Structural Breaks in Panel Data Models: A New Approach

Wei Liao

November 15, 2008

Abstract

This paper provides a new econometric framework to make inference about structural breaks in panel data models. The main contribution is twofold. First, different from most existing research which studies structural-change problem in a single time series setup, this paper investigates structural breaks by exploiting the rich information in the panel data. Second, in contrast to the common-break method that assumes all series change structure at the same time, my method assumes the change points of different time series follow a common distribution, allowing for the heterogeneity in the timing of structural changes across series while retaining the commonality among series. For the case of a single structural break in each time series, I propose a general nonparametric method by assuming that change points of different time series follow a multinomial distribution. If in addition, each time series is allowed to have multiple structural breaks, a special form of the common distribution is imposed to restrict the joint distribution of multiple breaks.

By incorporating both the similarity and the heterogeneity among series, this method not only improves the quality of change-point estimation, but it also reveals useful information about how different series respond to a common shock. This helps to answer important questions such as how stocks of different industries respond to a new policy or how different experimental groups respond to a new treatment. Monte Carlo simulations show that this method greatly increases the precision of change-point estimation, and the estimation algorithm is fast. I apply this method to investigate the volatility decline, or the Great Moderation, in 50 US states. Using quarterly personal income data, I find that the cross-sectional distribution of break dates has two peaks: one being around 1984, which has been generally recognized as the start of the Great Moderation; the other being around 1988, which corresponds to the stock market crash of the previous year. Only a small number of states had no evidence of structural change in the volatility.

Key Words Panel Data, multiple structural breaks, Bayesian inference, Hidden Markov Chain, Great Moderation

1 Introduction

Many recent studies highlight the presence of structural instability in key economic indexes and financial time series, such as output growth, unemployment, exchange rate, and stock returns (Stock and Waston, 1996). Various economic events can lead to structural changes in a wide range of time series, such as financial liberalization (Behaert, 2002), changes in exchange rate regimes, and the introduction of new monetary policies. The most striking example was the Great Moderation, which is well documented in the literature as the substantial reduction of volatility in major US macroeconomic time series since the 1980s.

The precise estimation of a change point has rich implications. First, it helps uncover the source of a structural change by spotting special events around the break dates. Second, it can be used to evaluate the impact of an event or a new policy by estimating the response time of the economy to the shocks.

In addition, from a statistical perspective, ignoring structural breaks in econometric modeling can lead to model misspecification and spurious estimation results of model parameters. For instance, Lamoureux and Lastrapes (1990) find that the extent of persistence in variance of stock-return data may be overstated if structural breaks are ignored. Pesaran and Timmermann (2006) point out that forecasting results based on such a misspecified model are unreliable.

While most research in the structural break literature addresses single time series problem, few studies have been done in a panel (or multivariate) data environment. Assume a series $\{y_t\}_{t=1,...T}$ is subject to a structural change at some time k^*. It is well known in econometric theory that the estimator of change point k^* for a single series is not consistent, and that only the fraction $\frac{\hat{k}}{T}$ converges in probability to the true value $\frac{k^*}{T}$. Even with more observations of y_t, it is not guaranteed that the estimation of k^* would be improved. When the magnitude of the structural break is not large enough or the true change point k^* is too close to the start or end of the sample, the single series approach might fail to detect any structural change even if there is one. Motivated by such observations, Bai (2006) proposes a panel data approach and proves that one is able to estimate the break point k^* itself consistently if a lot of time series are subject to a common break, which significantly improves the quality of the break point estimation. This is the so-called advantage of "borrowed power", which exploits the cross-section information. In the rest of the paper, I will refer to Bai (2006)'s method as "common-break approach".

A key question that arises in the context of panel data is how to build the link for structural changes across different series. The easiest way is to assume that all series experience the structural break at the same time. However, it is not realistic to assume that different series change their structure at exactly the same time. More generally, individual series may experience the structural break at different times, even after experiencing the same event, due to the cross-sectional heterogeneity in the transmission mechanism or response time. Therefore, it is crucial to find a way to relax the conventional common-break assumption while still retaining the common feature of structural changes across different series. This not only improves the estimation quality of change points, but it also provides useful and important information about the cross-sectional pattern of structural changes. Only by taking into account both the similarity and the heterogeneity across sections, is one able to answer important questions, such as how stocks of different industries respond to a new policy or how different experimental groups respond to a new treatment. It is thus imperative to develop a new and effective method for change-point detection in panel data.

This paper develops a new approach for estimating and making inference about structural changes in panel data models based on a Bayesian method. The new method takes into account two key factors. First, different series are subject to common shocks, which are assumed to be the source of the structural changes. Second, various series exhibit heterogeneous responses to these common shocks, and the timing of structural changes is allowed to be different across series. A key assumption is that change points for different time series follow a common distribution. The common-distribution assumption relaxes the conventional common-break restriction while still retaining the common feature of the structural changes. If there is only a single structural change for each series, a non-parametric multinomial distribution assumption is used to model the pattern of change points across sections. This single structural change framework follows Joseph and Wolfson (1993, 1997), who discuss the change-point estimation problem for Poisson panel data with a single structural break for each series. While they only focus on estimating the underlying common distribution of unknown change points, this paper focuses on estimating both the common distribution and the location of change point for each individual series.

In addition, if we allow each time series to have multiple structural breaks, the nonparametric common distribution becomes more complicated as the number of structural breaks increases. Such a complication arises because one has to model the joint distribution of multiple change points. Proper restriction on the joint distribution is needed to keep the model parsimonious and tractable. As a special form of such restrictions, I propose a nonreversible hidden Markov chain model, which is built on the structural break model proposed by Chib (1998). Chib's framework has been applied by many researchers to address various economic or econometric issues, e.g., Kim, Nelson and Piger (2004), and Pesaran, Pettenuzzo and Timmermann (2005). However, existing studies focus on the single series case, and thus require strong assumptions on both break magnitude and locations of the true break dates. As I will discuss later in more detail, Chib's framework is actually a special form of the common distribution model in the univariate case. In this paper, I extend the nonreversible hidden Markov chain model to a multivariate setup. With this modification, it becomes a powerful tool to analyze multiple change points in panel data models.

Finally, I apply the method to study the US state-level facts of the Great Moderation. It is well documented that major US macroeconomic time series, such as output growth and inflation, have experienced a substantial reduction in volatility since the 1980s. However, little has been done to examine whether such a Great Moderation exists in disaggregate data. Owyang, Piger and Wall (2008) conduct a study about the state-level Great Moderation using unemployment data. However, they estimate the structural change state-by-state without taking into account the strong correlation and co-movement among states. I apply the panel data methodology to 50 US state-level quarterly personal income data[1], spanning from 1952Q2 to 2008Q2. The cross-sectional distribution of break dates has two peaks, one around 1984, and the other around 1988. Only a few states show no evidence of structural breaks in the volatility. These findings continue to hold if one assumes two breaks for each state.

The contribution of this paper is twofold. First, it provides a new and effective methodology to analyze change points by using panel data information, which not only improves the estimation precision but also makes study of the cross-sectional pattern of structural breaks possible. Second, I extend the nonreversible hidden Markov chain model to the panel data models, as a special form of the general common-distribution framework. It greatly simplifies the form of the non-parametric common distribution while retaining enough flexibility.

The rest of the paper is organized as follows. Section 2 shows the motivation by

[1] We exclude the state Louisiana due to some data irregularities caused by Hurricane Katrina.

using Monte Carlo simulations. Section 3 describes the non-parametric method to estimate and make inference about structural changes in panel data. Section 4 provides an example of the parametric approach. In section 5, I extend the nonreversible hidden Markov chain model from a univariate to a multivariate data environment, as a special form of the non-parametric method to keep the model parsimonious in the presence of multiple structural changes. In section 6, I apply the panel data model to investigate the Great Moderation in the state-level data.

2 Motivating the Panel Data Approach: Common Break

Assume a series $\{y_t\}_{t=1,\ldots T}$ is subjected to a structure break at some time k^*, $\lambda^* = \frac{k^*}{T}$. It is well known that the estimate of change point for a single series is not consistent; only the fraction $\frac{\hat{k}}{T}$ converges in probability to λ^*. Even with more observations of y_t, it is not guaranteed that the estimation of k would be improved. In addition, the effectiveness of the single-series approach critically depends on two assumptions. First, the magnitude of the parameter change after a break must be large enough. Second, the true change point k^* must lie sufficiently far from both the start and the end of the sample. However, one can take advantage of "borrowed power" by noticing the cross-sectional pattern of the structure changes. An extreme case is the one in which many different series experience a structure break at the same time, in which the accuracy of the break point estimator is greatly increased. In particular, when the number of such series N is large enough, the consistent estimation of the break point itself could be achieved. This is so-called "borrowing power".

For a multivariate time series setup, the common-break model is well studied by Bai, Lumsdaine and Stock (1998). Bai (2006) studies the common-break model in a panel data framework, and proves that the consistency for the break-point estimate itself instead of the fraction is achieved. That paper also provides the convergence rate and limiting distribution. It has important potential applications in various economic research areas. The stock market saw many different equities' prices drop or soar together within one day or a short period in anticipation of the same event. Developed countries experienced the Great Moderation in the growth rate volatility around the same year. The behavior of a large number of consumers changed suddenly after a new tax return policy was announced (an example provided in Levitt and

Dubner, 2005). A group of patients' physical conditions change after taking a new treatment.

It is worth noting that it may sound restrictive to assume that different series are subject to a structural break at the same time. However, if we take a second look at the literature, we may find that many change-point studies focus on a single aggregate index. For instance, a lot of empirical work is dedicated to modeling the stock index and estimating the change point. The index is a weighted average of individual stock prices. By analyzing structural breaks in this way, it has been implicitly assumed that different stock prices are subject to a structural break at the same time. A vast majority of the macroeconomic literature studies the Great Moderation, or the volatility decline observed in output growth, employment growth and inflation. To identify the date when the Great Moderation occured, researchers generally use the country-level output data, which is a simple aggregate of the state-level outputs. Thus the common structural break for different states is implicitly assumed in these studies. Compared to the single-series method, the common-break method greatly improves the estimation precision of change points.

The larger the number of series, the more accurate the estimation. For models with mean breaks, Monte Carlo results can be found in Bai (2006). I use Monte Carlo simulations to show that one can still gain such a benefit for complex models like GARCH, by using multiple series to estimate the change points. The change-point analysis for GARCH models has always been a challenge in the literature, due to the difficulties of precise estimation of GARCH parameters. Another reason is that a CUSUM approach is generally used, which requires that unconditional volatilities change substantially after the break. The Bayesian method is a likelihood approach, and thus does not require restrictions on unconditional volatilities across structural regimes. Assume the GARCH(1,1) model is given by

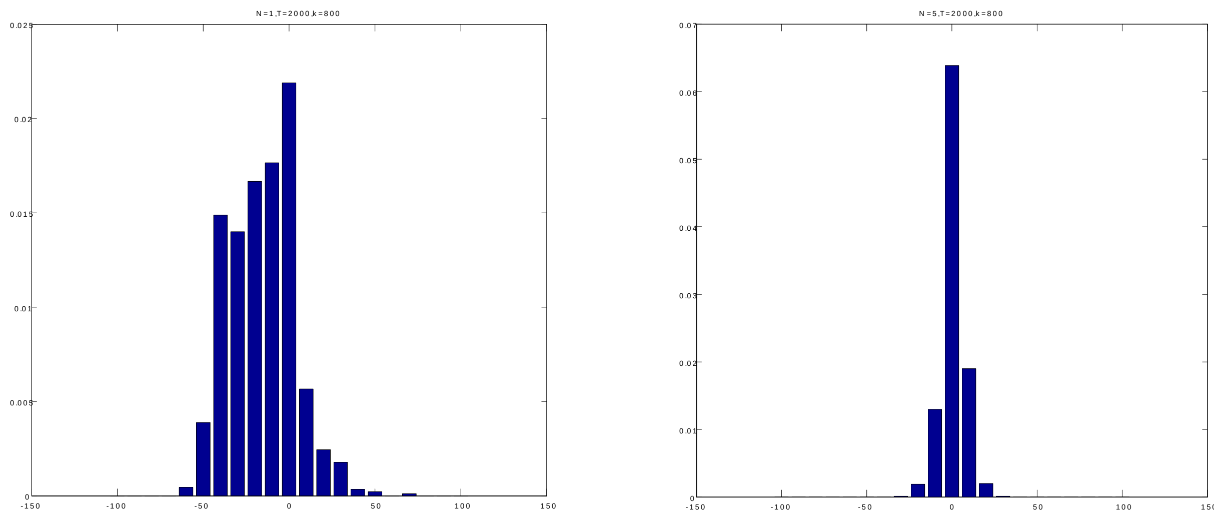
$$
\begin{aligned}
y_t &= \sqrt{h_t}\varepsilon_t, \text{ where } \varepsilon_t \sim i.i.d.N(0,1), t=1,...,T, \qquad (1)\\
h_t &= \begin{cases} \omega_1+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1}, \ t \le k^* \\ \omega_2+\alpha_2 y_{t-1}^2+\beta_2 h_{t-1}, \ t > k^* \end{cases}.
\end{aligned}
$$

where $T=2000$, $k^*=800$. Let

$$
\begin{aligned}
(\omega_1,\alpha_1,\beta_1) &= (0.10, 0.10, 0.80),\\
(\omega_2,\alpha_2,\beta_2) &= (0.15, 0.60, 0.35).
\end{aligned}
$$

I make 500 MCMC draws to calculate the posterior distribution of change points.

