Forecasting Under Structural Break Uncertainty

Jing Tian *

Abstract

This paper examines some existing forecasting techniques that can be used when the forecasting model has possibly undergone structural changes at unknown points in time. We also propose two new forecast methods that are designed to account for structural changes. The proposed combination forecasts are evaluated using Monte-carlo techniques, and they outperform forecasts based on other methods that try to account for structural change, including average forecasts weighted by the past forecasting performance and techniques that first estimate a break point and then forecast using the post break data. An empirical application based on a NAIRU Phillips curve model for the United States indicates that it is possible to outperform the random walk forecasting models when we employ the forecasting methods to account for break uncertainty.

Keywords: forecasting with structural breaks, parameter shifts, break uncertainty, structural break tests, the choice of estimation sample, forecasting combination, the NAIRU Phillips curve.
JEL code: C22, C53, E37.

*Correspondence to: Jing Tian, School of Economics, College of Business and Economics, The Australian National University, ACT, 0200, Australia. Phone: 61-2-61256126. Email: jing.tian@anu.edu.au. The author is grateful for the insight comments and supportive supervision from Professor Heather Anderson and Professor Farshid Vahid

1 Introduction

The forecasting of economic variables is often complicated by the possibility that the parameters in the underlying data generating process (DGP) might have changed at various points in time during the pre-forecast sampling period. In this paper, we define structural breaks as permanent shifts in parameters of a DGP, and we focus on the problem that structural breaks often affect the forecasts that rely on model estimation. The failure to identify in-sample breaks that change the data generating process produces biased parameter estimates and thus contaminates the model's out-of-sample forecasting performance. Ideally, if information on breaks, such as breakpoints and break sizes, is known, we can decide the estimation window size according to the trade off between the bias and forecast error variance to improve the out-of-sample means squared forecasting errors (see Pesaran and Timmermann, 2007). However, forecasters are commonly in the absence of the knowledge on structural breaks.

The main motivation for this paper is to determine how to choose the forecasts based on different estimation windows under break uncertainty. There are two main contributions of this paper. First, we propose a new weighting scheme that utilizes the recursively ordered Cusum squared (ROC) test results to average the forecasts based on different estimation windows. We also propose a second forecasting weighting technique that simply places more weights on the forecasts based on more recent samples. Second, we evaluate the forecasting ability of the NAIRU Phillips curve on the U.S. 12-month inflation changes using this method and the other existing approaches to deal with structural break uncertainty, and have showed improved forecasting performance.

The idea of combining forecasts across various estimation samples reflects the theoretical point in Pesaran and Timmermann (2007) that forecasting performance can be better off with including pre-break data. The reason to implement the ROC test proposed by Pesaran and Timmermann (2002) is to obtain some knowledge of the most recent breakpoint under the break uncertainty. We treat each in-sample time as a possible most-recent break and generate forecasts using data after those time. The averaging weights are built on the probability of each in-sample time being the last break, which we use the ROC test statistics and a prior function that indicates the location of each time point to approximate. We also examined a set of other methods producing either a single forecast or an averaging forecast to incorporate break uncertainty problem.

To investigate the forecasting performance of these techniques that account for structural breaks in practice, we employ the unemployment-based NAIRU Phillips curve model to forecast the U.S. 12-month inflation changes. Although Stock and Watson (1999) find statistically significant shifts of the coefficients in the NAIRU model, they claim that the existing methods that facilitate parameter instability such as rolling regression do not produce better forecasts than expanding window forecasts and thus ignore the breaks. It seems that the other authors follow the same reason and only focus on a full sample estimation (see Atkeson and Ohanian, 2001; Fisher et al., 2002). An important contribution of this paper is to show the usefulness of dealing with structural breaks in inflation forecasts. Our out-of-sample forecasting results show that the average forecast weighted by ROC-statistics and the prior function the breakpoint location achieves a reduction of $MSFEs$ comparing with expanding window forecasts. In particular, the location based weighting scheme that places more weight on the more recent information shows a powerful forecasting ability. In addition, the other break-dealing methods such as the cross-validation method, 2-stage ROC method, the Bai-Perron method, simple average forecasts and the MSFE-weighted forecasts can all improve the forecasting ability of the unemployment-based NAIRU Phillips curve model and defeat the random walk forecasts of zero inflation changes.

