariant. Thus, the conditional probability

$$\begin{aligned}\Pr(s_{t+1}|s_t) &= \int_{A_{t+1}} \int_{A_t} \Pr(A_{t+1}, s_{t+1}, A_t|s_t) dA_t dA_{t+1} \\ &= \int_{A_{t+1}} \int_{A_t} \Pr(A_{t+1}, s_{t+1}|A_t, s_t) \Pr(A_t|s_t) dA_t dA_{t+1} \equiv g(s_{t+1}, s_t)\end{aligned}$$

is also time invariant, where $g(\cdot, \cdot)$ is a function of s_{t+1} and s_t, not depending on time. To get intuition of the result, first notice that the transition matrix carries marginal distribution from the current to the next period. By assumption the unconditional distribution of joint state (A_t, s_t) already achieves stationarity, and thus will not change over time. As a result, the marginal distribution of regimes will also be the same over time. However, it is important to notice that although the conditional probability $\Pr(s_{t+1}|s_t)$ is time invariant given that the joint state has achieved its long-run distribution, the marginal process for s_t is not Markovian

[18] If we assume that elements of $P_1, ..., P_M$ and P^A are all strictly positive, then elements of P^Z are strictly positive, which implies the invariant distribution for joint state Z is also unique.

[19] Here $\pi_Z(A = m, s = n)$ means the probability of the event $(A = m, s = n)$ under invariant distribution π_Z.

as we have discussed. Even in the long run, the dynamic behavior will be different from that of a first Markov process with transition probabilities given by $g(s_{t+1}, s_t)$.

Generally, the conditional probability $\Pr(s_{t+1}|s_t)$ depends on the initial distribution of joint state and changes over time.

To see this, first notice that the conditional probability $\Pr(A_t|s_t) = \frac{\Pr(A_t,s_t)}{\Pr(s_t)}$ changes over time. The numerator is an element of the $MN \times 1$ probability vector $\pi_{Z,t}$, which is obtained via $\pi_{Z,t} = (P^Z)^t \cdot \pi_{Z,0}$[20]. The denominator, if we regard it as element of a $N \times 1$ probability vector $\pi_{s,t}$, is obtained directly by summing up corresponding elements of $\pi_{Z,t}$[21]. We can see that $\Pr(A_t|s_t)$ depends on both the initial distribution of joint state, which can be treated either as given or as model parameters, and the time t; thus so does P_t^S, whose elements is given by $\Pr(s_{t+1}|s_t) = \sum_{A_{t+1}} \sum_{A_t} \Pr(A_{t+1}, s_{t+1}, A_t|s_t) = \sum_{A_{t+1}} \sum_{A_t} \Pr(A_{t+1}, s_{t+1}|A_t, s_t) \cdot \Pr(A_t|s_t)$.

8 Appendix B: Estimation of the model

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

$$\begin{aligned} L(\theta; \widetilde{y}_T) = f_y(\widetilde{y}_T|\theta) &= \int_{\widetilde{(A,s)}_T} f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T|\theta) d\widetilde{(A,s)}_T \\ &= \sum_{(A,s)_T} \sum_{(A,s)_{T-1}} \cdots \sum_{(A,s)_1} f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T|\theta) \end{aligned}$$

where $(A, s)_t = (A_t, s_t)$. In the simple case with a binary structure and binary regime, involves computation and summation of 4^T terms. It will be a great computational burden to do so. Here we turn to the Expectation-Maximization method to estimate the model. A

[20] Notice here that if $\pi_{Z,0}$ is the stationary distribution of the joint state, then $\pi_{Z,t} = \pi_{Z,0}$ for all t, and the previous corollary applies.

[21] For example, for $n = 1, ..., N$, we have $\sum_{m=1}^{M} \pi_{Z,t}(A_t = m, s_t = n) = \pi_{s,t}(s_t = n)$.

by-product of the EM procedure is that the likelihood function is recursively computed.

8.1 EM Algorithm

Let θ be a vector of the model's unknown parameters[22]. The EM algorithm will iterate between expectation and maximization steps until some convergence criteria is met:

- Expectation step: Suppose we have complete data, then we multiply the log-likelihood of the complete data by the likelihood evaluated using last step parameter estimates (θ^{k-1}). The "expectation" is in the sense that we integrate out the effect of unobservables, to obtain the following

$$H(\theta; \widetilde{y}_T, \theta^{k-1}) = \int_{\widetilde{(A,s)}_T} \log[f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T|\theta)] \cdot f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T|\theta^{k-1}) d\widetilde{(A,s)}_T$$

- Maximization step: The objective function H is maximized with respect to parameters of the model, resulting in the step-k estimates θ^k.

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

One favorable property of EM algorithm is that each iteration results in a higher value of the likelihood function. 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[23].

[22] Here, if we ignore the initial condition, we have 6 parameters concerning probabilities, 4 parameters for mean, and 2 parameters for variance. And thus we have 12 parameters in total.

[23] The power of EM method is that we take "log" inside the integration, which will greatly simplify computation, as we are soon to see.