Figure 1. Left Panel: Bayesian Posterior Distribution for $\hat{k} - k_0$ when only using single series; Right Panel: Bayesian Posterior Distribution for $\hat{k} - k_0$, $N = 5$.



More generally, the model with M structural breaks is

$$\begin{aligned} \{y_t^{(i)}\}, t = 1, ..., T,\ i = 1, 2, ...N \\ y_t^{(i)} \sim F(\cdot, \theta_m),\ \text{for } t \in (k_{m-1}, k_m] \end{aligned} \tag{2}$$

The only link between different series is that they are subject to structural breaks at the same time, $\{k_m\}_{m=1,...M}$.

Again, let us assume that the prior of $\{k_m\}$ is jointly uniform distribution on $[2, T-1]$ (discrete version). Given other change point k_{-m} and model parameters $\{\theta_m\}_{m=1,...M}$, the conditional posterior density for the break point k_m follows the form

$$p(k_m = t | y, \theta, k_{-m}) = \frac{\prod_{i=1}^{N} l(\theta | y, k_{-m}, k_m = t)}{\sum_{s=2}^{T} \{ \prod_{i=1}^{N} l(\theta | y, k_{-m}, k_m = s)} \tag{3}$$

3 Nonparametric Approach to Change-Point Estimation

The common-break assumption provides a prospective for estimating structural breaks using panel data. However, it is generally not realistic to assume that different series change structure at exactly the same time. Individual series may experience the structural break at different time even after attack of the same event, due to the cross-sectional heterogeneity in the response time. Therefore, it is crucial to find a way to incorporate such a heterogeneity, while retaining the common feature of structural changes across different series. This will not only benefit the estimation but can also provide useful information about the cross-sectional pattern of structural changes. Only by taking into account both the similarity and the heterogeneity across sections, is one able to answer important questions such as how stocks of different industries respond to a new policy or how different experimental groups respond to a new treatment.

In this section, I relax the common-break assumption, by properly building a link for structural changes across different series. My approach provides a general framework to handle structural-break problem in panel data, allowing different series to experience a structural break at different time.

This new method takes into account two key factors. First, the different series are subject to common shocks, which are assumed to be the source of the structural changes. Second, various series exhibit heterogeneous responses to such common shocks and the timing of structural changes can be different across sections. A key assumption is that change points for different time series follow a common distribution. The common-distribution assumption relaxes the conventional common-break restriction while still retaining the common feature of the structural changes. If there is only a single structural change for each series, a non-parametric multinomial distribution is used to model the pattern of change points across sections.

In this section, I discuss the panel data model in which each different time series has a single structural break at k_i. Now, the break points $\{k_i\}$, $i = 1, ..., N$, follow a common distribution $F(\cdot, \pi)$, where π is the parameter characterizing the CDF. Notice that $F(\cdot, \pi)$ is now part of the model, not a prior. A hierarchical layer needs to be added for the parameter π. Since change points can only be integers, $F(\cdot, \pi)$ should be a discrete distribution.

A natural way is to assume that k_i follows a multinomial distribution governed by $\pi = (\pi_t)_{t=1,2,...,T}$.

Denote

$$\pi_t = prob(k_i = t), \ \ for \ t = 1, 2, ..., T \tag{4}$$

Then $\pi = (\pi_t)_{t=1,2,...,T}$, where $\sum_{t=1}^{T} \pi_t = 1$. In practice, one does not need to assign positive probability to each time period between 1 and T, but can limit the support of $F(\cdot, \pi)$ to a much narrower interval depending on one's prior belief about when the breaks happened. For example, to study how the stocks of different industries respond to a new policy, one could only search in a short period after the policy is applied, therefore reducing the dimension of parameters to be estimated. It must be emphasized that N must be large enough to consistently estimate π. It is worth pointing out that one may not need very large N to acchieve a good estimate of π if the true distribution is very concentrated. This can be caused by strong comovement among different series, which tend to change their structure in a short period. The estimation of π will soon concentrate around that period while leaving $prob(k_i = t|Y)$ near zero for the dates outside the interval. The US state-level Great Moderation application in section 6 corroborates such an intuition.

In this setup, different series are linked by $F(\cdot, \pi)$, of which each series's break point k_i is a random realization. It is easy to incorporate more correlation between series. For instance, one can also assume that $(\theta_m^{(i)})_{i=1,...,N,m=1,...M+1}$ are random numbers generated by the same distribution, and we are able to estimate this distribution as well as $(\theta_m^{(i)})_{i=1,...,N,m=1,...M+1}$ by applying the same trick. Here I only assume that change points are subject to a common distribution while leaving other parameters independent across sections without loss of generality.

To formally state the algorithm, let us take a simple linear model as an example. Data

$$Y = \begin{pmatrix} y_{11}, y_{12}, ..., y_{1T} \\ y_{21}, y_{21}, ..., y_{2T} \\ .. \\ y_{N1}, y_{N2}, ..., y_{NT} \end{pmatrix}$$

are observations for N different time series. Series i spans from 1 to T and experiences a single structural break at time k_i. Different from the conventional common-break assumption for panel data, I allow $k_i \neq k_j$, for any $i \neq j$.

For series i,

$$\begin{aligned} y_{it} &\sim N(u_{i1}, \sigma_{i1}^2), \text{ for } t \leq k_i \\ y_{it} &\sim N(u_{i2}, \sigma_{i2}^2), \text{ for } k_i + 1 < t \leq T \end{aligned} \tag{5}$$

For this simple model, $\theta = \{u_{ij}, \sigma_{ij}^2\}_{i=1,...,N}^{t=1,...,T}$. $k = (k_1, k_2, ..., k_N)$ are treated as hidden variables and are assumed to follow a multinomial distribution with $prob(k_i = t) = \pi_t$, for $t = 1, ..., T$. Both θ and π are model parameters and need to be estimated. To facilitate the computation, I choose conjugate priors for parameters $\{u_{ij}, \sigma_{ij}^2, \pi_t\}$, $i = 1, ...N, t = 1, ...T$, and $j = 1, 2$.

Likelihood	Parameters	Prior	Prior hyperparameters	Posterior hyperparameters
Normal with known mean μ	variance σ^2	Scaled inverse-chi	v δ_0^2	$v+T$ $\frac{v\delta_0^2+\sum_{t=1}^T(y_t-\mu)^2}{v+T}$
Normal with known variance σ^2	mean μ	Normal	μ_0 σ_0^2	$\frac{\mu_0/\sigma_0^2+\sum_{t=1}^T y_t^2/\sigma^2}{1/\sigma_0^2+T/\sigma^2}$ $(1/\sigma_0^2+T/\sigma^2)^{-1}$
Multinomial	probability π	Dirichlet	α_0 $(T \times 1)$	$\alpha_t = \alpha_{0,t} +$ $\sum_{i=1}^N I\{k_i = t\}$, for $t = 1, ..., T$

Here I use the Dirichlet distribution as the prior for π

$$\pi = (\pi_1, \pi_2, ..., \pi_T) \sim Dirichlet(\alpha) \tag{6}$$
$$\text{where } \alpha = (\alpha_1, ..., \alpha_T)$$

Thus the posterior conditional distribution for π is

$$\pi | k \sim Dirichlet(\alpha')$$
$$\text{where } \alpha'_t = \alpha_t + \sum_{i=1}^{N} I_{\{k_i^{(1)}=t\}}$$

$\alpha = (\alpha_1, ..., \alpha_T)$ is called hyperparameter. For instance, if we set $\alpha_1 = ... = \alpha_T = 1$, then equal weights are imposed on each period.

The Gibbs sampling procedure is given as follows. To save notation, a variable with an arrow on the top means a vector consisting of all elements of that variable. For example, $\overrightarrow{k}^{(0)} = (k_1^{(0)}, ..., k_N^{(0)})$.

Step 1: Given initial values $\left(\mu_{ij}^{(0)}, \sigma_{ij}^{2\,(0)}, k_i^{(0)}, \pi_t^{(0)}\right)$, $i = 1, 2, ...N$, $j = 1, 2$, $t = 1, 2, .., T$

Step 2: Update parameter μ_{ij}

Given the initial values of other parameters $\left(\mu_{ij}^{(0)}, \sigma_{ij}^{2\,(0)}, k_i^{(0)}, \pi_t^{(0)}\right)$, $i = 1, 2, ...N$, $j = 1, 2$, $t = 1, 2, .., T$ and data Y, the posterior conditional distribution of $\mu_{ij}^{(1)}$ is

$$
\begin{aligned}
&\mu_{i1}^{(1)} | \overrightarrow{\sigma}^{2\,(0)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
&\sim Normal\left(\left(\frac{\mu_{i10}}{\sigma_{i10}^2} + \frac{\sum_{t=1}^{k_i^{(0)}} {y_t}^2}{\sigma_{i1}^{2\,(0)}}\right) / \left(\frac{1}{\sigma_{i10}^2} + \frac{k_i^{(0)}}{\sigma_{i1}^{2\,(0)}}\right), \left(\frac{1}{\sigma_{i10}^2} + \frac{k_i^{(0)}}{\sigma_{i1}^{2\,(0)}}\right)^{-1}\right) \\
&\mu_{i2}^{(1)} | \overrightarrow{\sigma}^{2\,(0)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
&\sim Normal\left(\left(\frac{\mu_{i20}}{\sigma_{i20}^2} + \frac{\sum_{t=k_i^{(0)}+1}^{T} {y_t}^2}{\sigma_{i2}^{2\,(0)}}\right) / \left(\frac{1}{\sigma_{i20}^2} + \frac{T - k_i^{(0)}}{\sigma_{i2}^{2\,(0)}}\right), \left(\frac{1}{\sigma_{i20}^2} + \frac{T - k_i^{(0)}}{\sigma_{i2}^{2\,(0)}}\right)^{-1}\right)
\end{aligned}
\tag{7}
$$

Step 3: Update parameter σ_{ij}^2

Given $\left(\mu_{ij}^{(1)}, k_i^{(0)}, \pi_t^{(0)}, X\right)$, $i = 1, 2, ...N$, $j = 1, 2$, $t = 1, 2, .., T$, $\sigma_{ij}^{2\,(1)}$ has the following posterior conditional distribution

$$
\begin{aligned}
&\sigma_{i1}^{2\,(1)} | \overrightarrow{\mu}^{(1)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
\sim\ &\text{Scaled inverse-chi-square}\left(\nu_{i1} + k_i^{(0)}, \frac{\nu_{i1}\delta_{i10}^2 + \sum_{t=1}^{k_i^{(0)}} (y_{it} - \mu_{i1}^{(1)})^2}{\nu_{i1} + k_i^{(0)}}\right) \\
&\sigma_{i2}^{2\,(1)} | \overrightarrow{\mu}^{(1)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
\sim\ &\text{Scaled inverse-chi-square}\left(\nu_{i2} + T - k_i^{(0)}, \frac{\nu_{i2}\delta_{i20}^2 + \sum_{t=k_i^{(0)}+1}^{T} (y_{it} - \mu_{i2}^{(1)})^2}{\nu_{i2} + T - k_i^{(0)}}\right)
\end{aligned}
\tag{8}
$$

