Markov Switching in Disaggregate Unemployment Rates*

Marcelle Chauvet, Chinhui Juhn and Simon Potter †

May 2001

Abstract

We develop a dynamic factor model with Markov switching to examine secular and business cycle fluctuations in U.S. unemployment rates. We extract the common dynamics among unemployment rates disaggregated for seven age groups. The framework allows analysis of the contribution of demographic factors to secular changes in unemployment rates. In addition, it allows examination of the separate contribution of changes due to asymmetric business cycle fluctuations. We find strong evidence in favor of the common factor and of the switching between high and low unemployment rate regimes. We also find that demographic adjustments can account for a great deal of the secular change in the unemployment rate, particularly the abrupt increase in the 1970s and 1980s and the subsequent decrease.

Keywords: Markov Switching, Unemployment, Common Factor, Asymmetries, Business Cycle, Baby Boom, Bayesian Methods.

*We would like to thank the editors and two referees for numerous useful comments and suggestions.

†The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System. Address first author: Department of Economics University of California, Riverside, CA 92521, chauvet@mail.ucr.edu. Address second author Department of Economics, Univesity of Houston, Houston, TX 77204, cjuhn@uh.edu. Address third author: Domestic Research, Federal Reserve Bank of New York, New York, NY 10045, simon.potter@ny.frb.org. Note: please send proofs to first author.

1 Introduction

The U.S. economic performance during the 1990s expansion has in some aspects been unprecedented. Not only is this the longest expansion in U.S. history, inflation and unemployment have been unusually low this far into the business cycle, even in the presence of the current slowdown. In particular, the unemployment rate is at its lowest levels since before the 1970 recession.

This paper examines the secular and business cycle movements in unemployment over the past 50 years. In particular, we focus on a variety of features that can be observed in Figure 1, which shows the U.S. total civilian unemployment rate along with shaded bands indicating the NBER-dated recessions. One observes several interesting patterns in the unemployment fluctuations. First, there is the usual cyclical movement related to the phases of the business cycle. Unemployment decreases slowly during expansions, reaching its lowest level around the beginning of recessions. During recessions unemployment rises sharply, reaching a maximum a couple of months after the economic trough. Second, unemployment exhibits a different secular pattern before and after 1969. In particular, unemployment fluctuations display a long upswing through the 1970s and early 1980s, with a high of over 10% at the end of the 1982 recession.

We develop a common factor model with Markov switching to analyze these secular and business cycle fluctuations in the U.S. unemployment rate.[1] In particular, we consider extracting the common dynamics amongst a collection of disaggregate unemployment rates which, given appropriate labor force weights, sum to the total unemployment rate. The framework proposed allows analysis of the contribution of demographic factors to secular changes in unemployment rates. In particular, we focus on the impact of the entry into the labor force and the subsequent aging of the baby-boom generation.[2] In addition, it allows examination of the separate contribution of changes in

[1] Linear dynamic factor models have been widely applied in economics. A classical exposition is Geweke (1977) and one of the most popular applications is Stock and Watson's (1989), who build coincident and leading indicators of the U.S. economy. Diebold and Rudebusch (1996) propose adding the Markov switching model of Hamilton's (1989) to this framework in order to capture asymmetries over the business cycle.

[2] Another potential application of our methods is to explain the skill-biased technological change in the 1970s and 1980s, and its differential effects on education groups (see Juhn, Murphy and Topel 1991). Alternatively, one might consider unemployment grouped by sex and race to examine secular changes in the labor market behavior of these groups, particularly the increasing participation of women in the labor force.

unemployment rates due to asymmetric business cycle fluctuations.

The behavior of unemployment over the business cycle has attracted a great deal of attention from economists using nonlinear time series models.[3] With a few exceptions [4] this large literature has focused on some measure of aggregate unemployment. The focus of this paper is to extend the nonlinear analysis of unemployment rates to the disaggregate level. There are two main motivations for this. The first is that some of the multiple regimes that have been found in the unemployment rate may be explained by demographic factors, particularly the baby boom. The second motivation is that nonlinear models ask a great deal of a single time series in terms of identifying Markov switching regimes. For example, in Neftci's (1984) original analysis, the evidence for asymmetry in the total unemployment was marginal. If the regimes are actually present, one would expect to find a common switch in disaggregate series. This would increase both the accuracy of the switch estimates as well as the evidence for their presence.

With respect to demographic factors, the significant increase in the number of births in the 1950s and early 1960s is often acknowledged as changing several aspects of the economy. As far as the overall unemployment rate is concerned, because younger labor market participants tend to have higher unemployment rates than older ones, their entry into the labor market produced an increase in the unemployment rate in the 1970s and 1980s, while their subsequent aging induced a decrease in the 1990s.[5] Our approach is to model a latent unemployment rate that captures common labor market conditions across groups, and then re-aggregate this latent unemployment rate using the time-varying labor force weights of the different age groups. This allows us to decompose changes in unemployment between baby-boom type effects and changes in the overall functioning of the labor market.

We find strong statistical evidence in favor of the common factor structure and of the switching between high and low unemployment rate regimes. In particular, the latent factor exhibits the stylized business cycle asymme-

[3] See for example, Neftci (1984), Rothman (1993), Boldin (1994), Franses (1995), Montgomery, Zarnowitz, Tsay, and Tiao (1998), Abbring, Berg and Ours (1999), Vredin and Warne, (2000), or Skalin and Teräsvirta (2001).

[4] Rothman (1993) and Abbring, Berg, Ours (1999).

[5] Some of the important studies in this subject include Perry (1970) and Gordon (1982). More recently, Shimer (1998) uses an accounting style analysis, which suggests that the aging baby boomer is an important cause of the reduction in unemployment in the 1990s. These results are corroborated in Katz and Krueger (1999)

tries found in unemployment. The factor displays a fast and steep growth and a slow and long decline, associated with the phases of the business cycle. In addition, the low unemployment state is less persistent and more volatile compared to the high unemployment phase. We also find that demographic adjustments can account for a great deal of secular changes in the U.S. unemployment rates, particularly the abrupt increase in unemployment in the 1970s and 1980s and the subsequent decrease in the last 18 years. The baby boom effect induces a steep increase in the unemployment rate at the beginning of recessions in the 1970s and 1980s as the labor force weights of young labor market participants were increasing at the time that these switches to the high unemployment regime occurred. On the other hand, the impact of age on the unemployment rate was substantially smaller in the 1990 recession, which is related to the subsequent aging of the baby-boom generation.

The model is estimated and analyzed using Bayesian methods. In particular, the Gibbs sampling is used to simulate and estimate the model while the Savage-Dickey Generalized Density Ratio is used to calculate Bayes factors, which allow evaluation of the sample evidence in favor of the Markov switches.

The organization of the paper is as follows. Section 2 develops the statistical model for disaggregate unemployment rates. Section 3 describes some Bayesian techniques to estimate and test the model. Section 4 discusses the features of the data on unemployment rates disaggregate by age. Section 5 discusses the prior used and the results. Section 6 offers some concluding remarks and directions for future research. The appendix summarizes the complete set of priors and conditional distributions used in the analysis.

2 Statistical Model and Methods

2.1 Common Factor Model

Let U_t be the $K \times 1$ vector of unemployment rates for different age groups used to estimate the common factor, C_t. The statistical model is:

$$\mathsf{U}_t = \boldsymbol{\lambda} C_t + \mathsf{V}_t, \tag{1}$$

where $\boldsymbol{\lambda}$ is the $K \times 1$ vector of factor loadings, which measures the sensitivity of group k's unemployment rate to the $1x1$ underlying factor C_t, and the $K \times 1$

random vector V_t represents a possibly autocorrelated measurement error. The common factor is given by the Markov switching model:

$$C_t = \begin{cases} \alpha_1 + \phi_{1p}(L)C_{t-1} + \sigma_1\varepsilon_t & \text{if } s_t = 1 \\ \\ \alpha_2 + \phi_{2p}(L)C_{t-1} + \sigma_2\varepsilon_t & \text{if } s_t = 2 \end{cases}, \tag{2}$$

with $P[s_{t+1} = 1|s_t = 1] = \rho_{11}$ and $P[s_{t+1} = 2|s_t = 2] = \rho_{22}$. The regime specific means are defined as:

$$\eta(1) = \frac{\alpha_1}{1 - \phi_{11} - \cdots - \phi_{1p}}, \eta(2) = \frac{\alpha_2}{1 - \phi_{21} - \cdots - \phi_{2p}}.$$

We identify state 1 as the high unemployment phase and state 2 as the low unemployment phase, that is, $\eta(1) > \eta(2)$. Further, we consider the observed asymmetries in unemployment of sharp increases during recessions followed by slow decreases during expansions using restrictions on the autoregressive coefficients in the two regimes. These restrictions are easier to describe in the case of a first order process. In order for unemployment to increase quickly when it switches from state 2 to 1, ϕ_{11} needs to be relatively small so that α_1 is close in size to $\eta(1)$. Alternatively, when there is a switch from state 1 to state 2, unemployment declines slowly if ϕ_{21} is close to 1 and, therefore, α_2 is very different from $\eta(2)$.