The outline of the paper is as the follows. The next section explains the details of the forecasting methods that account for structural break uncertainty, including the new combination weighting scheme that implies the probability of each time point to the most recent break. These methods are firstly examined by the Monte Carlo simulations in section 3. We then turn in the section 4 to employ these approaches to conduct an out-of-sample forecasting exercise of the U.S. 12-month inflation changes based on the NAIRU Phillips curve model.

2 Forecasting Methods

Assume that the following linear model is subject to m structural breaks $(T_1, T_2, ..., T_m)$:

$$y_t = X_{t-1}^{'}\beta_j + u_t, \quad j = 1, 2, ..., m+1 \quad \text{and} \quad T_{j-1} + 1 \leq t \leq T_j \tag{1}$$

Here y_t is the dependent variable at time t and X_{t-1} is a $p \times 1$ vector of regressors at time $t-1$. The $1 \times p$ vector of β_j denotes the values of coefficients of regressors in each segment j. This setup, often named as the pure structural break model, implies that when a structure break occurs, all of p coefficients will shift permanently until the next breakpoint. Since we are not considering the model's misspecification problem caused by structural breaks, we let the vector of regressors X keep the same across all of the segments.

2.1 Single Forecast

Under breakpoint uncertainty, forecasters usually use structural break tests to identify break dates and obtain a better forecast, by working with a forecasting model that only uses data subsequent to the identified most recent break. Considering the likelihood of the occurrence of multiple breaks, we adopt a two-stage reversed ordered Cusum method and the Bai-Perron method to detect and estimate the most recent breakpoint.

An alternative approach to determine an appropriate estimation window and produce an associated single forecast is the cross-validation method. On the contrast to the approaches that explicitly identify breakpoints, the cross-validation method firstly calculates a set of forecasts based on various estimation windows and then search for the "best" one that produces an optimal criterion (e.g. MSFE) over a "testing" sample prior to the forecasting period.

2.1.1 Bai-Perron Method

The Bai-Perron Method (see Bai and Perron, 2003) estimates the number of multiple breaks and identifies the break locations. It requires assumptions on the maximum number of breaks and the minimum distance between every two breaks. Suppose that the maximum number of breaks is M and h denotes the minimum break distance. For every option of the number of breaks $m = 1, 2, ..., M$, the estimated locations of breaks $\{\hat{T}_j\}$ are derived by minimizing the global sum of squared residuals, such that:

$$(\hat{T}_1, .., \hat{T}_m) = argmin_{(T_1,...,T_m)} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - X'_{t-1}\hat{\beta}_j)^2 \tag{2}$$

where $T_j - T_{j-1} \geq h$ and the least squares estimates of coefficients $\{\hat{\beta}_j\}$ are associated with the m estimated breakpoints $\{\hat{T}_j\}$.

Bai and Perron (2003) introduce some approaches to determine the number of breaks from $\{m : |1 \leq m \leq M\}$. In this paper, we use the Schwartz information criterion (BIC) to select the number of breaks after the optimal locations of breakpoints have been determined for each value of m.

The Bai-Perron method can consistently estimate the number of breaks and the breakpoints in the presence of multiple parameter changes. However, the estimation results are subject to the assumptions on the maximum number of breaks M and the minimum distance between breaks h. Further, estimation of breakpoints can be imprecise when some breaks are small and a forecast that relies on imprecisely estimated breakpoints is unlikely to be reliable.

2.1.2 Two-Stage ROC Method

Two-stage reversed ordered Cusum (ROC) method is proposed by Pesaran and Timmermann (2002) for dealing with parameter instability problem in forecasting the U.S. stock returns. Based on a Cusum squared test that is often applied for testing a single structural break, they conduct a backward looking Cusum squared test to directly estimate the most recent break as the first step.