Step 3: Update parameter $k_i^{(1)}$

Draw $k_i^{(1)}$, given $\left(\mu_{ij}^{(1)}, \sigma_{ij}^{2\,(1)}, \pi_t^{(0)}, Y\right)$, $i = 1, 2, ...N$, $j = 1, 2$, $t = 1, 2, .., T$

$$\Pr(k_i^{(1)} = \tau | \overrightarrow{\mu}^{(1)}, \overrightarrow{\sigma}^{2\,(1)}, \overrightarrow{\pi}^{(0)}, Y) = \frac{\prod_{t=1}^{\tau} \frac{1}{\sqrt{2\pi\sigma_{i1}^{2\,(1)}}} \exp\{-\frac{(y_{it}-\mu_{i1}^{(1)})^2}{2\sigma_{i1}^{2\,(1)}}\} \prod_{k=t+1}^{T} \frac{1}{\sqrt{2\pi\sigma_{i2}^{2\,(1)}}} \exp\{-\frac{(y_{it}-\mu_{i2}^{(1)})^2}{2\sigma_{i2}^{2\,(1)}}\} \pi_t^{(0)}}{\sum_{t=1}^{T} (\prod_{k=1}^{t} \frac{1}{\sqrt{2\pi\sigma_{i1}^{2\,(1)}}} \exp\{-\frac{(x_{it}-\mu_{i1}^{(1)})^2}{2\sigma_{i1}^{2\,(1)}}\} \prod_{k=t+1}^{T} \frac{1}{\sqrt{2\pi\sigma_{i2}^{2\,(1)}}} \exp\{-\frac{(x_{it}-\mu_{i2}^{(1)})^2}{2\sigma_{i2}^{2\,(1)}}\} \pi_t^{(0)})} \tag{9}$$

Step 4: Update parameter π_t

Draw $\pi_t^{(1)}$, given $\left(\mu_{ij}^{(1)}, \sigma_{ij}^{2\,(1)}, k_i^{(1)}, Y\right)$, $i = 1, 2, ...N,\ \ j = 1, 2,\ \ t = 1, 2, .., T$

$$\overrightarrow{\pi}^{(1)} | \overrightarrow{k}^{(1)} \sim Dirichlet(\overrightarrow{\alpha}') \tag{10}$$

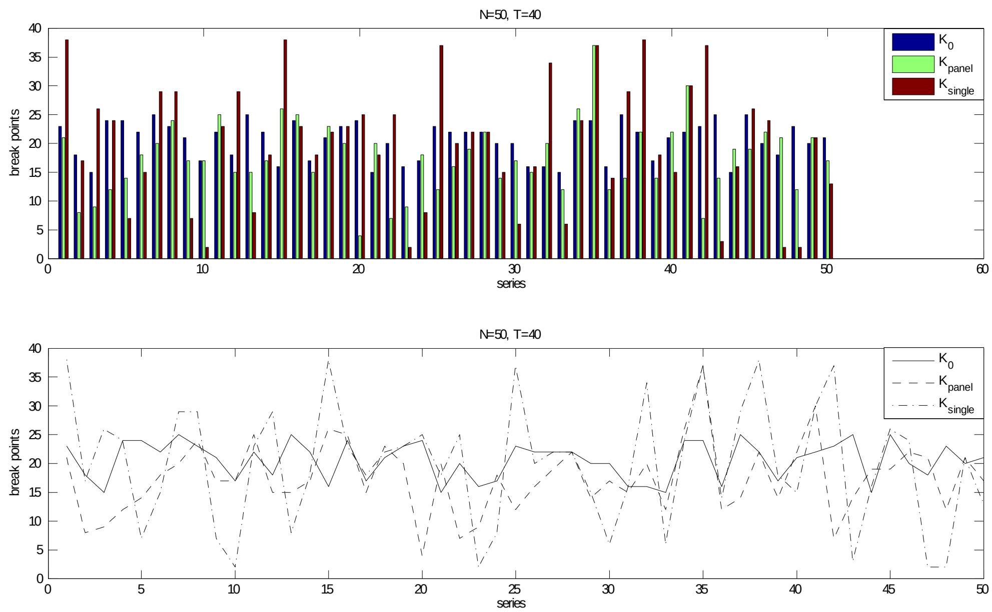
$$\text{where } \alpha'_\tau = \alpha_\tau + \sum_{i=1}^{N} I_{\{k_i^{(1)} = \tau\}}$$

To get random draws of $\{\pi_t\}$, one needs to draw Z_τ from $Gamma(\alpha'_\tau, 1)$, for each $\tau = 1, 2, ...T$ (or through the support of $F(\cdot, \pi)$ if one restricts the break points to a narrower interval). Then update $\pi_t = Z_t / \sum_{\tau=1}^{T} Z_\tau$.

It is worth mentioning that the existence of conjugate priors is not essential for the common distribution method to work. For more complicated models, which might not have conjugate priors, one can still use either Metropolis-Hastings sampler or Griddy-Gibbs sampler to generate random draws from the joint posterior distribution.

Monte Carlo simulations show that such a panel data method is able to improve the quality of estimates. Data is generated by model (5). In the rest of this paper, the point estimator $\hat{k}$ of change point is calculated as the posterior mean rounded to the nearest integer. In the lower panel of Figure 2, the solid line is the true value of change points for 50 time series, and the dashed line describes the change-point estimates using the non-parametric panel data model, while the dash-dotted line shows $\hat{k}_{\text{single}}$ by estimating univariate model series by series. $N = 50$, $T = 40$, and the true change points are independent random draws from uniform distribution over $[15, 25]$. $\hat{k}_{panel}$ is much closer to the true change point k_0 for most series than $\hat{k}_{\text{single.}}$

Figure 2. Panel Data V.S. Single series



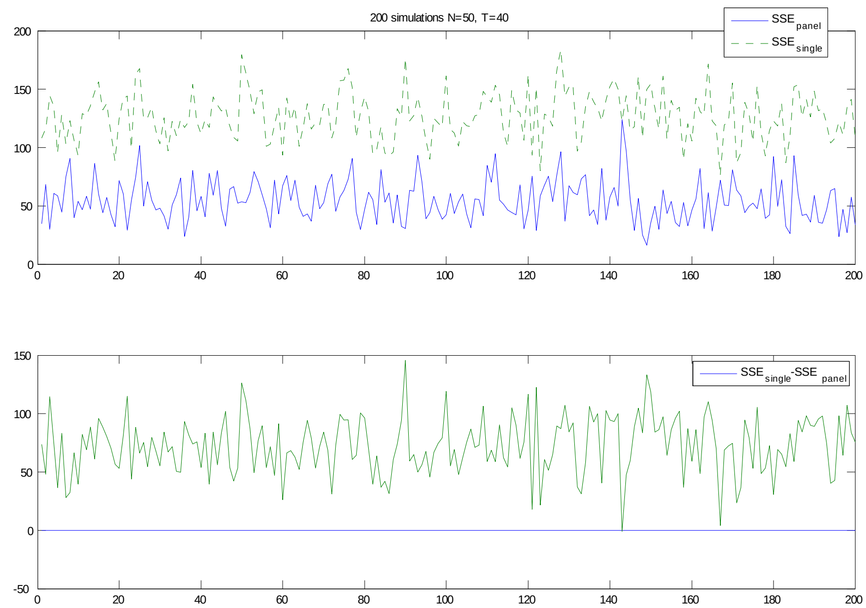
To check the robustness of the above result, I simulate 200 independent samples, each containing 50 time series and spanning 50 periods. The true change points are independent random draws from uniform distribution over $[15, 25]$. For each sample, I compute the sum of squared errors using the following equation,

$$SSE^N = \frac{1}{N}\sum_{i=1}^{N}(\hat{k}_i - k_i^*)^2,$$

where k_i^* is the true change point.

In Figure 3, I plot the SSE of estimation result for both panel data model and single series model in the upper panel. In the lower panel, the difference between $SSE_{\text{single}} - SSE_{panel}$ is plotted against zero line, thus the positive numbers represent the gain in precision by using panel data model. The graph shows that the panel data model uniformly outperforms the single series model in terms of significant reduction in estimation errors for all 200 independent samples.

Figure 3. 200 Simulations



4 A Parametric Method to Change-Point Estimation

When the number of series N is large but the time length T is not very large, the nonparametric method described in the previous section works very well and gives consistent estimation of both the model parameters and the distribution followed by the break points $k = \{k_i, i = 1, ..., N\}$. However, the estimation error would become big when N is small and T is large, because the large number of additional parameters $\pi = (\pi_1, \pi_2, ..., \pi_T)$.

Remark 1 *When using the nonparametric method of section 3, one may not need a very large N to achieve a good estimation of π in two cases. One case is that the true distribution of change points is very concentrated. This can be due to strong comovement among different series, and all the series tend to change their structure around the same period. The estimation of π will soon concentrate around that period while leaving $prob(k_i = t|Y)$ near zero for the dates outside a short interval. The US state-level Great Moderation application in section 6 corroborates such a intuition.*

Another case is that the magnitude of structural break is very large. Large break magnitudes will facilitate the detection of the change points, despite the large dimension of π.

To deal with the structural problem in this case, one needs to develop a parsimonious way to model the distribution of break points $k = \{k_i, i = 1, ..., N\}$, and the most vital part is to reduce the parameter dimension of π, which characterizes the distribution $F(\cdot, \pi)$. The choice of $F(\cdot, \pi)$ must satisfy the following properties:

1) The dimension of π is low;

2) $F(\cdot, \pi)$ has support on $[0, +\infty)$;

3) $F(\cdot, \pi)$ is a discrete distribution.

4) $F(\cdot, \pi)$ has a shape flexible enough since the structural break could happen anytime between period 1 and T.

5) It is easy to compute the posterior density.

The *Poisson* distribution is an ideal candidate. It is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time, provided these events occur with a known average rate and independently of the time since the last event, and it is governed by only one parameter λ, the expected number of occurrences during the given interval. The Gamma distribution is its conjugate prior, thus the posterior density is easy to achieve.

Koop and Potter(2004) use the *Poisson* distribution to model the duration of regimes for a single series model with structural breaks. The regime duration is just a reparametrization of the change points. When there exists only one structural break, the duration of the first regime is equal to the value of the change point. However, their work is based on the state-space model initialized by Chib (1998), and needs to make inference about the hidden states and transition matrix. Unlike Chib (1998), they allow the transition matrix to vary along with the duration of the most recent regime, which greatly complicates the computation. As they recognize in the paper, calculating the transition probability matrix will involve $O(T^3)$ calculations.

There are also other advantages of the *Poisson* distribution. First, k_i is not restricted between 1 and T since the Poisson distribution allows for $k_i = 0$ and $k_i \geq T$. This way, the series is allowed to have no structural break. Second, forecasting is feasible. For instance, it is reasonable to suppose that foreign countries' economy condition can shed some light on the home country's future economic performance.

The model is specified as

$$\begin{aligned}
Y &= \{y_{it}\}_{t=1,\dots,T}^{i=1,\dots,N} \sim f(Y|k,\theta) \\
\theta &= \{\theta_i^j\}_{i=1,\dots,N}^{j=1,2}, \text{ and } \theta_i = \begin{cases} \theta_i^1, \text{ for } t \le k_i \\ \theta_i^2, \text{ for } k_i < t \le T \end{cases} \\
k &= \{k_i, i = 1, \dots, N\}, \; k_i \sim Poisson(\lambda)
\end{aligned} \tag{11}$$

Again, k is treated as missing data or latent variables here. Because one does not observe k, inference of k is needed, which can be achieved by studying its posterior distribution. This is made possible by using Gibbs sampling. It draws the random samples from the marginal posteriors $p(\lambda|Y,k,\theta)$, $p(\theta|Y,k,\lambda)$ and $p(k|Y,\theta,\lambda)$ instead of from the joint posterior $p(\lambda,\theta,k|Y)$. An important assumption is $p(\lambda|k,\theta,Y) = p(\lambda|k)$. The procedure is given as follows.

Step 1. Specify priors for parameters λ and θ. The choice of priors for θ has been described previously. For convenience, I choose $Gamma(a,b)$ as prior for λ, where a and b are prespecified constant, say 2 and 1.

Step 2. Given observed data Y, and initial values of $k = \{k_i, i = 1, \dots, N\}$ *and* λ, make random draws of parameters θ from their posterior density. This step is the same as the procedure described in previous sections. Use Griddy-Gibbs method.