The measurement error vector V_t has an autoregressive structure:

$$\mathsf{V}_t = \Theta_1\mathsf{V}_{t-1} + \cdots + \Theta_q\mathsf{V}_{t-q} + \mathsf{M}_t, \tag{3}$$

where the innovations to the common factor, $\varepsilon_t \sim IIDN(0,1)$, and the measurement error, $\mathsf{M}_t \sim IIDN(\mathbf{0}, \Sigma_K)$, are independent of each other at all leads and lags, and Σ_K is diagonal. In addition, the autoregressive matrices are diagonal:

$$\Theta_i = \begin{bmatrix} \theta_{1i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{1K} \end{bmatrix}.$$

2.2 Latent Unemployment

Given our collection of K different age groups, we construct demographic weights for each category by:

$$\omega_{kt} = \frac{L_{kt}}{\sum_{k=1}^{K} L_{kt}}, \tag{4}$$

where L_{kt} is the total civilian labor force for age group k at time t. The overall unemployment rate for all groups UR_t is given by the estimate of the total number of workers unemployed, TU_t, divided by the total civilian labor force, $\sum_{k=1}^{K} L_{kt}$:

$$\begin{aligned} UR_t &= \frac{TU_t}{\sum_{k=1}^{K} L_{kt}} = \frac{\sum_{k=1}^{K} L_{kt} U_{kt}}{\sum_{k=1}^{K} L_{kt}}, \text{ or} \\ UR_t &= \sum_{k=1}^{K} \omega_{kt} U_{kt}, \end{aligned}$$

where U_{kt} is the unemployment rate for age group k. The latent aggregate unemployment rate implied by the common factor model is:

$$\mathcal{U}_t = C_t \sum_{k=1}^{K} \omega_{kt} \lambda_k, \tag{5}$$

The average or "natural rate" of unemployment fluctuates between the "high-mismatch" average,

$$\mathcal{U}_t(1) = \eta(1) \sum_{k=1}^{K} \omega_{kt} \lambda_k, \tag{6}$$

and the "low-mismatch" average,

$$\mathcal{U}_t(2) = \eta(2) \sum_{k=1}^{K} \omega_{kt} \lambda_k. \tag{7}$$

That is, $\mathcal{U}_t(1)$ and $\mathcal{U}_t(2)$ correspond to demographically adjusted high and low bounds for average unemployment. In the empirical analysis, these bounds allow analysis of periods in which average unemployment exceeds the high-mismatch and low mismatch averages in the sample.

Notice that, in contrast with an analysis of the aggregate unemployment rate, our estimate of $\{C_t, s_t\}$ is not influenced by variations over time in the demographic weights. In the case where the labor force weights are constant over time, we have:

$$\mathcal{U}_t - \mathcal{U}_\tau = (C_t - C_\tau)\sum_{k=1}^{K}\omega_k\lambda_k, \text{ for } \tau < t, \tag{8}$$

or, alternatively, the factor loadings are equal across groups at unity:

$$\mathcal{U}_t - \mathcal{U}_\tau = (C_t - C_\tau). \tag{9}$$

In both cases, demographic changes have no effect on the change in latent aggregate unemployment and there would be little advantage from examining disaggregate unemployment rates. This should be compared to the more general case in which the factor loadings vary across k or the demographic weights change over time. In these cases, changes in aggregate latent unemployment can be split into 2 different contributions – the ones arising from changes in the common factor and the ones from demographic changes:

$$\begin{aligned}\mathcal{U}_t - \mathcal{U}_\tau = & \\ & (C_t - C_\tau)\sum_{k=1}^{K}\omega_{kt}\lambda_k \text{ Factor Effect} \\ & + C_\tau \sum_{k=1}^{K}(\omega_{kt} - \omega_{k_\tau})\lambda_k \text{ Demographic Effect}\end{aligned} \tag{10}$$

2.3 State Space Form

The model can be written in state space form where we assume that $p = q+1$ for simplicity of notation. First we define the following:

1. Let $\mathsf{U}_t^* = (\mathsf{I}_k - \Theta(L))\mathsf{U}_t$.
2. Let $\mathsf{C}_t^* = [C_t, \ldots, C_{t-p+1}]^{\prime}$.

3. Define the $K \times (q+1)$ matrix H by:

$$\mathsf{H} = \begin{bmatrix} \lambda_1 & -\lambda_1\theta_{11} & \cdots & -\lambda_1\theta_{q1} \\ \lambda_2 & -\lambda_2\theta_{12} & \cdots & -\lambda_2\theta_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_K & -\lambda_K\theta_{1K} & \cdots & -\lambda_K\theta_{qK} \end{bmatrix}.$$

4. Define the $p \times p$ matrices A_i by:

$$\mathsf{A}_i = \begin{bmatrix} \phi_{i1} & \phi_{i2} & \cdots & \phi_{ip} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ 0 & \ddots & 1 & 0 \end{bmatrix}.$$

The state space form has measurement equation:

$$\mathsf{U}_t^* = \mathsf{H}\mathsf{C}_t^* + \mathsf{M}_t \tag{11}$$

and transition equation:

$$\mathsf{C}_t^* = \begin{cases} \mathsf{a}_1 + \mathsf{A}_1\mathsf{C}_{t-1}^* + \mathsf{W}_1\varepsilon_t & \text{if } s_t = 1 \\ \\ \mathsf{a}_2 + \mathsf{A}_2\mathsf{C}_{t-1}^* + \mathsf{W}_2\varepsilon_t & \text{if } s_t = 2, \end{cases} \tag{12}$$

where $\mathsf{W}_i = [\sigma_i, 0, \cdots, 0]'$ and $\mathsf{a}_i = [\alpha_i, 0, \ldots, 0]'$ are $(p \times 1)$ vectors.

Below we will also sometimes summarize the conditional mean coefficients in each regime by the $p \times 1$ vector $\boldsymbol{\phi}_i = (\phi_{1i}, \ldots, \phi_{pi})$, or by the $(p+1) \times 1$ vector $\boldsymbol{\beta}_i = (\alpha_i, \phi_{1i}, \ldots, \phi_{pi})$. Let φ represent all the parameters of the common factor model with Markov switching and χ represent the (smaller by $\boldsymbol{\beta}_2, \sigma_2, \rho_{11}, \rho_{22}$) set of parameters of the common factor model without Markov switching.

As is true of all single factor models there is an identification issue between the factor loadings and the scaling of the innovation to the common factor (see Chauvet 1998). We normalize one of the elements of the factor loading vector to 1. Also, unlike Stock and Watson (1989), we do not demean and standardize the observable variables before the analysis. Thus, not only will $\boldsymbol{\lambda}$ be estimated from the joint dynamics of the observed time series it will also depend on information on the relative means and variances of the unemployment rates. In order to capture movements between low and high unemployment regimes, it is crucial to keep the mean information in the series.

3 Bayesian Methods

If the sequence $\{s_t\}$ were known, estimation by classical or Bayesian methods would be standard. Both methods use the Kalman filter to construct the likelihood function.

3.1 Kalman Filter

The Kalman filter iterations are given by:

1. **Prediction Step**: The conditional mean of the factor is,

$$\mathsf{C}^*_{t+1|t} = \begin{cases} \mathsf{a}_1 + \mathsf{A}_1\mathsf{C}^*_{t|t} & \text{if } s_t = 1 \\ \\ \mathsf{a}_2 + \mathsf{A}_2\mathsf{C}^*_{t|t} & \text{if } s_t = 2 \end{cases}.$$

The conditional variance of the factor is,

$$\mathsf{P}_{t+1|t} = \begin{cases} \mathsf{A}_1\mathsf{P}_{t|t}\mathsf{A}_1^{\prime} + \mathsf{W}_1\mathsf{W}_1^{\prime} & \text{if } s_t = 1 \\ \\ \mathsf{A}_2\mathsf{P}_{t|t}\mathsf{A}_2^{\prime} + \mathsf{W}_2\mathsf{W}_2^{\prime} & \text{if } s_t = 2 \end{cases}.$$

Using the conditional mean of the factor and the measurement equation (11), we obtain the conditional forecast error:

$$\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t} = \mathsf{H}(\mathsf{C}^*_{t+1} - \mathsf{C}^*_{t+1|t}) + \mathsf{M}_{t+1},$$

and its conditional variance:

$$E\left[(\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t})(\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t})^{\prime}\right] = \mathsf{H}\mathsf{P}_{t+1|t}\mathsf{H}^{\prime} + \Sigma_K.$$

2. **Updating Step**: First, the Kalman gain matrix is constructed:

$$\mathsf{G}_{t+1} = \mathsf{P}_{t+1|t}\mathsf{H}^{\prime}\left\{E\left[(\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t})(\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t})^{\prime}\right]\right\}^{-1}.$$

Then, as new information about the factor is obtained after observing U^*_{t+1}, the Kalman gain is used to include it in the conditional mean of the factor:

$$\mathsf{C}^*_{t+1|t+1} = \mathsf{C}^*_{t+1|t} + \mathsf{G}_{t+1}\left(\mathsf{U}^*_{t+1} - \widehat{\mathsf{U}}^*_{t+1|t}\right),$$

and to update the conditional variance:

$$\mathsf{P}_{t+1|t+1} = \left(\mathsf{I}_p - \mathsf{G}_{t+1}\mathsf{H}\right)\mathsf{P}_{t+1|t}.$$

3.2 Posterior Draws of Markov States and Common Factor

The main estimation problem in the model proposed is that the switches in the Markov chain are not observable. As is well-known, this causes a computational problem for standard maximum likelihood approaches, since one has to keep track of the 2^T possible values of the Markov sequence in the sample. Kim (1994) suggests an maximum likelihood approach using an approximation that truncates the exponential increasing number of terms at each iteration in the Kalman filter. On the other hand, Bayesian simulation techniques can be used to obtain the exact likelihood function of the common factor Markov switching model, as proposed by Albert and Chib (1993) and Shephard (1994).