Suppose the observation matrices $y_{T:\tau}$ and $X_{T:\tau}$ are in a reversed time order given by:

$$y'_{T:\tau} = [y_T, y_{T-1}, ..., y_{\tau+1}, y_\tau], \quad X'_{T:\tau} = [X_T, X_{T-1}, ..., X_{\tau+1}, X_\tau].$$

With the minimum estimation window size $\bar{w} = T - \tilde{T}$, we derive the least squared estimates of β as:

$$\hat{\beta}_\tau = (X'_{T:\tau} X_{T:\tau})^{-1} X'_{T:\tau} y_{T:\tau}, \quad \tau = \tilde{T}, \tilde{T} - 1, ..., 2, 1. \tag{3}$$

The ROC test statistics s_τ are constructed based on the squares of standardized one-step-ahead recursive residuals v_τ from the reversed ordered regression:

$$s_\tau = \frac{\sum_{j=\tilde{T}}^{\tau} v_j^2}{\sum_{j=\tilde{T}}^{1} v_j^2}, \tau = \tilde{T}, \tilde{T} - 1, ..., 2, 1, \tag{4}$$

where v_τ is computed as:

$$v_\tau = \frac{y_\tau - \hat{\beta}'_{\tau+1} X_\tau}{\sqrt{1 + X'_\tau (X'_{T:\tau+1} X_{T:\tau+1}) X_\tau}}. \tag{5}$$

To estimate the most recent break, we choose the first time that the ROC test statistics sequence s_τ crosses one of the lines of critical values $(\tilde{T} - \tau + 1)/(\tilde{T}) \pm c_0$ in which c_0 can be simulated based on Brown et al. (1975) [1].

Conditional on the detection of parameter shifts, in the second step of the two-stage ROC method, we trim all the data prior to the estimated most-recent breakpoint and only use the post-break sample to estimate forecasting models. If no break is identified by *ROC* test, a full-sample estimation will be implemented. Therefore, a possible failure

[1] This ROC procedure does not produce consistent estimate of breaks since there is always α probability of falsely rejecting the null hypothesis of parameter constancy, where α is the test level.

for the ROC test to detect a mall in-sample parameter shift may result in a poor forecast from using the full sample to estimate forecasting models.

2.1.3 Cross-validation

The cross-validation method provides us with a way to choose the "best" estimation window for forecasting models without estimating breakpoints. Given a test sample $[T - \tilde{w}+1 : T]$ including $\tilde{w}$ observations, the optimal estimation window is chosen to start from time t_0^* if it minimizes some criterion such as the MSFE over the test sample:

$$t_0^*(T, \bar{w}, \tilde{w}) = argmin_{t0=1,..,T-\bar{w}-\tilde{w}}\{\tilde{w}^{-1}\sum_{\tau=T-\tilde{w}}^{T-1}(y_{\tau+1} - X_\tau'\hat{\beta}_{t0:\tau})^2\}. \quad (6)$$

This method can be less reliable if a break occurs during the predetermined test period. In other words, to apply this approach effectively, we need some pre-knowledge of the location of the most recent break, which is unlikely to happen in practice. After studying the evidence shown in many empirical papers, Timmermann (2005) concludes that a single forecast with the best tracked record often performs badly in out-of-sample experiments.

2.2 Forecasting Combination

A single forecast with one estimation window determined by the identification of the most-recent breakpoint may not be optimal. Apart from the problems of each method addressed above, even if one can estimate the locations of breaks accurately, due to the trade-off between the bias and forecast error variance, the forecasting performance measured by MSFE can be improved when pre-break data is included to estimate parameters of forecasting models, especially if the break is not large or the most-recent breakpoint is very close to the end of the sample.

Averaging forecasts based on different estimation windows is one approach that can be used to include more data other than post-break information. Then the question is how to weight forecasts that are based on different estimation windows. In this paper, we examine the forecasting performance of the equal weighting scheme and relative performance weighting scheme. We also propose a new combination method that incorporates the results of the reversed ordered Cusum squared test.

2.2.1 Equally Weighted Forecasts

A simple but powerful combination method that deals with break uncertainty is to equally average forecasts based on different estimation windows. Under the assumption that the last break τ can only occur in $[\tau_0 : T - \bar{w}]$ that is subject to the minimum estimation window size $\bar{w}$, we compute the equally weighted forecast as:

$$\hat{y}_{T+1} = \frac{1}{T - \bar{w} - \tau_0 + 1} \sum_{\tau=\tau_0}^{T-\bar{w}} (X_T' \hat{\beta}_{\tau+1:T}). \tag{7}$$

The value of τ_0 depends on forecasters' prior knowledge of structural breaks. Without any prior belief on the likelihood of break occurrence in the history, we can also consider the case of no structural break and thus set $\tau_0 = 0$. By doing so, essentially we put equal weights on the forecasts when breaks present in the past and the forecasts when no break occurs.