Step 3. Given updated value of θ,and initial value of λ,

$$k \sim f(k|\theta,\lambda,Y) \propto f(Y|\theta,k,\lambda)p(k|\theta,\lambda) \propto f(Y|\theta,k)p(k|\lambda) \tag{12}$$

In this example, for $i = 1, \dots, N$,

$$p(k_i = t|y,\theta,k_{-i}) = \frac{l(\lambda|Y,k_i = t,k_{-i}) \cdot p(k_i = t|\lambda)}{\sum_{s=1}^{T} l(\lambda|Y,k_i = s,k_{-i}) \cdot p(k_i = s|\lambda))} \tag{13}$$

Step 4. Given updated θ and $k = \{k_i, i = 1, \dots, N\}$, $\lambda \sim f(\lambda|\theta,k,Y) = p(\lambda|k)$ by assumption, which is $Gamma(a + \sum_{i=1}^{N} k_i, \left(N + \frac{1}{b}\right)^{-1})$.

Remark 2 *Although the Poisson distribution greatly reduces the dimension of the parameter vector, it has a bell-shaped density function. It is a good approximation if the true change points exhibit a unimodal pattern, i.e. a majority of series experience the structural changes on a tight time interval. This is a reasonable assumption when different series change structure due to a common shock and the change points* *are clustered. Otherwise, one needs to assume a more flexible shape of the common distribution for the change points to allow for a more flexible pattern. The mixture normal distribution is a good candidate if more flexibility is desired.*

5 The Nonreversible Hidden Markov Chain in Panel Data Models

5.1 The Modeling of Multiple Change Points in Panel Data

In addition to the difficulty of estimation when T is large but N is small, another problem arises when one further allows each series to have multiple structural changes. For the panel data

$$Y = \begin{pmatrix} y_{11}, y_{12}, \dots, y_{1T} \\ y_{21}, y_{21}, \dots, y_{2T} \\ \cdot\cdot \\ y_{N1}, y_{N2}, \dots, y_{NT} \end{pmatrix},$$

assume that each process y_{it} experiences M structural breaks during $[1, T]$. The change points for series i are $\{k_m^{(i)}\}_{m=1,\dots M}$, $i = 1, \dots N$. We need to make assumptions about the joint distribution of $\{k_m^{(i)}\}$ whenever the number of structural breaks is larger than 1. For illustration purpose, let us assume that each series y_{it} has two breaks at $k_1^{(i)}$ and $k_2^{(i)}$. The joint density of $k_1^{(i)}$ and $k_2^{(i)}$, $F(k_1^{(i)}, k_2^{(i)})$ is assumed to be the same across series, thus I drop the superscript i from the notation when the context is clear in the rest of the paper.

$$F(k_1, k_2) = f(k_2|k_1) \cdot f(k_1)$$

If one sticks to the nonparametric method, one needs to specify the probability for each possible pair of (k_1, k_2) on the whole interval $[1, T]$,

$$\pi_{t,\tau} = prob(k_1 = t_1,\ k_2 = t_2),\ \text{for any } 1 \le t_1 < t_2 \le T \tag{14}$$

which makes the dimension of vector π become $T \cdot (T-1)/2$. Likewise, when the number of breaks M increases, the dimension of the vector π also increases accordingly to C_T^M, which is of the order of T^M. Even with moderately large T, there would be too many parameters in the model and precise estimation would be a challenge to achieve.

Hence one needs to add some restriction to the form of $F(k_1, k_2, ...k_M)$ to reduce the number of parameters. In this section, I introduce a special form of common distribution $F(k_1, k_2, ...k_M)$ which proves to be a powerful tool to deal with multiple-break problems. It is built on the structural break model proposed by Chib (1998) which has been widely used in many applications, e.g., Nelson and Piger (2002), and Pesaran, Pettenuzzo and Timmermann (2006). However, all the existing works study the single series case.

As we willshortly see, the nonreversible hidden Markov chain model is actually a special form of the common-distribution model in the univariate case. Realizing this link, I extend the original univariate model to a multivariate data environment. With my modification, it becomes a powerful tool to estimate multiple change points in panel data.

It is worthwhile to begin by reviewing the influential structural break model for a single time series by Chib (1998), which provides an efficient way to estimate a univariate model with multiple structural breaks. The observed data is $\{y_t\}_{t=1,...T}$. If there are M structural changes during $[1, T]$, there would be $M + 1$ different regimes. Assume that $y_t \sim F(\cdot, \theta)$, while $\theta = \theta_m$ in the mth regime. Let $S_T = (s_{1,}\ s_{2,\ ...,}\ s_T)$, and s_t is a latent discrete state variable which takes values on $\{1, 2, ...M + 1\}$. The $\{s_t\}$ follows a Markov chain and the transition matrix P is given by

$$P = \begin{pmatrix} p_{11} & p_{12} & 0 & .... & \\ 0 & p_{22} & p_{23} & 0 & ... \\ ... & & & & \\ & & & p_{M,M} & p_{M,M+1} \\ 0 & ... & & & 1 \end{pmatrix} \tag{15}$$

where $p_{ii} = prob(s_t = i|s_{t-1} = i)$, and $p_{i,i+1} = prob(s_t = i + 1|s_{t-1} = i)$. Although this method does not explicitly model the change points $\{k_m\}_{m=1,...M}$, the estimation of k_m is obtained by counting the length of regime m. The structural change occurs when s_t jumps one step ahead, thus

$$k_m = \{t : s_{t+1} = m + 1, s_t = m\} \text{ for } m = 1, 2, ...M \tag{16}$$

Different from the conventional Markov switching model, the Markov chain is nonreversible, since s_{t+1} cannot jump backwards to $m-1$ once $s_t = m$, and s_t can only move forward step-by-step without skipping. This is an innovative way to look at the change-point problem and build the bridge between Markov switching literature and the change-point literature. It is particularly useful when the structural change actually does not have a recurring pattern, as is assumed in Markov switching literature.

To understand the link between the common distribution model and the nonreversible hidden Markov chain model, let us assume a single structural break for each series. Thus the transition matrix P is

$$P = \begin{pmatrix} p_{11} & 1-p_{11} \\ 0 & 1 \end{pmatrix}$$

The hidden states $\{s_t\}$ will only take value on $\{1, 2\}$.

The nonparametric method does not make any special assumption about $prob(k = t)$ and let $\pi_t = prob(k_i = t)$, for $t = 1, 2, ..., T$ to be the model parameters with $\sum_{t=1}^{T} \pi_t = 1$. If the nonreversible Markov chain structure is assumed, then

$$\pi_t = prob(k = t) = prob(s_t = 1, and\ s_{t+1} = 2) = p_{11}^t \cdot (1 - p_{11}) \tag{17}$$

The transition matrix P summarizes all the information about the distribution of change points. Each π_t now is a function of a single parameter p_{11}. The nonreversible hidden Markov chain assumption is actually a special form of the nonparametric distribution $F(k)$. This restriction greatly reduces the dimension of the parameter vector π (in the case of two structural breaks, from $T \cdot (T-1)/2$ to 2). Likewise, for multiple change-points problem, the transition matrix P governs the joint distribution $F(k_1, k_2, ...k_M)$, and $prob(k_1 = t_1, k_2 = t_2, ..., k_M = t_M)$ for any $1 \leq t_1 < t_2 < ... < t_M \leq T$ becomes a function of P. Instead of estimating π of C_T^M dimension, now only M parameters $\{p_{mm}\}_{m=1,...M}$ need to be estimated.

In the multivariate context, we modify the univariate nonreversible hidden Markov chain model by assuming that different series share the same transition matrix P. While the dimension of model parameters is significantly lowered, the nonreversible hidden Markov chain model still retains enough flexibility to allow each series to evolve in different ways over time, and thus experience structural change at different times. The algorithm for multivariate Markov chain is described in detail in the Appendix.

5.2 Monte Carlo simulations

To show the advantage of taking into account panel data information, I simulate 10 different series. $y_t = \mu + \sigma_t \cdot e_{t,}$ where $e_t \sim i.i.d.\ Normal(0,1)$, and assume that the mean μ does not vary across regimes. If there is also a structural break in the mean, it is even easier to detect the change points. Only variance σ_t^2 reduces 20% from regime 1 to regime 2. Also, different series changes regimes at different time with 2 periods' lag between two consecutive series. Figure 4 plots a typical realization for $\{y_{i,t}\}_{t=1,...T}^{i=1,...N}$. For instance, the first series changes regime at $t = 100$, the second one changes at $t = 102$, etc. In the following figure, I plot those series. For some series, the structural break is not that obvious.

Figure 5 shows that the change-point estimation for both multivariate analysis and single-series estimation. The latter treats individual series independently and does not consider the clustering pattern of the change points. Overall, the method assuming common transition matrix across series outperforms the single-series method.

Figure 4.

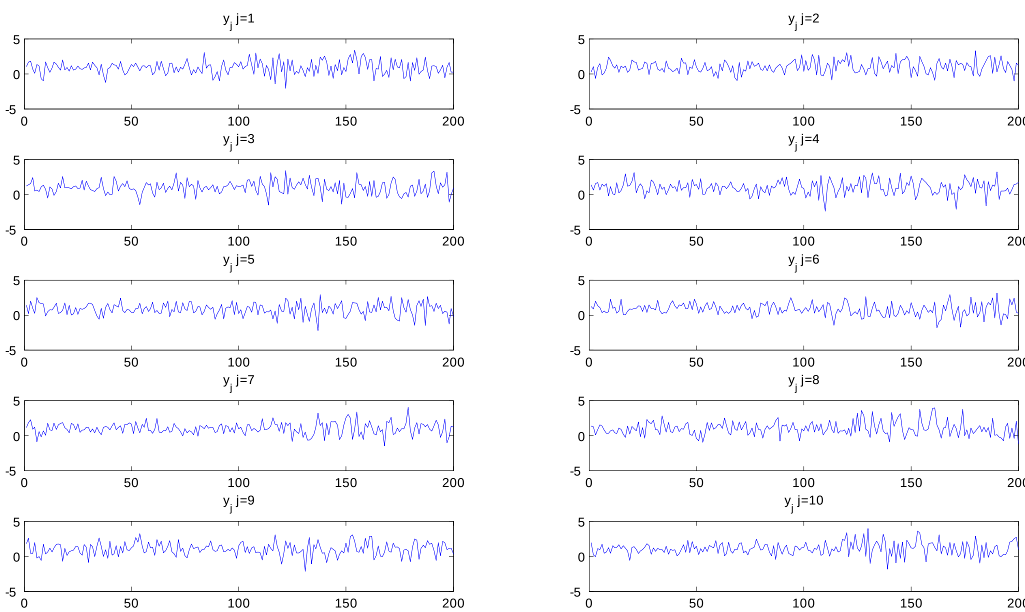
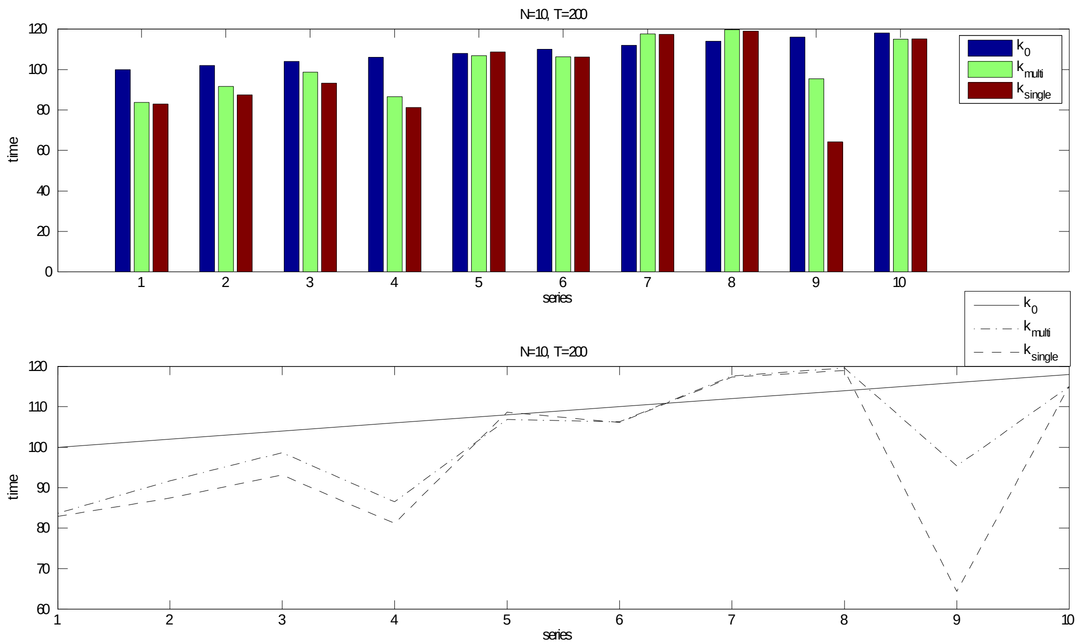


Figure 5.