In this paper we use Bayesian methods to extract the sample evidence about the Markov sequence. In our model it is not possible to work out analytically the properties of the posterior even if the common factor and Markov states were directly observed. In these cases, Bayesians have increasingly turned to simulation methods. These methods are based on the intuition that, given a large enough random sample from a distribution, it is possible to figure out the properties of that distribution (e.g. its mean, median, variance, etc.). Geweke (1999) and Chib (2001a, 2001b) provide recent surveys of the methods used for developing such posterior simulators.[6] The advantage of posterior simulators in our case is that they can also be developed to generate draws of the unobservables, which greatly simplifies the estimation problem.

[6] A textbook-type explanation of the method can also be found in Kim and Nelson (1999).

Posterior simulators for many models can be developed by successively drawing from a sequence of conditional posterior distributions. To motivate such methods, let X and Y be random variables and suppose that we are interested in the features of their joint distribution, $p(X, Y)$. Assume that $p(X, Y)$ is difficult to simulate, but that we can easily obtain draws from the marginal distributions $p(X|Y)$ and $p(Y|X)$. Consider the strategy where one selects an initial value, $Y^{(0)}$, and then successively draws $X^{(j)}$ from $p(X|Y^{(j-1)})$, and $Y^{(j)}$ from $p(Y|X^{(j)})$. The resulting sequence, $X^{(j)}, Y^{(j)}$ for $j = 1, \ldots, J$ will, under weak conditions, converge to a sample from $p(X, Y)$ as J increases. In practice, this means we can discard $\underline{J}$ initial draws to mitigate startup effects. We can then treat $X^{(j)}, Y^{(j)}$ for $j = \underline{J} + 1, .., J$ as an approximate sample from $p(X, Y)$, which can be used to estimate features of interest. This is a simple example of a Gibbs sampler, which belongs to a more general class of techniques known as Markov Chain Monte Carlo methods.

Many commonly-used econometric models with latent variables can be easily estimated using Gibbs sampling algorithms. In our particular model, the Gibbs sampler generates random draws of $\{s_t\}$, which allows analysis as if the sequence were known. However, in order to obtain inferences we also need to obtain a random draw of the common factor.

3.2.1 Common Factor

The recursion to generate the random draw of the common factor is as follows (see Fruhwirth-Schnatter, 1994, Carter and Kohn, 1994, or Shephard, 1994):

1. The last iteration of the Kalman filter gives:

$$\mathsf{C}_T^* \sim N(\mathsf{C}_{T|T}^*, \mathsf{P}_{T|T}).$$

   Thus, using standard methods a realization $\widetilde{\mathsf{C}}_T^*$ can be drawn from this multivariate normal. Then, the draw of the most recent value of the common factor is given by:

$$\widetilde{C}_T = \mathsf{e}\widetilde{\mathsf{C}}_T^*.$$

   where $\mathsf{e} = [1, 0, \ldots, 0]$ is a $p \times 1$ selection vector. In practice, one only needs to draw from the univariate normal with mean given by the first element of $\mathsf{C}_{T|T}^*$ and variance by the first diagonal element of $\mathsf{P}_{T|T}$.

2. Given a draw at $t+1$ based on draws from $t+2$ to T the information from the Kalman filter iterations are incorporated as if the filter were running backwards combining prior information from its initial forward run with the 'sample' information generated by the random draw:

$$f_t = \widetilde{C}_{t+1} - \begin{cases} \alpha_1 + \phi_{1p}(L)C_{t|t} & \text{if } s_t = 1 \\ \alpha_2 + \phi_{2p}(L)C_{t|t} & \text{if } s_t = 2 \end{cases},$$

$$p_t = \begin{cases} \boldsymbol{\phi}_1^{\prime} \mathsf{P}_{t|t} \boldsymbol{\phi}_1 + \sigma_1^2 & \text{if } s_t = 1 \\ \boldsymbol{\phi}_2^{\prime} \mathsf{P}_{t|t} \boldsymbol{\phi}_2 + \sigma_2^2 & \text{if } s_t = 2 \end{cases},$$

$$\mathsf{g}_t = \begin{cases} \mathsf{P}_{t|t} \boldsymbol{\phi}_1 / p_t & \text{if } s_t = 1 \\ \mathsf{P}_{t|t} \boldsymbol{\phi}_2 / p_t & \text{if } s_t = 2 \end{cases},$$

$$\mathsf{C}_{t|T}^* = \mathsf{C}_{t|t}^* + \mathsf{g}_t f_t,$$

$$\mathsf{P}_{t|T} = \begin{cases} \left(\mathsf{I}_p - \mathsf{g}_t \boldsymbol{\phi}_1^{\prime}\right) \mathsf{P}_{t|t} & \text{if } s_t = 1 \\ \left(\mathsf{I}_p - \mathsf{g}_t \boldsymbol{\phi}_2^{\prime}\right) \mathsf{P}_{t|t} & \text{if } s_t = 2 \end{cases}.$$

Thus, after observing the whole sample we obtain $\mathsf{C}_t^* \sim N(\mathsf{C}_{t|T}^*, \mathsf{P}_{t|T})$, and standard methods can be used to obtain a random draw.

3. This iteration stops with $\mathsf{C}_p^* \sim N(\mathsf{C}_{p|T}^*, \mathsf{P}_{p|T})$, which is used to simultaneously draw the first p observations of the common factor.

3.2.2 Markov States

1. Given this realization of the common factor, the methods described in Chib (1996) can be directly applied to investigate the pattern of Markov states contained in the draw of the dynamic factor. Similarly to the analysis of the common factor, the recursion starts by filtering the "observations" on the common factor to construct a sequence of prior and posterior distributions for the Markov state.

In our case, we assume that the chain starts in the low mismatch state 2. Thus, we have $S_1 = \cdots = S_p = 2$. The prior distribution for the state at time $p+1$ is then:

$$P[S_{p+1} = 2 | S_1 = \cdots = S_p = 2] = \widehat{b}_{p+1} = \rho_{22}.$$

The relevant sample information is contained in the likelihood for $\widetilde{C}_{p+1}$, which takes on the value:

$$\ell_1(\widetilde{C}^{p+1}) = (2\pi\sigma_1^2)^{-0.5} \exp\left[-0.5(\widetilde{C}_{p+1} - \alpha_1 - \phi_{1p}(L)\widetilde{C}_p)^2/\sigma_1^2\right]$$

if $S_{p+1} = 1$, and:

$$\ell_2(\widetilde{C}^{p+1}) = (2\pi\sigma_2^2)^{-0.5} \exp\left[-0.5(\widetilde{C}_{p+1} - \alpha_2 - \phi_{2p}(L)\widetilde{C}_p)^2/\sigma_2^2\right]$$

if $S_{p+1} = 2$. Thus, the posterior distribution $P[S_{p+1} = 2 | S_1 = \cdots = S_p = 2, \widetilde{C}^{p+1}]$ is given by

$$\widehat{b}_{p+1} = \frac{\ell_2(\widetilde{C}^{p+1})\rho_{22}}{\ell_2(\widetilde{C}^{p+1})\rho_{22} + \ell_1(\widetilde{C}^{p+1})(1-\rho_{22})}.$$

This posterior is used to form a prior for $S_{p+2} = 2$ from:

$$\widehat{b}_{p+2} = b_{p+1}\rho_{22} + (1 - b_{p+1})(1 - \rho_{11}).$$

This process continues through the end of the sample and we obtain:

$$b_T = P[S_T = 2 | S_1 = \cdots = S_p = 2, \widetilde{C}^T].$$

Next, the posterior distribution over the last sample value of the Markov state is used to generate a draw of S_T. This draw is easy to obtain from the generation of a uniform random variable. If the draw is less than or equal to the posterior probability that $S_T = 2$, then we have $\widetilde{S}_T = 2$, otherwise $\widetilde{S}_T = 1$.