Despite its simplicity, many researchers find that the equal-weighting scheme often performs superior to the other more elaborated methods (see Stock and Watson, 1999; Pesaran and Timmermann, 2007; Clark and McCracken, 2006).

2.2.2 Average Forecasts Weighted by the Relative Performance

One forecast combination method is to base the weights on the relative forecasting performance. For example, under the squared error loss function, the weight for a forecast using data $[t_0 : T]$ is proportional to the inverse of its associated test sample $MSFE_{\tilde{w},t_0} = \tilde{w}^{-1} \sum_{\tau=T-\tilde{w}}^{T-1} (y_{\tau+1} - X_\tau' \hat{\beta}_{t_0:\tau})^2$, which is computed over the window of $\tilde{w}$ periods previous to time T. Then we consider the whole range of the values of $t_0 \in 1, 2, ..., T - \bar{w} - \tilde{w}$, assuming the minimum estimation window $\bar{w}$, and compute the forecasts $\hat{y}_{T+1,t_0} = X_T' \hat{\beta}_{t_0:T}$ for each value of t_0. The weighted average forecast thus is given by:

$$\hat{y}_{T+1} = \frac{\sum_{t_0=1}^{T-\tilde{w}-\bar{w}} \hat{y}_{T+1,t_0} (\frac{1}{MSFE_{\tilde{w},t_0}})}{\sum_{t_0=1}^{T-\tilde{w}-\bar{w}} (\frac{1}{MSFE_{\tilde{w},t_0}})} \tag{8}$$

Since this method does not require any break test or identification of break points, it avoids the imprecise break estimation problem. However, to obtain a reliable measure of the historical forecasting performance, a test sample containing a moderate number of data is necessary. Otherwise the combination weights based on MSFE over a small period can be very misleading. In practice, when only limited number of data is available, after taking some for an out-of-sample experiment, making a sensible size for test samples becomes difficult. Moreover, similar with the cross-validation method, this weighting scheme requires an implicit assumption that no break occurs during the test period, which might not be realistic.

2.2.3 Average Forecasts Weighted by the ROC Statistics

We propose an alternative combination method that combines forecasts derived from different post-break estimation windows, treating each past time as a possible most-recent breakpoint. The weights are based on the probability of the break occurrence at each time point. Under structural break uncertainty, the break tests provide us an approach to evaluate the likelihood for each time location to be most-recent breakpoint. In this paper, we adopt the ROC and construct the weights ww_τ on each choice of the most-recent breakpoint τ by the ROC statistics s_τ given as the following equation (4)[2] :

$$ww_\tau = \frac{|s_\tau - (\frac{T-\bar{w}-\tau-1}{T-\bar{w}})|\pi_\tau}{\sum_{\tau=1}^{T-\bar{w}} |s_\tau - (\frac{T-\bar{w}-\tau-1}{T-\bar{w}})|\pi_\tau}. \tag{9}$$

With a small number of observations $\bar{w}$ for the minimum estimation window, we consider all past dates $\tau \in [1 : T - \bar{w}]$ as a sequence of choices for the last breakpoint and approximate the associated probabilities as the absolute values of the distance between the calculated ROC statistics s_τ and the mid-points of two critical values given by $\frac{T-\bar{w}-\tau+1}{T-\bar{w}}$. The intuition is that the farther the ROC statistic is away from the mid-point of two critical values, the more likely it goes across the critical value lines, implying a higher probability of parameters shifts at the associated time point. This combination method provides an averaged forecast

$$\hat{y}_{T+1} = \sum_{\tau=1}^{T-\bar{w}} ww_\tau (X_T' \hat{\beta}_{\tau+1:T}). \tag{10}$$

The value of π_τ functions as a prior weight that indicates a prior belief on the probability of time τ to be the most recent break. The specification of π_τ depends on forecasters' knowledge of structural breaks. For instance, for a shorter period T when the presence of a single break is possible, we can define $\pi_\tau = 1$ for any value of τ, which makes the weight ww_τ entirely depend on the magnitude of the ROC statistics. In the paper, we refer the average forecasts from this weighting method as the ROC-weighted forecasts. However, if we have longer historical data and believe that multiple structural breaks may present

[2]The use of the reciprocals of p-values from break tests provides an alternative approach for building the weights.