To assess the robustness of the result, I simulate 200 independent random samples. Each sample contains N different time series, and each series has one structural break during $[1, T]$. Different series have different change points. We assume that the true change points k_i^* are uniformly distributed over $[a, b]$ independently, for $i = 1, 2, ...N$, where $1 < a < b < T$. For each sample, we estimate the change points $\{\hat{k}_{i,multi}\}_{i=1,...N}$ using the multivariate algorithm (panel data analysis) as well as $\{\hat{k}_{i,\text{single}}\}_{i=1,...N}$ by univariate method (series by series analysis) as a benchmark model. The sum of squared errors is computed as

$$
\begin{aligned}
SSE_{multi}^N &= \frac{1}{N}\sum_{i=1}^{N}(\hat{k}_{i,multi} - k_i^*)^2 \\
SSE_{\sin gle} &= \frac{1}{N}\sum_{i=1}^{N}(\hat{k}_{i,\text{single}} - k_i^*)^2
\end{aligned}
$$

For each sample, we compute both SSE_{multi} and SSE_{single}, and I plot the SSE_{multi} and SSE_{single} for the 200 independent samples in the following graphs. The Monte Carlo simulations are performed in the following procedure:

Step 1: Simulate N time series assuming that $\{k_i^*\}_{i=1,...N}$ are i.i.d uniformly distributed over $[a, b]$.

Step 2: Estimate the change points using multivariate model. Compute the sum of squared errors $SSE_{multi}^N = \frac{1}{N}\sum_{i=1}^{N}(\hat{k}_{i,multi}^N - k_i^*)^2$

Step 3: Estimate the change points for each series separately using univariate model. Compute the sum of squared errors

$$SSE_{\sin gle}^N = \frac{1}{N}\sum_{i=1}^{N}(\hat{k}_{i,\text{single}} - k_i^*)^2 \tag{18}$$

Step 4: Repeat the experiment 200 times.

First, I choose $N = 5$, $T = 200$, $y_{i,t} = \begin{cases} \mu + \sigma_{i1} \cdot e_{it,} \text{ for } 1 \leq t \leq k_i^* \\ \mu + \sigma_{i2} \cdot e_{it,} \text{ for } k_i^* < t \leq T \end{cases}$, $a = T/2 = 100$, $b = 125$, thus the change points $\{k_i^*\}_{i=1,...5}$ are uniformly distributed over $[100, 125]$. σ_{i1}^2 is randomly generated from uniform $[0.5, 0.75]$, while $\sigma_{i2}^2 = 0.8 \cdot \sigma_{i1}^2$. To compare the result of multivariate model with that of the univariate method, I plot $SSE_{multi}^{N=5}$ and $SSE_{\sin gle}^{N=5}$ (upper panel) as well as the difference between them (lower panel) in Figure 6. The difference is calculated as $SSE_{\sin gle}^{N=5} - SSE_{multi}^{N=5}$, thus a positive number implies improvement in the estimation precision of multivariate model over univariate model. Figure 6 shows that the multivariate model outperforms the single series model with only a few exceptions for the 200 independent samples.

In Figure 7, I set $a = T/2 = 100$, $b = 150$, thus the change points follow a more diffuse distribution. Again, the multivariate model reduces the estimation errors for almost all simulations, despite less similarity among different series implied by the more diffuse change-point distribution.

Furthermore in Figure 8, I increase the number of series to $N = 10$, $T = 200$, $a = T/2 = 100$, $b = 150$,other parameters are generated in the same way as in the $N = 5$ case. The true distribution of change points $\{k_i^*\}_{i=1,...10}$ are now uniform$[100, 150]$. The upper panel of Figure 8 plots $SSE_{multi}^{N=10}$ and $SSE_{\sin gle}^{N=10}$, and the lower panel shows the $SSE_{\sin gle}^{N=10} - SSE_{multi}^{N=10}$, which is plotted against the zero line. The graph shows that the difference $SSE_{\sin gle}^{N=10} - SSE_{multi}^{N=10}$ is above zero for most of the samples, which implies improvement in the change-point estimation by using the multivariate model.

The extensive simulation experiments show that the multivariate model uniformly reduces the estimation errors comparing with the single series model.

Figure 6. $N = 5$, $T = 200$, $a = T/2 = 100$, $b = 125$. Multivariate V.S. univariate model

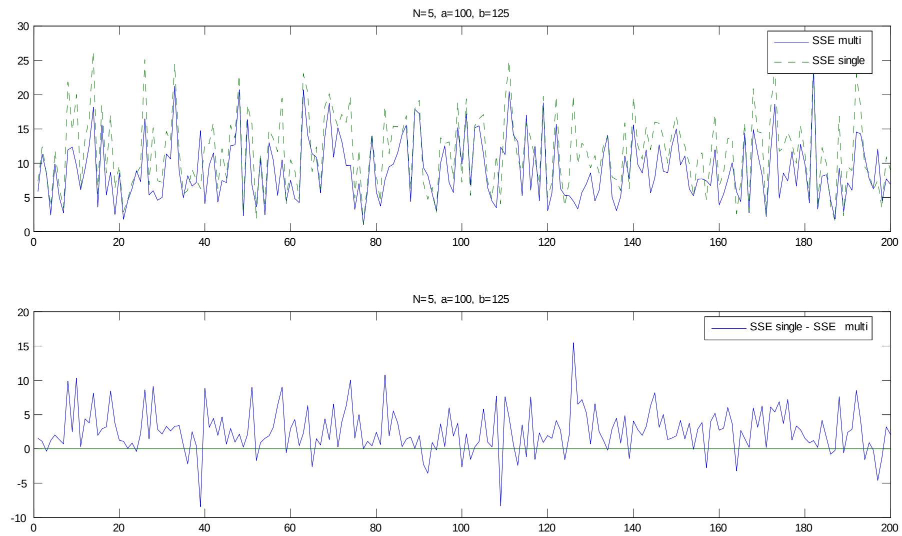


Figure 7. $N = 5$, $T = 200$, $a = T/2 = 100$, $b = 150$, Multivariate V.S. univariate model

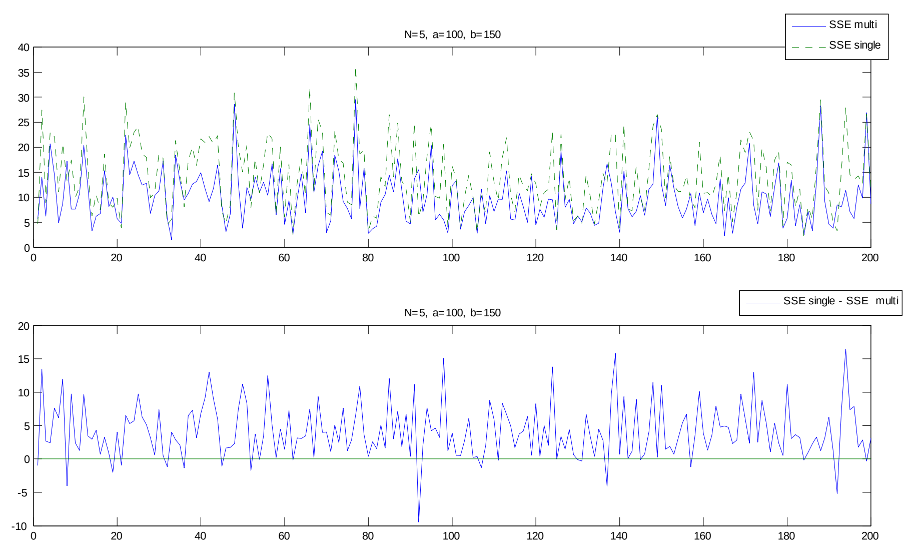
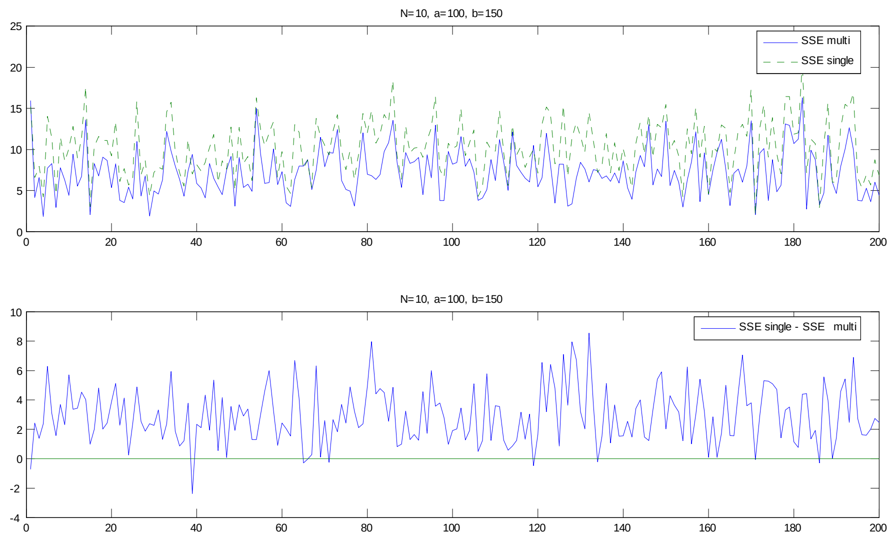


Figure 8. $N = 10$, $T = 200$, $a = T/2 = 100$, $b = 150$. Multivariate V.S. univariate model



6 Application

It is well documented that major US macroeconomic time series such as output growth and inflation have experienced a substantial reduction in volatility since the 1980s. This is commonly referred to as the Great Moderation. There has been a heated debate about the causes of the Great Moderation. Some argue that better monetary policy stabilized the economy, while others consider that improved inventory control helped stabilization.

One way to track down the source of the Great Moderation is to study the Great Moderation evidence at the state level. Different states have different economic structures and are exposed to various state-specific shocks. However, they also bear great similarity and are subject to both common nation-wide and global shocks. Knowledge about the dates of the Great Moderation for all the states is thus helpful to identify the source of it.

However, little has been done to examine whether such a Great Moderation exists in disaggregate data. A recent work by Owyang, Piger and Wall (2008) studies the state-level facts of Great Moderation in unemployment growth rate. However, they conduct the structural change analysis by univariate method, i.e., they estimate the change point series-by-series without taking into account the similarity among different states.

In this section, I use my multivariate model for change-point detection at the state level, which could exploit the cross-sectional information. By doing this, two key factors are taken care of. First, the state-level series are subject to common shocks, which are assumed to be the source of the volatility decline. Second, different states exhibit heterogeneous response to such common shocks and the timing of structural changes must be allowed to differ. By taking into account both the similarity and the heterogeneity across states, one is able to answer important questions such as how different states respond to a new policy or an economic shock.

6.1 Data description

I use 50 US state-level quarterly personal income data, spanning from 1950Q1 to 2008Q2. Louisiana is excluded due to the huge impact of the Hurricane Katrina, which was a major disaster for the local economy. The data is published by the Bureau of Economic Analysis. The growth rate is computed as the log difference

$$y_{it} = \log(\text{personal income}_{it}) - \log(\text{personal income}_{it-1}) \tag{19}$$

Each state has 233 observations, starting from the second quarter of 1950 and ending at the second quarter of 2008, thus $N = 50$, and $T = 233$ in my model.

The data $\{y_{it}\}_{t=1,\dots,T}^{i=1,\dots,N}$ are plotted below.