Similarly to the draw of the common factor, a sequence of draws of $\{S_t\}$ is also generated going backwards through the sample. From the Markov property and the exogeneity of the Markov chain, none of the future realizations of the observed time series $\{\widetilde{C}_s : s > t\}$ conditional on observing the value of tomorrow's state, $s_{t+1,}$ are relevant for the estimate of today's state.

Using this restriction and ignoring the dependence on estimated parameters and the restriction on the initial states, we have:

$$
\begin{aligned}
& P[S_{T-1} = 2 | S_T = 2, \widetilde{C}^{T-1}] \\
&= \frac{P[S_{T-1} = 2, S_T = 2 | \widetilde{C}^{T-1}]}{P[S_T = 2 | \widetilde{C}^{T-1}]} \\
&= \frac{P[S_{T-1} = 2 | \widetilde{C}^{T-1}] P[S_T = 2 | S_{T-1} = 2]}{\widehat{b}_T} \\
&= \frac{b_{T-1} \rho_{22}}{\widehat{b}_T},
\end{aligned}
$$

and also:

$$
\begin{aligned}
& P[S_{T-1} = 2, S_T = 2 | \widetilde{C}^T] \\
&= P[S_{T-1} = 2 | S_T = 2, \widetilde{C}^T] P[S_T = 2 | \widetilde{C}^T] \\
&= P[S_{T-1} = 2 | S_T = 2, \widetilde{C}^{T-1}] b_T \\
&= \rho_{22} \frac{b_{T-1}}{\widehat{b}_T} b_T .
\end{aligned}
$$

This last expression can be used to generate a smoothed probability $\widetilde{b}_t$ by averaging out over the values of S_T :

$$
\widetilde{b}_{T-1} = b_{T-1} \left(\rho_{22} \frac{\widetilde{b}_T}{\widehat{b}_T} + (1 - \rho_{11}) \frac{1 - \widetilde{b}_T}{1 - \widehat{b}_T} \right).
$$

The first expression gives a direct method for drawing $\widetilde{S}_{T-1}$ using a random uniform number. This iteration is repeated until the draw of $\widetilde{S}_{P+1}$ using $P[S_p = 2, S_{p+1} = 2 | \widetilde{C}^T] = 1$ is obtained.

3.3 Estimation by Gibbs Sampler

We initialize the Gibbs sampler running the Kalman filter on the observed data with the following parameters: a Markov sequence given by the NBER business cycle dates, factor loadings proportional to the sample means of the unemployment rates, measurement error equal to 1/4 of the observed variances of the unemployment rates with autoregressive coefficients equal to 1/3, and the prior means used for the common factor dynamic model and loadings. The results of the Kalman filter are then used to draw a initial sequence of realizations for the common factor $\{\widetilde{C}_t\}$. Given this sequence and the sequence of Markov states, it is relatively simple to update the draws of the remaining parameters using the Gibbs sampler with appropriate choices of prior distributions.

A summary of our choices is as follows: for the parameters of C_t we use a restricted normal inverted gamma prior; for the factor loadings $\boldsymbol{\lambda}$ we use independent normal priors; for the measurement error variances, $\{\sigma_{kk}\}$ we use independent normal gamma priors; for the autoregressive coefficients in the measurement error process we use restricted independent normal priors; for the transition probabilities we use a Beta distribution. Section 5.1 contains information of the exact prior distributions used and the appendix contains more detailed information on the priors and the implied conditional posteriors. Here we focus on some of the restrictions imposed on the parameters.

For both the parameters of C_t and the measurement error processes we impose a stationarity condition. In the case of the measurement error process the roots of the individual lag polynomials $1 - \theta_{kq}(L)$ all lie outside the unit circle. For the common factor Markov switching model we use the sufficient condition that the roots of each lag polynomial $1 - \phi_{ip}(L)$ lie outside the unit circle. For both sets of restrictions we use simple rejection sampling.

In addition to the stationarity restriction, we also impose the restriction that $\eta(1) > \eta(2) > 0$. Since the regime specific means are nonlinear functions of the underlying autoregressive parameters and the intercept, the restriction is imposed sequentially as follows. First, we draw $\boldsymbol{\beta}_2$ and check whether the stationarity and non-negativity conditions are satisfied. If they are not we make a new draw. This process continues until a satisfactory draw is obtained. Next, we draw $\boldsymbol{\beta}_1$and check whether both the stationarity and mean inequality conditions hold. Again, we reject this draw if it does not satisfy the conditions, continuing until a statisfactory draw is obtained.

3.4 Evidence for Markov Switching

We assess the observed sample evidence in favor of a Markov switching in the common factor model by comparing the average (often called the marginal) likelihood of the observed time series with and without switching. The ratios of these two average likelihoods is the Bayes factor and it provides a direct 'test' of the usefulness of the additional complexity of the Markov switching model. On the other hand, classical testing of Markov switching models is a somewhat unresolved area due to various non-standard aspects of the model. We do not attempt to solve these issues here. Instead, we provide some classical-type information by calculating F-statistics at each iteration of the Gibbs sampler. Note that under the null hypothesis of no Markov switching effect and uninformative priors we would expect these statistics to be draws from a F-distribution with appropriate degrees of freedom. Since we are not using non-informative priors the exact sampling distribution of this sequence of F-statistics is unknown.[7] However, we focus on the minimum value of this statistic across iterations of the Gibbs sampler.

Although the calculation of marginal likelihoods involves multiple integration, it can be simplified using the following tricks. The Bayes factor is the marginal likelihood of the 'no-switching model' divided by the marginal likelihood of the 'switching model':

$$B_{\text{No Switching vs.Switching}} = \frac{\int l(\mathsf{U}|\chi, \{C_t\})b(\chi,)d\{C_t\}d\chi}{\int l(\mathsf{U}|\varphi, \{s_t\}, \{C_t\})b(\varphi, s_1 = 2)d\{C_t\}d\{s_t\}d\varphi},$$

if this ratio is larger than 1 the sample favors the simple model with no switching. Notice that a ratio of 0.05 is not equivalent to a classical critical value. Instead, it implies that the Markov switching model is 20 times more likely than the simple model, given the observed data on unemployment rates.

Using the basic likelihood identity (see Chib 1995) we have:

$$\int l(\mathsf{U}|\varphi, \{s_t\}, \{C_t\})b(\varphi, s^p = 2)d\{C_t\}d\{s_t\}d\varphi = \frac{l(\mathsf{U}|\varphi, \{s_t\}, \{C_t\})b(\varphi, s^p = 2)}{p(\varphi|\mathsf{U})},$$

for all points in the parameter space. In particular, consider the transformation of the parameter space for the Markov swtiching common factor model from $(\boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_2, \sigma_2, \rho_{11,}\rho_{22})$ to $(\boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1, \sigma_2/\sigma_1, \tau)$. If we evaluate the transformation at $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 0, \sigma_2/\sigma_1 = 1$, then there is no information in the likelihood function about $\{s_t\}$. As discussed in Koop and Potter (1999), one can use this lack of identification to simplify marginal likelihood calculations using the Savage-Dickey Density ratio. In this case, conditional on the

[7] The informative priors on the parameters of the Markov switching model used here rule out some of the computationally based non-standard problems directly.

generated sequence of $\{\widetilde{C}_t, \widetilde{s}_t\}$ we have:

$$\frac{\int l(\{\widetilde{C}_t, \widetilde{s}_t\}|\boldsymbol{\beta}_1, \sigma_1) b(\boldsymbol{\beta}_1, \sigma_1) d\boldsymbol{\beta}_1 d\sigma_1}{\int l(\{\widetilde{C}_t, \widetilde{s}_t\}|\boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_2, \sigma_2) b(\boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_2, \sigma_2) d\boldsymbol{\beta}_1 d\sigma_1 d\boldsymbol{\beta}_2 d\sigma_2}$$
$$= \frac{p(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 0, \sigma_2/\sigma_1 = 1|\{\widetilde{C}_t, \widetilde{s}_t\}, \mathsf{U}, \boldsymbol{\varphi}^-)}{b(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 0, \sigma_2/\sigma_1 = 1|\boldsymbol{\varphi}^-)},$$

where $\boldsymbol{\varphi}^-$ signifies the parameter space excluding the parameters of the common factor model.[8] Using the methods of Koop and Potter (2000), the LHS of this expression can be directly calculated at each iteration of the Gibbs sampler, for normal inverted gamma prior distributions that are independent across regimes. If this quantity is averaged across draws of $\boldsymbol{\varphi}^-$ and $\{\widetilde{C}_t, \widetilde{s}_t\}$ from the Gibbs sampler we will have:

$$\frac{p(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 0, \sigma_2/\sigma_1 = 1|\mathsf{U})}{b(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 0, \sigma_2/\sigma_1 = 1)},$$

which is the Savage Dickey ratio for the Bayes factor of a 'no-switching common factor' versus a 'switching common factor model'.