in the past, this specification becomes less sensible. The reason lies in the following. If the ROC statistics are observed to be far from the mid-point of the critical-value lines at some early time points, with $\pi_\tau = 1$, the average forecasts will heavily rely on the forecasts generated based on the information subsequent to the early breakpoints indicated by the ROC statistics. To incorporate the idea that the identification of the most-recent break is more helpful in the forecasting context, we can also define the prior weight π_τ to be a function of the location of time τ in the full sample $[1 : T]$, such that

$$\pi_\tau = \frac{\tau}{T}. \tag{11}$$

This function suggests that heavier weights are placed on the forecasts based on more recent sample because the most-recent break is more likely to happen at the end. After setting $\pi_\tau = \frac{\tau}{T}$, we reduce the weights on the forecasts using information from the beginning of the sample even if the ROC statistics suggest the presence of structural breaks at the some early time points. In order to distinguish from the ROC weights that we have discussed, we name this weighting scheme adjusted-ROC, and the associated forecasts are called adjusted-ROC-weighted forecasts.

3 Monte Carlo Experiments

To evaluate each forecasting method that deals with different levels of break uncertainty, we consider the following bivariate data generating process and operate Monte-Carlo simulations:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_{yt} \\ \mu_{xt} \end{pmatrix} + A_t \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}. \tag{12}$$

where the variance covariance matrix of errors is set to be $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ across the whole sample. In a two-breaks case where the first and the second break occur at $p1 \times T$ and $p2 \times T$, respectively, parameters in the matrix A_t shift permanently after these two breakpoints:

$$A_t = \begin{cases} \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} & t \le p1 \times T \\ \begin{pmatrix} a_{11}+d_{11} & a_{12}+d_{12} \\ 0 & a_{22}+d_{22} \end{pmatrix} & p1 \times T + 1 \le t \le p2 \times T \\ \begin{pmatrix} a_{11}+d_{11}+d_{11}^* & a_{12}+d_{12}+d_{12}^* \\ 0 & a_{22}+d_{22}+d_{22}^* \end{pmatrix} & t \ge p2 \times T + 1 \end{cases} \tag{13}$$

This DGP follows the simulation setup in Clark and McCracken (2005), where they use it for evaluating the small sample properties of various in-sample predictive ability tests in the presence of structural breaks. Pesaran and Timmermann (2007) also adopt this DGP for comparing different methods on the choice of estimation windows when structural break occurs. This specification allows all of the coefficients as well as only parts or one of the coefficients to change when breaks occur. The constants in both equations will adjust according to the shifts of coefficients of regressors to keep the whole DGP stationary.

Suppose the full sample contains $T = 100$ observations and we are interested in the forecast of y_{101}. The loss function is squares of the forecasting errors. After repeating the simulation 5000 times, we evaluate various forecasting methods that have been discussed in the last section by comparing their MSFEs with the benchmark forecasts based on full sample estimations when structural breaks are ignored. Tables 1 records the ratios of the MSFEs relative to the benchmark forecasts. In the first three columns, we consider single forecasts that deal with structural breaks, where estimation windows are determined by the cross-validation method, the two-stage ROC method, and the Bai-Perron method. The columns headed *Equal-W*, *MSFE-W ROC-W1*, and *ROC-W2* relate to combination forecasts with the equal weights, the weights measured by historical *MSFE*s, the ROC weights, as well as the adjusted-ROC weights. We set the minimum size of estimation windows $\bar{w}$ to be $0.1 \times T$ through the whole exercise, thus the minimum distance between two breaks in the Bai-Perron method is also $0.1 \times T$. Bai and Perron (2003) suggest that a maximum of 5 breaks is sufficient in most empirical work, so we consider the possibilities up to 5 breaks when we estimate the breakpoints using the Bai-Perron method. As required in conducting *MSFE*-weighted forecasts and cross-validation forecasts, a test sample prior to time T is set to contain $\tilde{w} = 0.25 \times T$ observations.

In the simulations reported in table 1, both y_t and x_t are persistent, with their autoregressive parameters $a_{11} = 0.9$ and $a_{22} = 0.9$. The marginal effect of x_{t-1} on y_t is initially one unit. Here we consider various scenarios where the changes of coefficients, denoted as d_{ij} and d^*_{ij} for $i, j = 1, 2$, have different sizes and signs. We assume two true breaks happen at the one fourth and three fourths of the full sample (i.e $p1 = 0.25$ and $p2 = 0.75$).