Figure 9. State-level personal income, Part 1

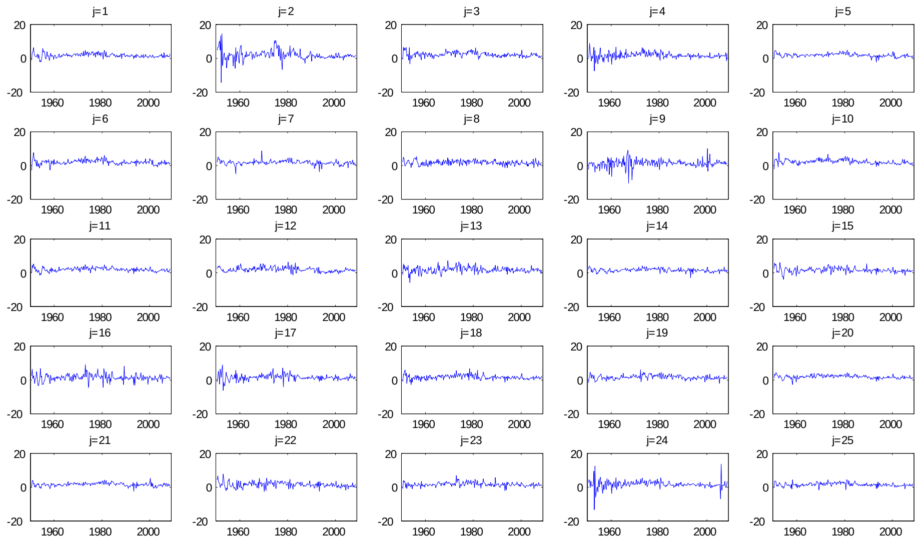
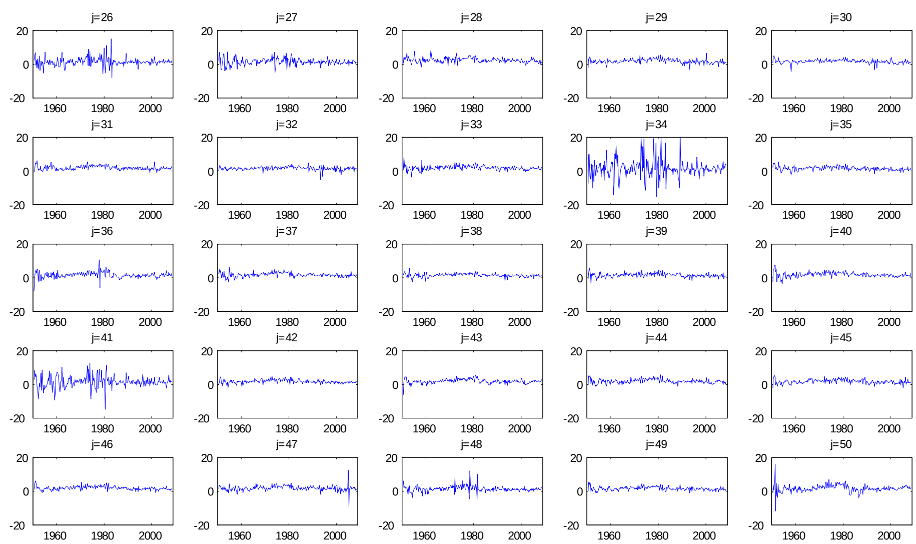


Figure 10. State-level personal income, Part 2



6.2 Estimation Results

The cross-section distribution of break dates shows two peaks, one near 1984 and the other around 1988. Only several states have no evidence of structural breaks in the volatility.

Figure 11. The estimated distribution of change points

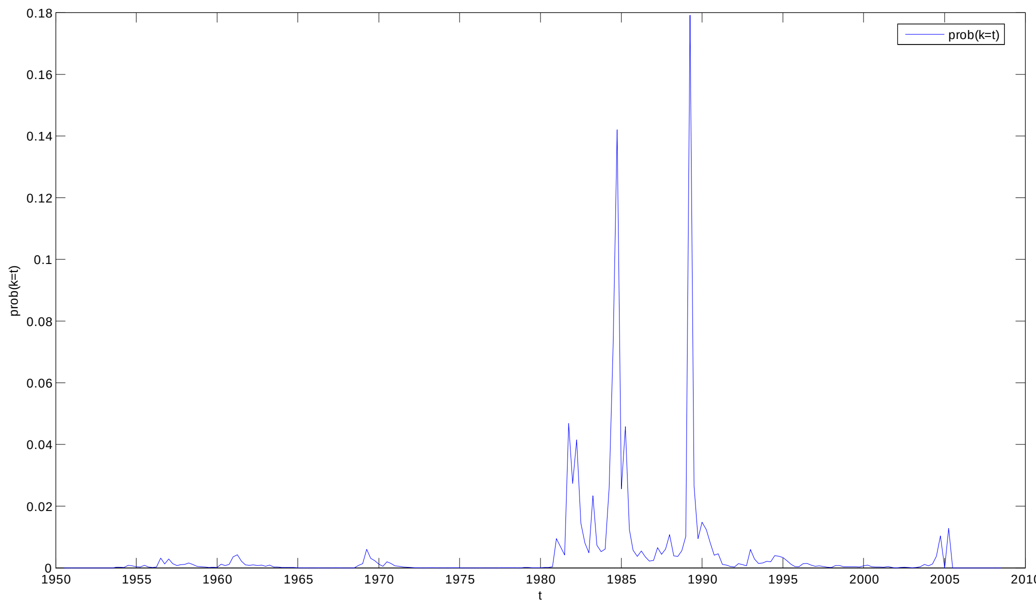
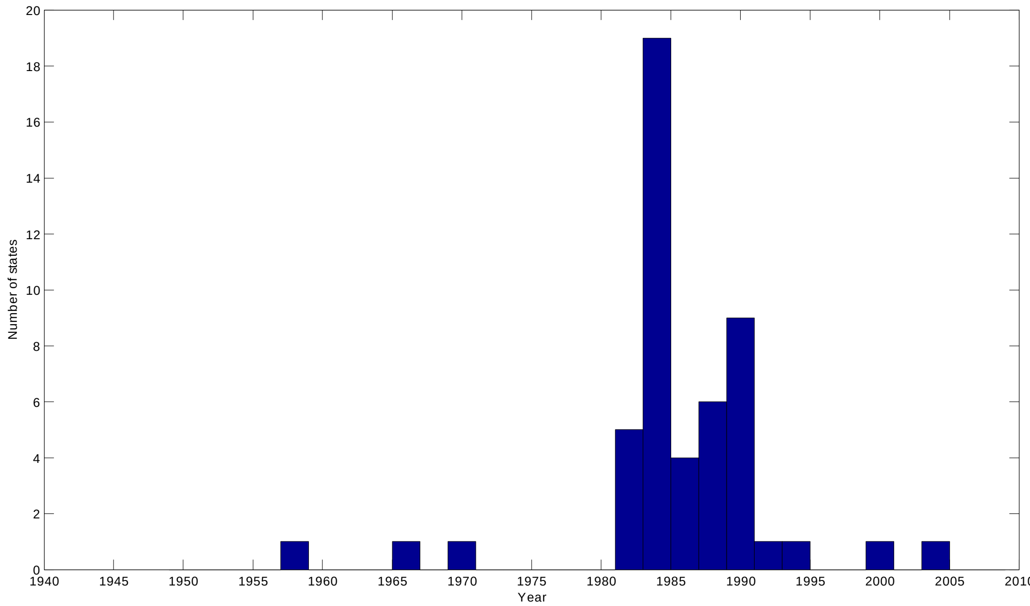


Figure 12. The histogram for estimated change points using panel data model



Estimating the change-point state by state

Figure 13.

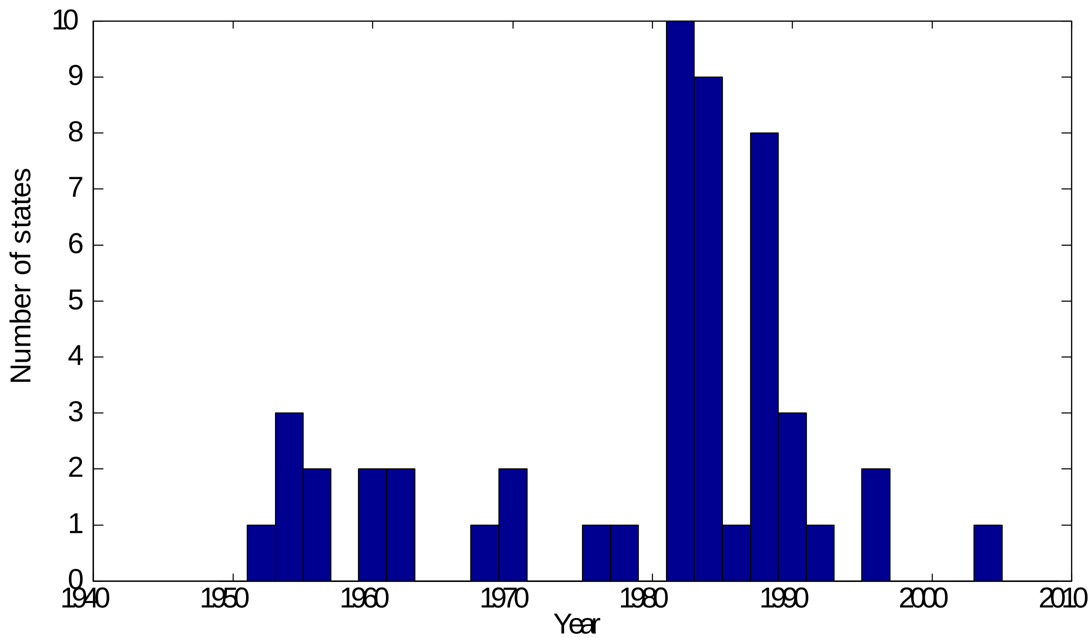


Table 1 to 7 report the estimated break dates with their standard errors for each states.

Table1.

State	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware
Panel	1984Q3	1988Q2	1985Q1	1984Q1	1988Q4	1982Q3	1988Q4	1983Q1
	10.777	9.3299	3.4564	4.2217	7.4622	4.61	6.1603	42.789
Single	1982Q4	1988Q2	1981Q1	1983Q4	1989Q1	1962Q4	1987Q3	1968Q3
	19.87584	8.969675	42.9066	4.74224	7.320957	52.57815	30.84343	57.03934

Table2.

State	District of Columbia	Florida	Georgia	Hawaii	Idaho	Illinois	Indiana	Iowa
Panel	1969Q3	1988Q3	1985Q3	1991Q3	1983Q1	1987Q2	1984Q4	2000Q3
	3.0899	6.4854	5.9694	9.1404	5.0146	18.713	9.5002	25.72
Single	1969Q3	1988Q2	1981Q2	1992Q1	1982Q4	1975Q3	1977Q4	1996Q3
	2.972544	7.38009	42.86543	6.91422	5.01432	71.15429	47.62047	19.04184

Table 3.

State	Kansas	Kentucky	Maine	Maryland	Massachusetts	Michigan	Minnesota	Mississippi
Panel	1984Q1	1984Q3	1989Q2	1989Q1	1989Q1	1983Q2	1994Q1	1957Q1
	3.1418	2.6597	2.3383	3.1435	3.0266	9.2293	15.355	6.7652
Single	1983Q4	1984Q3	1989Q2	1988Q4	1988Q4	1982Q4	1995Q1	1956Q4
	3.453476	2.654953	3.200971	4.202545	4.716993	10.10239	12.68627	6.658901

Table 4.

State	Missouri	Montana	Nebraska	Nevada	New Hampshire	New Jersey	New Mexico	New York
Panel	1984Q4 3.2478	1983Q3 2.9649	1984Q2 2.1026	1981Q3 2.9437	1989Q1 4.0799	1989Q1 4.8236	1984Q2 5.745	1987Q3 4.4605
Single	1984Q4 2.923832	1983Q2 2.07449	1984Q2 3.009992	1981Q3 2.775937	1988Q4 6.677632	1959Q4 35.46539	1984Q2 6.075391	1987Q3 5.344509

Table 5.

State	North Carolina	North Dakota	Ohio	Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina
Panel	1984Q4 2.2277	1989Q1 0.86597	1984Q4 7.0858	1982Q3 4.279	1982Q1 3.5157	1966Q4 38.626	1989Q1 4.5092	1984Q3 4.6258
Single	1959Q1 4.185137	1989Q2 1.43018	1984Q4 7.652469	1982Q3 3.185034	1982Q1 3.427007	1961Q2 3.214681	1953Q4 10.80853	1955Q3 5.112438

Table 6.

State	South Dakota	Tennessee	Texas	Utah	Vermont	Virginia	Washington	West Virginia
Panel	1983Q3 5.0838	1985Q4 13.006	1983Q2 7.1316	1983Q1 5.8304	1989Q2 2.6243	1986Q2 8.3995	2003Q4 5.5255	1981Q4 2.4154
Single	1983Q1 5.484646	1987Q2 16.78302	1952Q1 5.670415	1982Q4 5.799764	1953Q4 8.730882	1986Q2 7.561718	2003Q4 5.786927	1981Q4 2.086744

Table 7.