The joint density of $\widetilde{y}_T$, $\widetilde{A}_T$ and $\widetilde{s}_T$ can be written as,

$$\begin{aligned} f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T; \theta) &= f(\widetilde{y}_T | \widetilde{A}_T, \widetilde{s}_T; \theta) \cdot \Pr(\widetilde{A}_T, \widetilde{s}_T; \theta) \\ &= \prod_{t=1}^{T} f(y_t | A_t, s_t; \theta) \prod_{t=1}^{T} \Pr(A_t, s_t | A_{t-1}, s_{t-1}; \theta) \\ &= \prod_{t=1}^{T} f(y_t | Z_t; \theta) \prod_{t=1}^{T} \Pr(Z_t | Z_{t-1}; \theta) \end{aligned}$$

where $Z_t \equiv (A_t, s_t)$ is the joint state.

And then the log likelihood is given by sum of a set of conditional probabilities,

$$\log f(\widetilde{y}_T, \widetilde{A}_T, \widetilde{s}_T; \theta) = \sum_{t=1}^{T} \log f(y_t | Z_t; \theta) + \sum_{t=1}^{T} \log \Pr(Z_t | Z_{t-1}; \theta)$$

It is useful to notice that the maximization of H is equivalent to maximize Q, with

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

where $\Pr(\widetilde{Z}_T | \widetilde{y}_T; \theta^{k-1})$ can be obtained using a filtering-smoothing procedure as described in Kim and Nelson (1999).[24]

[24] Here we can treat the objective function Q as the conditional expectation of complete data likelihood function, with the expectation taken over unobservables conditional on all information available.

8.1.1 Inside the likelihood function

Now we compute $\int_{\widetilde{Z}_T} \log[f(\widetilde{y}_T|\widetilde{Z}_T;\theta)]\cdot\Pr(\widetilde{Z}_T|\widetilde{y}_T;\theta^{k-1})d\widetilde{Z}_T$ and $\int_{\widetilde{Z}_T} \log[\Pr(\widetilde{Z}_T|\theta)]\cdot\Pr(\widetilde{Z}_T|\widetilde{y}_T;\theta^{k-1})d\widetilde{Z}_T$ to facilitate our deduction of first order conditions.

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

where $\log[f(y_t|Z_t;\theta_1] = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2(A_t)) - \frac{1}{2}\frac{(y_t-\mu(Z_t))^2}{\sigma^2(A_t)}$.

Similarly, we have

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

8.1.2 Closed-form solution

First order condition concerning structure transition yields

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

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

The first order condition with respect to conditional transition matrix for regimes gives us

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

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

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

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

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

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

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

In order to get the above parameter solutions, we still need 3 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 $\Pr(A_t, s_t, s_{t-1}|\widetilde{y}_T; \theta^{k-1})$. The following sections describe the expectation step, featured by a filtering and smooth procedure.

8.2 Filtering

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 joint state $Z = (A, s)$.[25]

Step 1. Given $\Pr(Z_{t-1} = i|I_{t-1})$, $i = 1, 2$, at the beginning of time t iteration, the weighting terms $\Pr(Z_t = j|I_{t-1})$, $j = 1, 2$, are calculated as

$$\begin{aligned}\Pr(Z_t|I_{t-1}) &= \sum_{Z_{t-1}} \Pr(Z_t, Z_{t-1}|I_{t-1}) \\ &= \sum_{Z_{t-1}} \Pr(Z_t|Z_{t-1}) \cdot \Pr(Z_{t-1}|I_{t-1})\end{aligned}$$

Step 2. Once y_t is observed at the end of time t, we can update the probability term in the following way:

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

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

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

As a by-product, we also obtain the likelihood function as $Likelihood = \sum_{Z_t} f(y_t|Z_t) \cdot \Pr(Z_t|I_{t-1})$, which is recursively computed during the above filtering steps, treating θ^{k-1} as an unknown parameter θ.

[25] In the simple case we considered, the state number is four.

8.3 Smoothing

We can make inference on Z_t, based on the entire available information in the sample, to obtain the smoothed probability $\Pr(Z_t|I_T)$ and $\Pr(Z_t, Z_{t-1}|I_T)$, $t = 1, 2, ..., T$.

Consider the following joint probability,

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

Notice that

$$
\begin{aligned}
\Pr(Z_t = j|Z_{t+1} = k, I_T) &= \Pr(Z_t = j|Z_{t+1} = k, \widetilde{I}_{t+1,T}, I_t) \\
&= \frac{\Pr(Z_t = j, \widetilde{I}_{t+1,T}|Z_{t+1} = k, I_t)}{\Pr(\widetilde{I}_{t+1,T}|Z_{t+1} = k, I_t)} \\
&= \frac{\Pr(Z_t = j|Z_{t+1} = k, I_t) \cdot \Pr(\widetilde{I}_{t+1,T}|Z_t = j, Z_{t+1} = k, I_t)}{\Pr(\widetilde{I}_{t+1,T}|Z_{t+1} = k, I_t)} \\
&= \Pr(Z_t = j|Z_{t+1} = k, I_t)
\end{aligned}
$$

where $\widetilde{I}_{t+1,T} = \{y_{t+1}, ..., y_T\}$, and the last line of the above equation comes from the fact that y_{t+1} and its future depends on Z_{t+1} only; once Z_{t+1} is known, Z_t and I_t contain no further information for $\widetilde{I}_{t+1,T}$.

The other smoothed probability is calculated as follows,

$$
\Pr(Z_t = j|I_T) = \sum_{k=1}^{4} \Pr(Z_t = j, Z_{t+1} = k|I_T)
$$

Thus given $\Pr(Z_T|I_T)$ at the last iteration of the filtering process, the above can be iterated for $t = T-1, T-2, ..., 1$ to get the two types of smoothed probabilities we need.

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 the terms s_t and s_{t-1}.