Our choice of the prior distribution presents a difficulty when implementing this approach, since the prior imposes some nonlinear restrictions on the parameter space. Instead of attempting to directly incorporate these restrictions in the calculation of the conditional marginal likelihood, we calculate the conditional marginal likelihood for the unrestricted case. In practice, the differences in the conditional marginal likelihoods are so large that the computer is rarely able to distinguish the conditional Bayes factor from zero. Since we find few violations of the nonlinear restrictions on the parameter space, an adjustment for the restrictions on the prior would not change the overall result of strong support for the Markov switching model in the data.

4 Data Description

We use unemployment rates for seven age groups for the period from 1948Q1 to 2000Q2.[9] The age groups are $16-19, 20-24, 25-34, 35-44, 45-54, 55-$

[8] In practice, we use different priors between the no switching and switching models, which require a generalization of the density ratio. For simplicity, we ignore this complication here.

[9] This is the earliest available starting point for high frequency and disaggregate unemployment rates in the United States. For a descriptive analysis of unemployment statistics

64, and 65 and over. We constructed the unemployment rates from the three-month averages of the estimated total unemployment for each group, divided by the three-month average of the estimated labor force for each group:

$$UR_{kt} = \frac{TU_{kt}}{L_{kt}}.$$

These underlying data are not seasonally adjusted. We adjusted the age-specific unemployment rates using a seasonal factor given by the ratio of the published seasonally adjusted to the unadjusted total civilian unemployment rates. We also constructed weights for each group given by equation (4). Note that these weights are not seasonally adjusted. Our adjustment procedure produces an overall unemployment rate that is virtually identical to the published series. This would obviously not be the case if we also seasonally adjusted the labor force weights.

Our focus in on the impact of demographic changes related to the entry into the labor market and subsequent aging of the baby-boom generation, which is illustrated in Figure 2. In particular, the figure shows the time path of the civilian labor force for different age groups. The aging of the baby boomers can be observed in the pattern of changes in the age composition of the labor force over time. The wave of young workers (age 16-19) had its peak at the end of the 1970s. Ten years later, some of these workers fell into the 25-35 age category that peaked in 1990. Currently, the majority of the workers is now between 35-44.

What are the implications of these demographic changes in the labor market? In order to answer this question it is important to understand how the unemployment rate behaves differently across different age groups. Table 1 contains some sample statistics on unemployment rates for different age groups, while Table 2 contains information on the relative weights of each group in the labor force, with the weights scaled by 100. The unemployment rate for teenagers is much higher and more volatile than for the other age groups. In fact, the mean and variance of the unemployment rate decrease steadily as workers age. For example, teenagers have an average unemployment rate of 15.5%, workers between 35-45 have an average of 4%, while the aggregate rate is 5.7% (Table 1).

over 120 years, see Denman and McDonald (1996).

Table 1:Statistics for Unemployment Rates

U_{kt}\Statistic	Mean	St. Dev.	Min	Max
16-19	15.5	3.6	6.3	25.0
20-24	9.0	2.4	3.9	16
25-34	5.6	1.7	2.2	10.8
35-44	4.0	1.2	1.8	8.2
45-54	3.6	1.0	1.8	6.7
55-64	3.7	1.0	1.7	6.2
65+	3.6	0.8	1.7	6.1

As a result of these differences, secular changes in the relative participation of young workers in the labor force has had a significant impact on aggregate unemployment. In particular, as can be seen in Figures 1 and 2, the long secular upswing in the aggregate unemployment rate coincides with the entry of young workers into the labor market. The participation of teenagers in the labor market increased substantially in the 1960s reaching a peak in the mid 1970s, while the participation of workers between 35-45 fell substantially during this period. The subsequent aging of the baby boom generation in the 1980s and 1990s is associated with a reduced fraction of teenagers in the labor force and a substantial increase in the proportion of workers between age 35 and 55 (Figure 3 and Table 2).

Table 2: Statistics for Labor Force Weights

ω_{kt}\Statistic	Mean	St. Dev.	Min	Max
16-19	7.3	1.4	5.1	10.7
20-24	11.8	2.0	8.4	15.3
25-34	24.2	3.2	18.7	29.6
35-44	22.7	2.2	17.6	27.5
45-54	18.7	2.2	14.8	21.8
55-64	11.8	1.7	8.8	14.2
65+	3.6	1	2.5	5.3

Finally, we present the contemporaneous correlation matrix for the unemployment rates ordered from the youngest to the oldest:

Table 3: Correlation Matrix for Unemployment Rates

	16-19	20-24	25-34	35-44	45-54	55-64	65+
16-19	1						
20-24	0.88	1					
25-34	0.84	0.96	1				
35-44	0.79	0.92	0.97	1			
45-54	0.68	0.86	0.91	0.95	1		
55-64	0.50	0.70	0.76	0.84	0.92	1	
65+	0.28	0.43	0.39	0.45	0.53	0.64	1

It is interesting to notice how the unemployment rates are correlated between groups close in age, and much less so when comparing younger workers with older workers. The reason is that unemployment rates for younger workers are very volatile over the entire sample, and display particularly accentuated oscillations over the business cycle. On the other hand, the volatility of unemployment decreases monotonically as workers age. Table 1 shows that the standard deviation of teenager's unemployment is more than 3 times higher than the standard deviation for workers age 35 or older. In addition, the unemployment rates for older workers exhibit much smaller oscillations around business cycle turning points. As argued by Shimer (1998), this may be explained by the fact that although younger workers do not have trouble finding jobs, they are more frequently fired.[10]

Anticipating our empirical results, the resulting dynamic factor model is broadly consistent with the sample moments given in Table 1 for the group of workers $35-44$. Notice that this is not only the mid-group, but it is also the one with the highest relative participation in the labor force. On the other hand, as it turns out to be somewhat striking given the contemporaneous correlation between unemployment rates, our measure of latent unemployment constructed from the factor loadings and demographic weights tracks almost exactly the overall unemployment rate.

[10] Young workers in the U.S. have a mean and median unemployment duration smaller than older workers.

5 Priors and Results

5.1 Properties of the Prior Distribution

We start by considering the choice of the hyperparameters of the normal-inverted gamma priors for the common factor parameters in each regime, under the restriction used in the estimation that the lag length is 1. These priors are the most important ones for interpretation of the sample evidence. We begin by eliciting a prior that is relatively noninformative, but accords with our subjective prior beliefs. To simplify matters, the prior covariances are assumed to be zero for all the conditional mean parameters in the model. As we work with one lag in each regime, this does not seem controversial. Since we are examining unemployment rates, it does not make economic sense to assume that the prior means are zero. Instead, we specify the prior means in terms of common empirical findings regarding asymmetries in the unemployment rate.

The prior is specified in two steps. First we describe the unrestricted prior, then we discuss various restrictions on the prior that are imposed as in Geweke (1986). For the low unemployment regime, we specify a mean of 0.4 for the intercept, and of 0.8 for the autoregressive parameter. The respective variances of these parameters are 1 and 0.3. For the high unemployment regime, we specify a mean of 2.5 for the intercept, and of 0.5 for the autoregressive parameter. The variances of these parameters are 0.8 and 0.95, respectively. These are the priors used in generating candidate random draws from the conditional posterior. These draws must satisfy stationarity as well as the restriction $\eta(1) > \eta(2) > 0$. Finally, the degrees of freedom for the inverted gamma priors are 3 for both regimes, and the prior mean of the variance is set to 0.1.[11]

We simulate from this combined normal inverted gamma prior for two regimes in order to obtain some prior features of interest. The prior means of η are around 1% for the low unemployment regime, and around 13.5% for the high unemployment regime. These values are reasonable, given that the η's are upper and lower bounds under the AR(1) assumption. About 32% of the joint priors have $\eta(1) > 6\%$ and $\eta(2) < 4\%$.

For the factor loadings, we use a normal prior centered at 1 with a diagonal variance matrix, individual variance of 4 for unemployment rates of workers

[11] The variance for the conditional mean parameters already takes into account the prior mean for the variance of the errors.

age 25 and over, and variance 1 for the unemployment rates of workers 16-19 and 20-24. Once again, this is an uninformative choice. Naturally, we impose the degenerate prior that the factor loading is 1 for the unemployment rates of workers age 35-44 in order to normalize the dynamic factor.

For the measurement error innovation variance, we use the minimally informative inverted gamma prior with degrees of freedom 3 and mean 1. For the autogressive structure of the measurement errors, we use a more informative Gaussian prior. The prior mean is set to zero and the variance is a diagonal matrix. The standard deviation of the autoregressive coefficient for the group age 25 and older is set to 0.07, while for the 16-19 and 20-24 groups it is equal to 0.1. These priors are based on the fact that the Current Population Survey currently follows a "4 month in 8 month out, 4 month in" rotation for households. Thus, the measurement error in the survey across households should not be too strong.