[TABLE 1 ABOUT HERE]

For the first row, the persistency of y_t drops by a large proportion at both breaks. The effects of single forecasts on average are almost as good as combination forecasts except for the two-stage ROC method. The average forecast weighted by the adjusted-ROC is the most accurate with over 50% gain compared with the forecasts that ignore the breaks. The single forecast when breakpoints are identified by the Bai-Perron method produces nearly the same result. With the same final value of the autoregressive coefficient of y_{t-1}, the second and the third row only differs with respect to the size of each break. When the second break is bigger than the first break, the Bai-Perron method successfully identifies the breakpoints and generates the best forecasts among all the methods. The relative MSFE under *ROC-W2* is the second lowest, which continuously beats the equally weighted forecasts. In the next row where the second break is smaller than the first, the overall forecasting performance is better than the reversed scenario, but forecasts from the Bai-Perron method become one of the worst.

The next two rows report the results for the scenario that the autoregressive coefficient a_{11} decreases at the first breakpoint and then increases at the second one. When half of the drop is recovered after the second break, the combination forecasts dominate the single forecasts, with the best performance from the simple equal-weights. When a_{11} totally recovers after the second break, the single forecast using the Bai-Perron method to estimate breakpoints and the adjusted-ROC-weighted forecast deliver the smallest two MSFEs.

The seventh and the eighth rows report the benefits of using these forecasting methods for structural break uncertainty when the marginal coefficient of x_{t-1} changes at each break. With a large upwards shift, the gains from considering breaks are considerable. Among all of the weighted forecasts, MSFE under *ROC-W2* is the lowest whereas the single forecast using the Bai-Perron method is the most accurate. However, the performance of the Bai-Perron method depends on the size and the direction of the shifts on a_{12}. For the seventh row where a_{12} declines by a smaller amount each time, weighted forecasts dominate the single forecasts.

The forecasting results when no breaks present in the past are shown in the last row. The two-stage ROC method achieves the MSFE closest to 1. This result is expected since we apply the full sample to estimate forecasting models if no break can be detected by the ROC test. Without the presence of breaks, the sequence of the ROC statistics could display as a rather flat line, resulting in similar distance between each of the ROC statistics and the mid point of two critical lines and thus similar weights for the forecasts using different estimation windows. This could explain the reason that we observe similar MSFEs produced by *ROC-W1* and the equal weighting scheme. Although no break has occurred, the Bai-Perron method detects one break with a high frequency of 88.3%, leading the worst forecasting result produced by the Bai-Perron method.

Overall, the results in table 1 suggest the follows. The forecast weighted by the adjusted-ROC performs well and on average beats the equally weighted forecast. Between the two methods that both require a test sample, the cross-validation method that produces a single forecast generally performs better than a MSFE-weighted forecast. The largest MSFEs of the single forecast based on the two-stage ROC method result from the use of full sample estimations if we fail to reject the null hypothesis of parameter constancy. The single forecast using the Bai-Perron method to determine the estimation window performs worse than the combined forecasts when small breaks or no breaks have occurred. The reasons might lie in the follows. Firstly, small breaks are difficult to detect and estimated accurately by the Bai-Perron method. Secondly, as suggested in Pesaran and Timmermann (2007), when breaks are small, it is not optimal to use only the post-most-recent break data to estimate forecasting models under a squared-error loss function.

Notice that the adjusted-ROC that considers a a differential prior belief on each time to be the most-recent breakpoint constantly produces better forecasts than the ROC-weight that is only determined by the ROC statistics. Thus we conjecture that the gain results from weighting more on the forecasts based on more recent information. To check it, we consider the forecasting combination weights to be proportional to the location of time τ in the whole sample $[1 : T]$, i.e τ/T. The associated average forecasts, named as the location-weighted forecasts, are given by

$$\hat{y}_{T+1} = \sum_{\tau=1}^{T-\bar{w}} \frac{\tau/T}{\sum_{\tau=1}^{T-\bar{w}} \tau/T} \left(X_T' \hat{\beta}_{\tau+1:T} \right). \tag{14}$$

We evaluate this weighting scheme using the same simulation setup and record the relative MSFEs in table 2. By comparing table 1 and table 2, we find that in the presence of structural breaks, the location-based weight outperforms all the other weighting schemes including the "superior" simple average and the weights based on the ROC structural break test. Moreover, the location-weighted forecast is more accurate than single forecasts in most cases. The only comparable forecasting method is the Bai-Perron method that seems to perform slightly better than the location-weighting method when the second break is bigger than the first.