State	Wisconsin	Wyoming
Panel	1984Q3	1989Q1
	7.44	1.3456
Single	1970Q3	1953Q1
	61.4632	3.201719

The panel data method not only results in a more concentrated pattern of break dates, but it also significantly reduces the standard error of the change-point estimates when some of them exhibit relatively large standard errors if using single series model. Pennsylvania is an exception, however, whose standard error of the estimate increases from 3.2 to 38.6 when using panel data model. Also, Alabama, Delaware, Iowa, Minnesota and Tennessee still have standard error larger than 10 (2.5 year) when using the panel data model, and the standard errors for them are even larger if using the univariate model. To investigate the reason behind such a phenomenon, one needs to examine the estimation results for each state in greater detail. In the following two figures, I show the posterior distribution resulting from the panel data model for each state.

Figure 14. The posterior distribution of the break date for each state, Part I

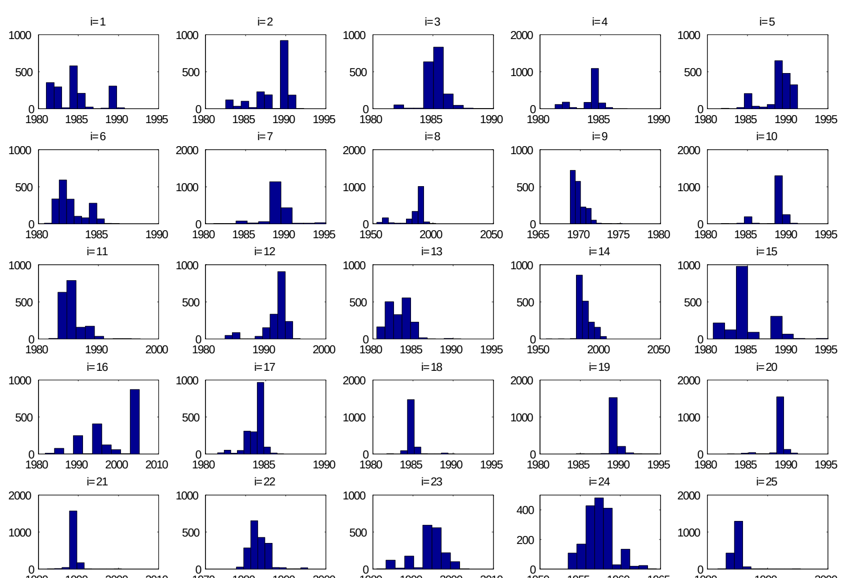
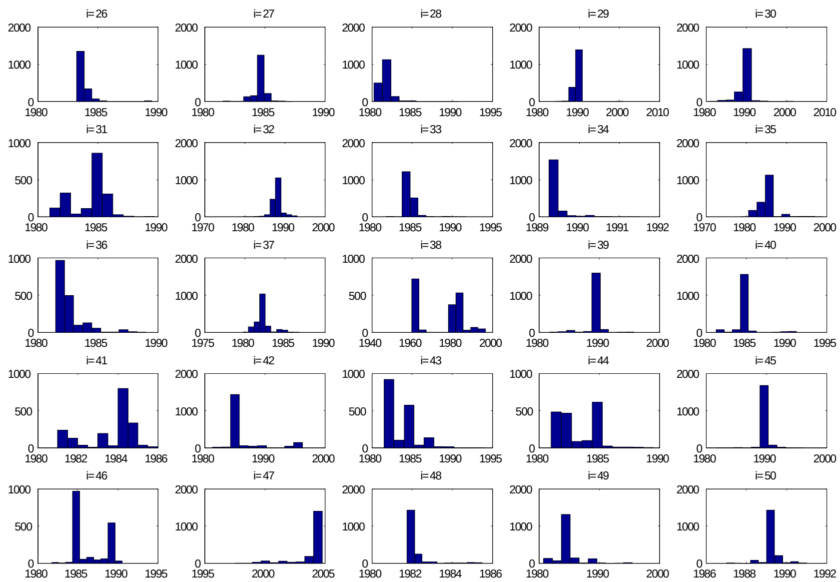


Figure 15. The posterior distribution of the break date for each state, Part II



After closely checking the posterior distributions of change points for all series, one may notice that although many states do experience the Great Moderation around 1984, others have structural change around 1988. Most of the posteriors are very concentrated around the means, but a few are less so, which causes large standard errors. There are two possible explanations for the large standard error of some states. First, such a state may not have a significant structural break at all during the sample periods, as Delaware and Illinois do. Second, a couple of states show very different dynamic patterns from the majority of the states. For instance, Alabama and Iowa seem to have more than one structural breaks, although the other states only show evidence of one break. In this instance, the standard error would increase too, since such states are far less similar than others.

Now assume that each time series has two structural breaks. Figure 16 shows the estimation results of the first structural break for 50 states, and Figure 17 shows the results for the second break date.

Figure 16. The first structural break

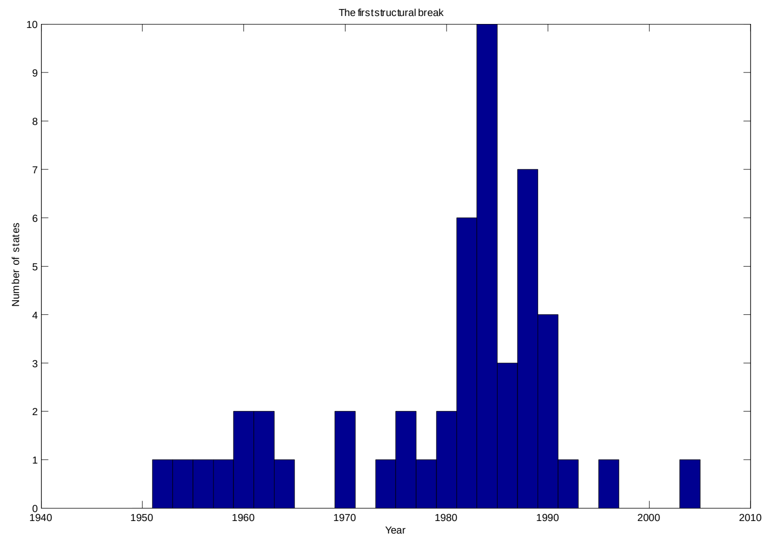
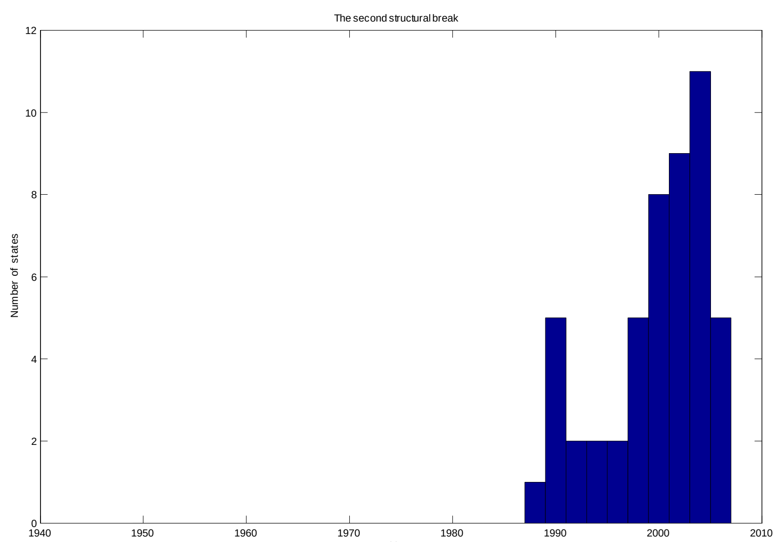


Figure 17. The second structural break



The above graph shows that only a few states were subject to two structural breaks, with the second break centered around the 1990s. There is little statistical evidence to support a second structural change for a majority of states, because most of the estimated second break points for those states hit the boundary. This suggests that most states experienced only one structural break.

7 Conclusion

This paper provides a new and effective Bayesian method to estimate and to make inference about structural breaks for panel data models. This method takes into account two key factors: first, all series are subject to common shocks, which are the source of a possible structural change; second, such common shocks have heterogeneous impact on different series, and thus the timing of structural changes can vary across series. By incorporating both the similarity and the heterogeneity among series, this method not only improves the quality of change-point estimation, but it also provides useful information about the cross-sectional pattern of structural changes. Monte Carlo simulations show that this method greatly increases the precision of change-point estimation. A key assumption is that change points for different time series follow a common distribution, which relaxes the conventional common-break restriction while retaining the commonality among series. Furthermore, if each time series has multiple structural breaks, a special form of the common distribution is imposed to restrict the joint distribution of multiple breaks. I apply this method to investigate the volatility decline, or the Great Moderation, in 50 U.S. states. Using quarterly personal income data, I find that the cross-sectional distribution of break dates has two peaks, one being around 1984, which is generally recognized as the start of the Great Moderation, and the other being around 1988, which corresponds to the stock market crash of the previous year. Only a couple of states showed no evidence of structural change in the volatility.

It is worth noting that the panel data method is easily adjusted to incorporate either more similarity or more heterogeneity among different series. In this paper, the only link among various series is the common distribution of the change points. One can assume that the model parameters of different series follow a common distribution by adding another layer to my Bayesian algorithm, especially if strong similarity exists across sections. On the other hand, one can also loosen the link among series by assuming different but correlated distribution of the change points. For example, in the nonreversible Markov chain model, one can assume that different series have different transition matrices P, which follows a common distribution. These are all interesting topics to explore.

8 Appendix: The Gibbs sampling algorithm for nonreversible hidden Markov chain

8.1 Sampling of the hidden states $\{s_{it}\}_{t=1,...T}^{i=1,...N}$ and the change points $\{k_i^m\}_{i=1,...N}^{m=1,...M}$

The algorithm for sampling $\{s_{it}\}_{t=1,...T}^{i=1,...N}$ is based on a modified version of the procedure in Chib(1998). Given the common transition matrix P, each series in the panel has its own dynamics and thus the states are sampled independently series by series.

To fixed the notation, let $S_{it} = (s_{i1,....,}s_{it})$, $S_i^{t+1} = (s_{it+1,....,}s_{iT})$, $Y_{it} = (y_{i1,....,}y_{it})$, and $Y_i^{t+1} = (y_{it+1,....,}y_{iT})$, for any series i. We need to draw a sequence of values $\{s_{it}\}_{t=1,...T}$ from the posterior distribution function $f(S_{iT}|Y_{iT},\ \theta,\ P)$.

First, we write the joint posterior density in reverse time order using the conditional probability rule

$$f(S_{iT}|Y_{iT},\theta,P) = f(s_{i,T-1}|Y_T, s_{i,T},\theta,P)\times\cdots\times\ f(s_{i,t}|Y_T, S_i^{t+1},\theta,P)\times\cdots\times\ f(s_{i,1}|Y_T, S_i^2,\theta,P) \tag{20}$$

To sample the hidden states $\{s_{it}\}_{t=1,...T}$ for any given time series in the panel, we follow the following procedure. Since we have assumed that there are M structural breaks for each time series i, then $s_{i,T} = M+1$.

1. draw $s_{i,T-1}$ from $f(s_{i,T-1}|Y_T, s_{i,T} = M+1,\theta,P)$

2. draw $s_{i,T-2}$ from $f(s_{i,T-2}|Y_T, s_{i,T} = M+1, s_{i,T-1},\theta,P)$, here $s_{i,T-1}$ takes the value drawn from the previous step.

.

.

.

N-2. draw $s_{i,2}$ from $f(s_{i,3}|Y_T, S_i^3,\theta,P)$, where S_i^3 take the values drawn from the previous steps.

The $s_{i,1}$ has been fixed as 1. Chib(1996) shows that

$$f(s_{i,t}|Y_T, S_i^{t+1},\theta,P) \propto prob(s_{i,t}|Y_{it},\theta,P)\cdot prob(s_{i,t+1}|s_{i,t},P) \tag{21}$$

$prob(s_{i,t+1}|s_{i,t},P)$ is just the transition probability given P. The first term $prob(s_{i,t}|Y_{it},\theta,P)$ is obtained by a recursive procedure called "filtering" starting from $t=1$. Given the value of $prob(s_{i,t-1}|Y_{it-1},\theta,P)$, one could update the $prob(s_{i,t}|Y_{it},\theta,P)$ following

$$prob(s_{i,t}=m|Y_{it},\theta,P)=\frac{prob(s_{i,t}=m|Y_{it-1},\theta,P)\times f(y_{it}|Y_{i,t-1},\theta_m)}{\sum_{l=1}^{M+1} prob(s_{i,t}=l|Y_{it-1},\theta,P)\times f(y_{it}|Y_{i,t-1},\theta_l)} \quad (22)$$

Notice that

$$\begin{aligned}
&prob(s_{i,t}=m|Y_{it-1},\theta,P) \quad (23)\\
&\quad= prob(s_{i,t}=m|s_{i,t-1}=m,P)\times prob(s_{i,t-1}=m|Y_{it-1},\theta,P)+\\
&\quad prob(s_{i,t}=m|s_{i,t-1}=m-1,P)\times prob(s_{i,t-1}=m-1|Y_{it-1},\theta,P)
\end{aligned}$$

Again, both $prob(s_{i,t}=m|s_{i,t-1}=m,P)$ and $prob(s_{i,t}=m|s_{i,t-1}=m-1,P)$ are the transition probabilities.