For the parameters of the Markov chain transition matrix, we use a common beta prior. For the probability of unemployment staying in the low regime ρ_{22} or staying in the high regime ρ_{11}, the parameters are $(9, 1)$. This is equivalent to having observed about 10 observations of the Markov chain. Thus, it will be easily dominated by the sample information if there are repeated switches. Using standard formula for the beta distribution, this implies a very vague prior on the expected duration $1/(1-\rho_{22})$ or $1/(1-\rho_{11})$, since the first moment does not exist. The median duration obtained from the simulation is 13.6 quarters, with a probability equal to 0.09 that the duration is longer than 25 years, and a probability of 0.08 that it is less than 1 year.

5.2 Empirical Results

The model is estimated with the order of the autogressive parameters $q = 1, p = 1$. The unemployment rate of the age group $35 - 44$ is chosen as the variable with factor loading equal to unity. The Gibbs sampler was run with a burn-in phase of 200 iterations and a further 4000 iterations. Repeated runs of the sampler with different initial conditions produced very few changes in the posterior features suggesting that the sampler has converged.

As noted above, the difference in marginal likelihoods is so large between a model with switching and no switching (a Bayes factor of 10^{19} in favor of switching) that there is little information in the conditional Bayes factors. Figure 4 shows evidence on the F-Statistics calculated at each iteration of the Gibbs sampler. Here we can see that the F-statistics are uniformly higher than a 0.1% critical value of around 7 obtained from a F-distribution.

The resulting latent unemployment is highly correlated with total unemployment rate (0.995). In addition, the latent factor is also centered with the actual unemployment rate. With respect to the posterior mean of the factor loadings, they show a pattern consistent with the differences in means and standard deviations reported in Table 1. In particular, young labor market participants have the highest factor loading, and the loadings decline with age.[12]

Figure 5 plots the actual unemployment rate with posterior means of the demographically adjusted bounds, calculated from equations (6) and (7). The unemployment rate fluctuates inside the high-mismatch and low mismatch averages almost the entire sample. In two main instances the bounds are violated, which correspond to periods in which the unemployment rate reached its minimum and maximum values in the sample. One was in the early 1950s, when unemployment rates were very low (2.6%). The other was in the early 1980s, when unemployment rates peaked at 10.7%. The role of the demographic adjustment to the upper and lower bounds can also be seen in this figure. The high and low mismatch averages display a hump in the 1970s, reflecting the entry and subsequent aging of the baby boom in the labor market.

In order to further investigate the subsequent influence of the aging baby boomers on unemployment, we compare changes in the unweighted and weighted latent unemployment factors. First, recall that the weighted latent unemployment can be decomposed into a factor effect and a demographic effect, as described in Section 2.2. We consider the change in unemployment for 3 cases: from the earlier dates 1978, 1982, and 1992 to the present (2000). For example, comparing unemployment in 2000 and 1982, equation (10) becomes:

[12] Similar patterns are also found in the posterior means of the autogressive model for the individual measurement errors.

$$
\begin{aligned}
\mathcal{U}_{2000} - \mathcal{U}_{1982} &= \\
& (C_{2000} - C_{1982}) \sum_{k=1}^{K} \omega_{kt} \lambda_k \text{ Factor Effect} \\
& + C_{1982} \sum_{k=1}^{K} (\omega_{k2000} - \omega_{k1982}) \lambda_k \text{ Demographic Effect}
\end{aligned}
$$

For changes in unemployment from 1978 and 1982 to the present, we find as a demographic effect a reduction of 75 basis points in the latent unemployment rate. This is is very close to the ones obtained in Shimer's (1998) analysis of the actual unemployment rates. Since 1992, there has been a mild demographic induced reduction in the latent unemployment of 10 basis points.

Second, consider the case in which demographic changes have no effect on the change in latent aggregate unemployment, that is, the factor without demographic or factor loading weights, as described in equations (8) and (9). Since we normalize the factor loading of the 35-44 year old group to unity, the factor without demographic\factor loading weights is a measure of the underlying state of the labor market for this group. In this case, equation (8) becomes:

$$
C_t \left(\sum_{k=1}^{K} \omega_{kt} \lambda_k - 1 \right),
$$

which gives the difference between latent unemployment and a hypothetical labor market where all participants are in the 35-44 age group. Figure 6 shows the factor with and without demographic\factor loading weights, while Figure 7 plots the difference between these two series, which corresponds to the series derived from the equation above.

There are some interesting findings from this analysis. First, changes in the composition of different age workers raise the aggregate unemployment rate from a minimum of 0.61% in the early 1950s to a maximum of 3.07% in 1982. The average impact of age composition on unemployment is 1.55% with a standard deviation of 0.54%. The entry of the baby boom into the labor market in the 1960s and 1970s increased the aggregate unemployment rate by about 1.8%.[13] The subsequent aging of the baby boomers has decreased unemployment by around 1.7%.[14] These findings are also similar to the ones obtained by Shimer's (1998).

A second interesting observation is that the baby boom effect induces a steep increase in the unemployment rate at the beginning of recessions in the 1970s and 1980s. In particular, the values of the difference between the unweighted and weighted factors doubled during the 1970, the 1975, and 1980-82 recessions. The reason is that the labor force weights of young labor market participants were increasing at the time that these switches to the high unemployment regime occurred. One can also observe that in the 1990 recession the impact of age on the unemployment rate was substantially smaller.

The properties of the dynamic factor stochastic model are very similar to those expected and broadly consistent with the sample moments given in Table 1 for the group of workers 35-44. Notice that this is not only the mid-group, but also the one with the highest relative participation in the labor force. If we evaluate the parameters at their posterior means, the implied model is:

$$C_t = \begin{cases} 1.41 + 0.75 C_{t-1} + \varepsilon_{1t} & \text{if } s_t = 1 \\ 0.16 + 0.94 C_{t-1} + \varepsilon_{2t} & \text{if } s_t = 2 \end{cases},$$

with $P[s_{t+1} = 1 | s_t = 1] = \rho_{11} = 0.81$ and $P[s_{t+1} = 2 | s_t = 2] = \rho_{22} = 0.93$, $\varepsilon_{1t} \sim N(0, 0.48)$ and $\varepsilon_{2t} \sim N(0, 0.15)$.

However, analysis of the posterior means can be misleading, since they ignore the amount of uncertainty about each parameter and their joint variation. In order to gain some insight into the importance of this, we compare the posterior means of $\eta(1)$ and $\eta(2)$ and the average duration in each regime to those implied by the model evaluated at the posterior means of the parameters. In the high unemployment state, $\eta(1)$ =5.81 (compared to $\alpha_1/(1-\phi_1) = 5.64$), while $\eta(2) = 2.55$, in state 2 (compared to

[13]More specifically, since the baby boomers correspond to births between 1946 and 1964, here we calculate the average difference between the two factors using sample from 1962 to 1983. This period corresponds to the years in which the first and last generations of the baby boomers were beween age 16-19.

[14]This is calculated using sample from 1981 on, which correponds to the year in which the first generation of the baby boomers reached the age of 35.

$\alpha_2/(1-\phi_2) = 2.66$). For the posterior mean of duration, state 2 lasts 16.9 quarters (compared to $1/(1-\rho_{22}) = 14.3$ quarters), while state 1 lasts 5.9 quarters (compared to $1/(1-\rho_{11}) = 5.3$). Thus, the posterior means provide a reasonably accurate summary of the differences between regimes in terms of average dynamics.

Now, consider the following experiments with impact multipliers to illustrate the asymmetries in unemployment, which consist of sharp increases followed by long slow declines as implied by the model. Assume that the dynamic factor is in regime 1 with an unemployment rate of 5.5%, and that it switches to the low unemployment regime (state 2) with no future shocks. At first unemployment stays high ($\alpha_2 + \phi_2 \times 5.5 = 5.33$) and then it gradually declines. After 4 years, which is the typical duration of the low missmatch regime, the unemployment rate is equal to $[\alpha_2/(1-\phi_2)] + \phi_2^{16} \times [5.5 - \alpha_2/(1-\phi_2)] = 3.7$. That is, it drops 63% of the way to the long run low unemployment level. In contrast, if we start the dynamic factor in regime 2 with an unemployment rate of 2%, upon a switch to regime 1 unemployment immediately jumps to $(\alpha_1 + \phi_1 \times 2) = 2.91$. After 6 quarters, which is the average duration of the high mismatch regime, unemployment reaches a level of $[\alpha_1/(1-\phi_1)] + \phi_1^{6} \times [2 - \alpha_1/(1-\phi_1)] = 5.0$, which corresponds to a 82% increase of the way to the long run high unemployment level.

One final aspect of interest in the estimated model is the large diference in the size of the shocks hitting the high unemployment regime versus the low unemployment regime. This difference implies that one should expect quite smooth downward movements in unemployment but more irregular upward movements, which is depicted in Figure 6. The asymmetries found in unemployment are also summarized in Figure 8, which plots the posterior mean of the high mismatch state. Unemployment switches to the high phase right at the beginning of recessions, and switches back to a low unemployment regime a couple of quarters after economic troughs, which illustrates the fast and steep growth of unemployment and its slower and more gradual decline around business cycle turning points. In addition, the high unemployment state is less persistent and corresponds to periods in which unemployment is more volatile, compared to the low unemployment state.

Overall, as shown by the Bayes factor and the collection of F-statistics displayed in Figure 6, there are large differences between the two regimes of unemployment. Furthermore, these differences are sensible and economically important.

6 Conclusions

This paper examines the secular and business cycle movements in unemployment over the past 50 years. We develop a common factor model with Markov switching to analyze disaggregate unemployment rates of workers grouped by age. We find a strong statistical evidence in favor of the common factor structure and of the switching between high and low unemployment rate regimes. In particular, the latent factor exhibits the stylized asymmetries in unemployment associated with the phases of the business cycle. In addition, the model captures the low frequency fluctuations in the U.S. unemployment rate associated with secular demographic changes. We find that demographic adjustments to the unemployment rate can account for a great deal of secular changes in the U.S. unemployment rates. In particular, they explain some of the increases in unemployment in the 1970s and 1980s, and the subsequent decrease in the last 18 years.

The framework we have constructed is being extended in two main ways. First, we are examining the introduction of a second factor, in order to model short term fluctuations in disaggregate unemployment rates. In particular, the factor is identified by including changes in inflation rates and in output amongst the observed variables to identify business cycle patterns. The model implies a time-varying Phillips curve associated with the phases of the business cycle. Second, we are extending the framework to include a more detailed demographic decomposition of the total unemployment rate, using observations on unemployment rates by age, sex, education, and race.

7 Appendix

This appendix gives information on the prior distributions used in the Gibbs sampler and the resulting conditional posterior distributions.

1. **Factor Loadings:** we assume that the factor loadings are a priori normally distributed and independent of each other with mean $\mu(\underline{\lambda}_k)$ and precision $\tau(\underline{\lambda}_k)$. Conditional on $\{\widetilde{C}_t\}$, $\Theta(L), \Sigma_K$ we draw the $K \times 1$ vector of factor loadings $\boldsymbol{\lambda}$ from (independent) normal distributions.

For the generic loading λ_k we have the sample information:

$$\left[\sum_{t=q+1}^{T} C_{kt}^{*2}\right]^{-1}, \sum_{t=q+1}^{T} C_{kt}^{*} U_{kt}^{*},$$

where $C_{kt}^{*} = \widetilde{C}_t - \theta_{1k}\widetilde{C}_{t-1} - \cdots - \theta_{qk}\widetilde{C}_{t-q}$. Thus, the conditional posterior is normal with mean:

$$\frac{\tau(\underline{\lambda}_k)\mu(\underline{\lambda}_k) + \sum_{t=q+1}^{T} C_{kt}^{*} U_{kt}^{*}/\sigma_{kk}}{\tau(\underline{\lambda}_k) + \sum_{t=q+1}^{T} C_{kt}^{*2}/\sigma_{kk}}$$

and precision:

$$\left[\tau(\underline{\lambda}_k) + \sum_{t=q+1}^{T} C_{kt}^{*2}/\sigma_{kk}\right].$$

For one of the elements of $\boldsymbol{\lambda}$ we impose the prior belief that it is equal to 1.

2. Innovation Variance to Measurement Error: we assume that the innovation variances to the measurement errors are a priori independent inverted Gamma distributions with degrees of freedom of $\underline{\xi}$ and scale $\underline{s}^{-2}$. Conditional on $\{\widetilde{C}_t\}$, $\Theta(L), \boldsymbol{\lambda}$ we draw the posterior of the measurement error innovations variances as independent gamma distributions. For the generic measurement error variance σ_{kk} we have the sample information:

$$\sum_{t=q+1}^{T} (U_{kt}^{*} - \lambda_k C_{kt}^{*})^2, T - q,$$

which is combined with the prior degrees of freedom of ξ and "prior sum of squares" $\underline{\xi s}^2$ to obtain the posterior degrees of freedom $\underline{\xi} + T - q$ and sum of squares $\underline{\xi s}^2 + \sum_{t=p+1}^{T}(U_{kt}^{*} - \lambda_k C_{kt}^{*})^2$.

3. Measurement Error Autoregressive Parameters: we assume that the measurement error autoregressive parameters are truncated normal with the truncation determined by the stationarity condition. Since we use rejection sampling to implement the truncated prior, we start by

describing the proposal distribution. The location and scale parameters of the truncated normal are given by the vector $\mu(\underline{\boldsymbol{\theta}}_k)$ and matrix $\tau(\underline{\boldsymbol{\theta}}_k)^{-1}$. Thus, without the stationarity condition we would have the Gaussian prior. Conditional on $\{\widetilde{C}_t\}$,$\boldsymbol{\lambda}, \Sigma_K$ we draw a proposed set of the measurement autoregressive coefficients from independent multivariate Gaussian distributions. For the generic measurement error autoregression k the sample information is:

$$\left[\mathsf{Z}_k^{\prime}\mathsf{Z}_k\right]^{-1}, \mathsf{Z}_k^{\prime}\mathsf{W}_k,$$

where $\mathsf{W}_k = [\overline{U}_{kq+1}, \cdots, \overline{U}_{kT}]'$ and

$$\mathsf{Z}_k = \begin{bmatrix} \overline{U}_{kq} & \cdots & \overline{U}_{k1} \\ \overline{U}_{kq+1} & \cdots & \overline{U}_{k2} \\ \vdots & \vdots & \vdots \\ \overline{U}_{kT-1} & \cdots & \overline{U}_{kT-q} \end{bmatrix},$$

and $\overline{U}_{kt} = U_{kt} - \lambda_k C_t$. This is combined with the unrestricted Gaussian "prior" on the autoregressive coefficients in the standard way to obtain a normal distribution with variance matrix:

$$\left[\tau(\underline{\boldsymbol{\theta}}_k) + \mathsf{Z}_k^{\prime}\mathsf{Z}_k/\sigma_{kk}\right]^{-1}$$

and mean:

$$\left[\tau(\underline{\boldsymbol{\theta}}_k) + \mathsf{Z}_k^{\prime}\mathsf{Z}_k/\sigma_{kk}\right]^{-1}\left[\tau(\underline{\boldsymbol{\theta}}_k)\mu(\underline{\boldsymbol{\theta}}_k) + \mathsf{Z}_k^{\prime}\mathsf{W}_k/\sigma_{kk}\right].$$

This distribution is used to generate draws until the stationarity condition is satisfied.

4. **Bayes Factor Calculation:** conditional on $\{\widetilde{C}_t\}$, $\{s_t\}$ we calculate the marginal likelihood of $\{\widetilde{C}_t\}$ under both the Markov switching model and the no switching model priors. As discussed in the main text this is an approximate calculation.

5. **Dynamic Factor Model Parameters:** as noted in the main text, assuming that $\{\beta_i, \sigma_i\}$ are a priori distributed normal inverted gamma and independently of each other simplifies the calculation of Bayes factors considerably. Conditional on $\{\widetilde{C}_t\}, \{s_t\}$ we make candidate draws

of the autoregessive model parameters for the two regimes from the posterior distributions, which are inverted-gamma normal distribution. These draws are made sequentially starting with the low unemployment regime first as described in the main text. Denote the prior precision matrix of $\boldsymbol{\beta}_i$ by $\tau(\underline{\beta}_i)\sigma_i^2$ and its prior mean by $\mu(\underline{\beta}_i)$. Let the prior degrees of freedom of σ_i^{-2} be $\underline{\nu}_i$ and its mean be $\underline{s}_i^{-2}$. The sample information is contained in the draw of the common factor and the Markov states. Let $y_{it} = 1(s_t = i)\widetilde{C}_t$ and define the vector $\mathsf{y}_i = [y_{i1+p}, \ldots y_{iT}]'$ and the matrix:

$$\mathsf{X}_i = \begin{bmatrix} 1 & y_{ip} & \cdots & y_{i1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & y_{iT-1} & \cdots & y_{iT-p} \end{bmatrix}.$$