Given updated $\{s_{it}\}_{t=1,...T}^{i=1,...N}$, it is straightforward to get $\{k_i^m\}_{i=1,...N}^{m=1,...M}$. For any $i=1,...,N$, $k_i^m=k_i^{m-1}+\sum_{t=1}^{T} I(s_{i,t}=m)$.

8.2 Sampling of P

Assume the prior distribution of p_{mm} is Beta. And p_{mm} and $p_{l.l}$ are independent draws from the same Beta distribution, i.e.

$$p_{mm}\sim i.i.d\ Beta(a,b) \quad (24)$$

Since all series share the same P, p_{mm} is the same across sections and one must use panel data information to update the value of p_{mm}.

Given updated $\{s_{it}\}_{t=1,...T}^{i=1,...N}$, the posterior distribution of p_{mm} is assumed to be independent from data Y.

$$p_{mm}|\{s_{it}\}_{t=1,...T}^{i=1,...N}\sim Beta(a+n_{mm},b+1) \quad (25)$$

Where $n_{mm}=\sum_{i=1}^{N} I(s_{it}=m)$, and $I(s_{it}=m)=1$ if and only if $s_{it}=m$, $I(s_{it}=m)=0$ otherwise.

8.3 Sampling of other model parameters

For illustration purpose, I describe the sampling algorithm for the simple model $y_{i,t} = \begin{cases} \mu + \sigma_i^1 * e_{it,} \text{ for } 1 \le t \le k_i^* \\ \mu + \sigma_i^2 * e_{it,} \text{ for } k_i^* < t \le T \end{cases}$, $i = 1, ..., N$, and assume only one structural break. However, it is straightforward to apply the same algorithm to more complicated models. For more than one structural break, just divide the time series into more regimes and perform the same procedure for parameters in each regime.

In the above two sections, I have simulated the hidden states $\{s_{it}\}_{t=1,...T}^{i=1,...N}$ thus the change points $\{k_i^m\}_{i=1,...N}^{m=1,...M}$ as well. To update the common transition matrix P, one needs to take into account the updated information about $\{k_i^m\}_{i=1,...N}^{m=1,...M}$ for all the series in the panel. Once that is done, one can update other model parameters such as $\left(\mu_{ij}, \sigma_{ij}^2, k_i\right)$, $i = 1, 2, ...N,\ \ j = 1, 2,\ \ t = 1, 2, .., T$

Step 1: Given initial values $\left(\mu_{ij}^{(0)}, \sigma_{ij}^{2\ (0)}, k_i^{(0)}\right)$, $i = 1, 2, ...N,\ \ j = 1, 2,\ \ t = 1, 2, .., T$

Step 2: Update parameters μ_{ij}

Given the initial values of other parameters $\left(\mu_{ij}^{(0)}, \sigma_{ij}^{2\ (0)}, k_i^{(0)}, \pi_t^{(0)}\right)$, $i = 1, 2, ...N,\ \ j = 1, 2,\ \ t = 1, 2, .., T$ and data Y, the posterior conditional distribution of $\mu_{ij}^{(1)}$ is

$$\begin{aligned}
&\mu_{i1}^{(1)}|\overrightarrow{\sigma}^{2\ (0)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
&\sim Normal((\frac{\mu_{i10}}{\sigma_{i10}^2} + \frac{\sum_{t=1}^{k_i^{(0)}} y_t{}^2}{\sigma_{i1}^{2\ (0)}})/(\frac{1}{\sigma_{i10}^2} + \frac{k_i^{(0)}}{\sigma_{i1}^{2\ (0)}}), (\frac{1}{\sigma_{i10}^2} + \frac{k_i^{(0)}}{\sigma_{i1}^{2\ (0)}})^{-1}) \\
&\mu_{i2}^{(1)}|\overrightarrow{\sigma}^{2\ (0)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
&\sim Normal((\frac{\mu_{i20}}{\sigma_{i20}^2} + \frac{\sum_{t=k_i^{(0)}+1}^{T} y_t{}^2}{\sigma_{i2}^{2\ (0)}})/(\frac{1}{\sigma_{i20}^2} + \frac{T - k_i^{(0)}}{\sigma_{i2}^{2\ (0)}}), (\frac{1}{\sigma_{i20}^2} + \frac{T - k_i^{(0)}}{\sigma_{i2}^{2\ (0)}})^{-1})
\end{aligned} \tag{26}$$

Step 3: Update parameters σ_{ij}^2

Given $\left(\mu_{ij}^{(1)}, k_i^{(0)}, \pi_t^{(0)}, X\right)$, $i = 1, 2, ...N,\ \ j = 1, 2,\ \ t = 1, 2, .., T$, $\sigma_{ij}^{2\ (1)}$ has the following posterior conditional distribution

$$\begin{aligned}
& \sigma_{i1}^{2\,(1)} | \overrightarrow{\mu}^{(1)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
\sim \quad & \text{Scaled inverse-chi-square}(\nu_{i1} + k_i^{(0)}, \frac{\nu_{i1}\delta_{i10}^2 + \sum_{t=1}^{k_i^{(0)}} (y_{it} - \mu_{i1}^{(1)})^2}{\nu_{i1} + k_i^{(0)}}) \\
& \sigma_{i2}^{2\,(1)} \,| \overrightarrow{\mu}^{(1)}, \overrightarrow{k}^{(0)}, \overrightarrow{\pi}^{(0)}, Y \\
\sim \quad & \text{Scaled inverse-chi-square}(\nu_{i2} + T - k_i^{(0)}, \frac{\nu_{i2}\delta_{i20}^2 + \sum_{t=k_i^{(0)}+1}^{T} (y_{it} - \mu_{i2}^{(1)})^2}{\nu_{i2} + T - k_i^{(0)}})
\end{aligned} \tag{27}$$

Here I choose conjugate priors for each parameter to relieve the computational burden. If the model bears a more complicated form and conjugate priors are not available for the model parameters, one can use Griddy-Gibbs sampling or standard Metropolis-Hasting Algorithm.

References

[1] Bai, J., 1994, "Least Squares Estimation of a Shift in Linear Processes," Journal of Time Series Analysis, 15, 453-470.

[2] ——— 1997, "Estimating Multiple Breaks One at a Time," Econometric Theory, 13, 315-352.

[3] ——— 2000, "Vector Autoregressive Models with Structural Change in Regression Coefficients and in Variance-Covariance Matrix," Annals of Economics and Finance, Vol. 1, 303-339.

[4] ——— 2006, "Common breaks in panel data," New York University, unpublished manuscript.

[5] Bai, J., R. Lumsdaine and J. Stock, 1998, "Testing for and Dating Common Breaks in Multivariate Time Series," Review of Economic Studies, 65, 395-432.

[6] Bai, J. and P. Perron, 1998, "Testing for and Estimation of Multiple Structural Changes," Econometrica, 66, 47–79.

[7] ——— 2003, “Computation and Analysis of Multiple Structural Change Models,” Journal of Applied Econometrics, 18, 1–22.

[8] Bauwens, L. and M. Lubrano, 1998, “Bayesian Inference on GARCH Models using the Gibbs Sampler,” Econometrics Journal, 1, C23–C46.

[9] Bekaert, G., C. R. Harvey, R. L. Lumsdaine, 2002, “Dating the integration of world equity markets,” Journal of Financial Economics, v65(2), 203-247.

[10] Carlin, B. P., A. E., Gelfand and A. F. M. Smith, 1992, “Hierarchical Bayesian Analysis of Change-point Problems,” Applied Statistics, Vol. 41, No. 2, pp. 389-405.

[11] Chib, S., 1995, “Marginal Likelihood from the Gibbs Output,” Journal of the American Statistical Association, 90, 1313–1321.

[12] ——— 1996, “Calculating Posterior Distribution and Modal Estimates in Markov Mixture Models,” Journal of Econometrics, 75, 79–97.

[13] ——— 1998, “Estimation and Comparison of Multiple Change Point Models,” Journal of Econometrics, 86, 221–241.

[14] Chib, S. and E. Greenberg, 1995, “Understanding the Metropolis–Hastings Algorithm,” American Statistician, 49, 327–335.

[15] Chib, S. and I. Jeliazkov, 2001, “Marginal Likelihood from the Metropolis–Hastings Output,” Journal of the American Statistical Association, 96, 270–281.

[16] Dubner, Stephen and S. Levitt, 2005, *Freakonomics*

[17] Gelman, A., J. B. Carlin, H. S. Stern and D. Rubin, 2002, *Bayesian Data Analysis*, second edition (London: Chapman & Hall).

[18] Giordani, P. and R. Kohn, 2006, “Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models,” Sveriges Riks Bank working paper series 196.

[19] Jarrett, R. G., 1979, “A Note on the Intervals Between Coal-Mining Disasters,” Biometrika, Vol. 66, No. 1, pp. 191-193.

[20] Joseph, L. and D. B. Wolfson, 1992, "Estimation in Multi-Path Change-Point Problems," Commun. Statist. Theory Meth., 21(4), 897–913.

[21] ——— 1993, "Maximum Likelihood Estimation in the Multi-Path Change-Point Problem," Ann. Inst. Statist. Math, Vol.45, No.3, 511-530

[22] Joseph, L., Wolfson, D. B., Berger, R. D., and Lyle, R. M., 1997, "Analysis of Panel Data with Change-points," Statistica Sinica, 7, 687-703.

[23] Kim, C., C. Nelson and J. 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.

[24] Koop, G. and S. Potter, 2004a, "Prior Elicitation in Multiple Change-point Models," Federal Reserve Bank of New York Staff Report, Number 197.

[25] ——— 2004b, "Forecasting and Estimating Multiple Change-point Models with an Unknown Number of Change-points," Mimeo, University of Leicester and Federal Reserve Bank of New York.

[26] Lamoureux, CG and W. D. Lastrapes, 1990, "Persistence in Variance, Structural Change, and the GARCH Model," Journal of Business and Economic Statistics, Vol. 8, No.2

[27] McConnell, M. M. and Perez-Quiros, G., 2000, "Output Fluctuations in the United States: what has changed since the early 1980's?" American Economic Review 90 (5), 1464–1476.

[28] McCulloch, R. E. and R. S. Tsay, 1993, "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," Journal of the American Statistical Association, Vol. 88, No. 423, pp. 968-978.

[29] Owyang, M. T., J. Piger and H. J. Wall, 2008, "A state-level analysis of the Great Moderation," Regional Science and Urban Economics 38, 578–589.

[30] Perron, Pierre, 2006, "Dealing with Structural Breaks," Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 278-352.

[31] Perron, Pierre and Z. Qu, 2006, "Estimating Restricted Structural Change Models," Journal of Econometrics, 134, 373-399.

[32] ——— 2007, "Estimating and Testing Structural Changes in Multivariate Regressions," Econometrica, 75, 459-502.

[33] Pesaran, H., D. Pettenuzzo and A. Timmermann, 2006, "Forecasting Time Series Subject to Multiple Structural Breaks," Review of Economic Studies, 73, 1057-1084.

[34] Pesaran, H. and A. Timmermann, 2005, "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks," Journal of Econometrics, 2005, 183-217.

[35] Raftery, A. E. and V. E. Akman, 1986, "Bayesian Analysis of a Poisson Process with a Change-Point," Biometrika, Vol. 73, No. 1, pp. 85-89.

[36] Stephens, D. A., 1994, "Bayesian Retrospective Multiple-Changepoint Identification," Applied Statistics, Vol. 43, No. 1, pp. 159-178.

[37] Stock, J. H. and M. Watson, 1996, "Evidence on Structural Instability in Macroeconomic Time Series Relations," Journal of Business and Economic Statistics, vol. 14(1), 11-30.

[38] ——— 2004, "Combination Forecasts of Output Growth in a Seven-Country Data Set," Journal of Forecasting, 23, 405–430.

[39] Western, B. and M. Kleykamp, 2004, "A Bayesian Change Point Model for Historical Time Series Analysis," Princeton University, unpublished manuscript.