The posterior degrees of freedom of the variance in regime i are:

$$\overline{\upsilon}_i = \underline{\nu}_i + \sum_{t=1}^{T} 1(s_t = i).$$

The posterior scale of the inverted gamma is given by:

$$\begin{aligned} \overline{\upsilon s_i^2} &= \underline{\nu}_i \underline{s}^2 + \upsilon s^2 + \left[\mu(\overline{\boldsymbol{\beta}}_i) - \boldsymbol{\beta}_i\right]' \mathsf{X}_i'\mathsf{X}_i \left[\mu(\overline{\boldsymbol{\beta}}_i) - \mathsf{X}_i\right] \\ &+ \left[\mu(\overline{\boldsymbol{\beta}}_i) - \mu(\underline{\beta}_i)\right]' \tau(\underline{\beta}_i) \left[\mu(\overline{\boldsymbol{\beta}}_i) - \mu(\underline{\beta}_i)\right], \end{aligned}$$

where the posterior mean of $\boldsymbol{\beta}_i$ is:

$$\mu(\overline{\boldsymbol{\beta}}_i) = \tau(\overline{\boldsymbol{\beta}}_i) \left[\tau(\underline{\beta}_i)\mu(\underline{\beta}_i) + \mathsf{X}_i'\mathsf{y}_i\right],$$

the posterior precision matrix is:

$$\tau(\overline{\boldsymbol{\beta}}_i) = \tau(\underline{\beta}_i) + \mathsf{X}_i'\mathsf{X}_i$$

and sample sum of squared errors is:

$$\upsilon s^2 = \left[\mathsf{y}_i - \mathsf{X}_i\widehat{\boldsymbol{\beta}}_i\right]' \left[\mathsf{y}_i - \mathsf{X}_i\widehat{\boldsymbol{\beta}}_i\right].$$

In addition,

$$\widehat{\boldsymbol{\beta}}_i = \left[\mathsf{X}_i'\mathsf{X}_i\right]^{-1} \mathsf{X}_i'\mathsf{y}_i.$$

The candidate draws are obtained by first drawing a value of σ_i^2 from a Gamma distribution with posterior degrees of freedom and scale given above. This draw is then used to obtain the posterior variance of $\boldsymbol{\beta}_i$ as $\sigma_i^2\tau(\overline{\boldsymbol{\beta}}_i)^{-1}$. This variance together with the mean described above is used to obtain a candidate draw of $\boldsymbol{\beta}_i$. If the draw is rejected, we draw σ_i^2 and $\boldsymbol{\beta}_i$ again and repeat until a succesful draw is obtained.

6. Markov States: conditional on $\{\widetilde{C}_t\}, \rho_{11}, \rho_{22}, \beta_1, \beta_2, \sigma_1, \sigma_2$, we draw the sequence $\{s_t\}$ as discussed in the main text using the methods of Chib (1996) with the restriction that $S_1 = \cdots = S_p = 2$.

7. Transition Probabilities: we assume that the transition probabilities are a priori independent with a Beta distribution. Under the Beta distribution prior, the posterior is also in the Beta family. We focus on updating the transition probabilities for state 1, as the case for state 2 follows analagously. The Beta density is proportional to:

$$\rho_{11}^{\underline{\delta}_{11}-1}(1-\rho_{11})^{\underline{\delta}_{12}-1}, \underline{\delta}_{11} > 0, \underline{\delta}_{12} > 0.$$

The updating of the parameters of the Beta distribution is direct with:

$$\begin{aligned} \overline{\delta}_{11} &= \underline{\delta}_{11} + \sum_{t=p+1}^{T} P[S_t = 1, S_{t-1} = 1 | \widetilde{C}^T], \\ \overline{\delta}_{12} &= \underline{\delta}_{12} + \sum_{t=p+1}^{T} P[S_t = 2, S_{t-1} = 1 | \widetilde{C}^T]. \end{aligned}$$

8. Common Factor: conditional on $\Theta(L), \boldsymbol{\lambda}, \Sigma_K, \beta_1, \beta_2, \sigma_1, \sigma_2$, the Kalman filter is run on the observed data. The Kalman filter is initialized at the stationary distribution for $\{C_t\}$ implied by $\boldsymbol{\beta}_2, \sigma_2$. Then, using the recursions described above, a draw of $\{\widetilde{C}_t\}$ is obtained and we return to step 1 above. We also calculate various features of the latent unemployment rate using the draw of the common factor and factor loading.

References

[1] Abbring, J.H., G.J.V.D. Berg and J.C.V. Ours, 1999, "Business Cycles and Compositional Variation in U.S. Unemployment," *Tinbergen Institute Working Paper* 97-050/3.

[2] Albert, J. and S. Chib, 1993, "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts," *Journal of Business and Economic Statistics*, 11, 1-15.

[3] Boldin, M., 1994, "Dating Turning Points in the Business Cycle," *Journal of Business*, 1, 97-131.

[4] Carter, C. and P. Kohn, 1994, "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.

[5] Chauvet, M., 1998, "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches," *International Economic Review*, 39, 4, 969-96.

[6] Chib, S., 1995, "Marginal Likelihood from Gibbs Output," *Journal of American Statistical Association*, 90, 1313-1321.

[7] Chib, S., 1996, "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models, *Journal of Econometrics*, 75, 79-97.

[8] Chib, S., 2001a, "Monte Carlo Methods and Bayesian Computation: Overview," in S.E. Fienberg and J.B. Kadane (eds.), *International Encyclopedia of the Social and Behavioral Sciences: Statistics*, Amsterdam: Elsevier Science, in press.

[9] Chib, S., 2001b, "Markov Chain Monte Carlo Methods: Computation and Inference," in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, Amsterdam: North Holland, in press.

[10] De Jong, P. and N. Shephard, 1995, "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339-350.

[11] Denman, J. and P. McDonald, 1996, "Unemployment Statistics from 1881 to the Present Day," *Labour Market Trends*, 5-18.

[12] Diebold, F.X. and G.D. Rudebusch, 1996, "Measuring Business Cycles: A Modern Perspective," *Review of Economics and Statistics*, 78, 67-77.

[13] Franses, P. H., 1995, "Quarterly U.S. Unemployment: Cycles, Seasons and Asymmetries," *Empirical Economics*, 20, 717-725.

[14] Fruhwirth-Schnatter, S., 1994, "Data Augmentation and Dynamic Linear Models," Journal of Time Series Analysis, 15, 183-202.

[15] Geweke, J., 1977, "The Dynamic Factor Analysis of Economics Time Series Model," in D. Aigner and A. Goldberge (eds.), Latent Variables in Socioeconomic Models, 365-383. Amsterdam: North-Holland.

[16] Geweke, J., 1986, "Exact Inference in the Inequality Contrained Normal Linear Regression Model," Journal of Applied Econometrics, 1, 127-142.

[17] Geweke, J., 1999, "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," Econometric Reviews, 18, 1-126.

[18] Gordon, R., 1982, "Inflation, Flexible Exchange Rates and the Natural Rate of Unemployment," in Martin Baily (ed.), Workers, Jobs and Inflation Brookings, 89-152.

[19] Hamilton, J., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica, 57, 357-384.

[20] Juhn, C., K., M. Murphy, and R. Topel, 1991 "Why has the Natural rate of Unemployment Increased over Time?" Brookings Papers On Economic Activity, 75-126.

[21] Katz, L. F and A. B. Krueger, 1999, "New Trend in Unemployment?" Brookings Review, 4, 4-8.

[22] Kim, C. J., 1994, "Dynamic Linear Models with Markov-Switching," Journal of Econometrics, 60, 1-22.

[23] Kim, C-J. and C. Nelson, 1999, "State Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications," Cambridge: MIT Press.

[24] Koop, G. and S. M. Potter, 1999, "Bayes Factors and Nonlinearity: Evidence from Economic Time Series," Journal of Econometrics, 88, 251-281.

[25] Koop, G. and S. M. Potter, 2000, "Nonlinearity, Structural Breaks or Outliers in Economic Time Series?" in Nonlinear Econometric Modeling in Time Series Analysis, William Barnett (ed.), Cambridge: Cambridge University Press, 61-78.

[26] Montgomery, A.L., V. Zarnowitz, R S. Tsay, and G. C. Tiao, 1998, "Forecasting the U.S. Unemployment Rate, Journal of the American Statistical Association, 442, 478-493.

[27] Neftci, S., 1984, "Are Economic Time Series Asymmetric over the Business Cycle," Journal of Poltical Economy, 92, 307-328.

[28] Rothman, P., 1993, "Further Evidence on the Asymmetric behavior of Unemployment Rates over the Cycle," Journal of Macroeconomics, 13, 291-298.

[29] Shephard, N., 1994, "Partial Non-Gaussian State Space," Biometrika, 81, 115-131.

[30] Shimer, R., 1998, "Why is the Unemployment Rate So Much Lower?" in B. Bernanke and J. J. Rotemberg (eds.), NBER Macroeconomics Annual, Cambridge, MIT Press 11-61.

[31] Skalin, J. and T. Teräsvirta, 2001, "Modelling Asymmetries and Moving Equilibria in Unemployment Rates," forthcoming in Macroeconomic Dynamics.

[32] Stock, J.H. and M.W. Watson, 1989, "New Indices of Coincident and Leading Indicators," in O. Blanchard and S. Fischer (eds.), NBER Macroeconomics Annual, Cambridge, MIT Press.

[33] Vredin, A. and A.Warne, 2000, "Unemployment and Inflation Regimes," mimeo, Research Department, Sveriges Riksbank.