4 Forecasting Inflation

4.1 NAIRU Phillips Curve Models

Inflation forecasts have important implications for monetary policy makers. Among various models for inflation, a Phillips curve which connects the top two domestic economic burdens, unemployment and inflation, attracts the most attention. The early version of the Phillips curve implies a durable tradeoff between unemployment and inflation, but nowadays more and more economists advocate a "natural rate" of inflation that guides the economy back to equilibrium (Tobin, 1972). For instance, a specification of the Phillips curve called NAIRU (non-accelerating inflation rate of unemployment) presents the idea that inflation will increase if unemployment stays below its natural rate. A textbook version of the NAIRU model for 12-month ahead inflation changes is,

$$E_t(\pi_{t+12} - \pi_t) = \beta \times (u_t - \bar{u}) = -\beta\bar{u} + \beta u_t \tag{15}$$

where π_t and u_t denotes the inflation and unemployment rate, respectively. The NAIRU here is time invariant $\bar{u}$ that merges to the intercept. This model not only provides researchers a method to estimate the baseline unemployment rate, i.e. NAIRU, but also becomes a popular inflation forecasting model because of its simplicity and backward-looking specification.

The usefulness of the NAIRU Phillips curve for inflation forecasts has been discussed in several papers. Under a simulated out-of-sample framework, Stock and Watson (1999) compare the forecasts of the U.S. inflation rates at the 12-month horizon from 1970 to 1996 based on a variety of the NAIRU Phillips curve-based models. The basic 12-month ahead forecasting models used in their paper can be generalized as

$$\pi^{12}_{t+12} - \pi_t = \phi + \beta(L)x_t + \gamma(L)\Delta\pi_t + e_{t+12} \tag{16}$$

where π^{12}_t is 12-month inflation at time t whereas π_t is monthly inflation at an annual rate; x_t is defined as the unemployment rate, another macroeconomic variable, or a diffuse index measuring aggregate real activity at time t. They find that the conventional model with an unemployment rate gap produces no better forecasts than models based on the measures of real aggregate activities.

Atkeson and Ohanian (2001) provide evidence that the 12 month-ahead U.S. inflation forecasts from 1985 to 2001 based on the NAIRU Phillips curve models cannot be better than "flipping a coin". The benchmark forecasting model, which subsequent literature often calls Atkeson and Ohanian model, indicates no change in 12 month ahead inflation and is set to be,

$$E_t\left(\pi_{t+12}^{12}\right) = \pi_t^{12} \tag{17}$$

In order to make the NAIRU Phillips-curve-based inflation forecasts comparable with the benchmark directly, they revise Stock and Watson's model to

$$\pi_{t+12}^{12} - \pi_t^{12} = \alpha + \beta(L)x_t + \gamma(L)\Delta\pi_t + \varepsilon_{t+12}. \tag{18}$$

Triggered by the debate on the usefulness of the NAIRU Phillips curve, Fisher et al. (2002) examined the U.S. inflation forecasts generated from equation (18) within three distinct sample periods during which the inflation changes present different volatilities. Their results confirm those from Atkeson and Ohanian (2001) only within the low volatility periods that the Phillips curve models perform poorly, However, once they change the measurements of inflation or revise the models for a 24-month-ahead forecast horizon, the NAIRU Phillips curve models become favorable.

4.2 Instability of NAIRU Phillips Curve

The discussion of the validity of using NAIRU Phillips curve to forecast inflation focuses on model stability. Based on the well-known specification of autoregressive distributed lagged models for the NAIRU Phillips curves, the analysis of parameter stability in previous study includes the statistical relationship between inflation changes and unemployment rates (or other variables revealing real activity), the persistence of inflation changes, and the NAIRU level that is closely related to the intercept of the model[3].

Atkeson and Ohanian (2001) argue that the relationship between the current unemployment rates and the future inflation should vary when the economic environment changes because individuals often adjust their expectation on economic variables according

[3]Note that as we use the means of squared forecasting errors to evaluate forecasts, any variance change should also results in model instability. However, this type of structural breaks is beyond the scope of this paper.

to policy changes. By simply plotting changes in inflation against current unemployment rates from 1960 to 1999, they observe a flatter negative slope after mid 80's, meaning a weaker relationship between inflation changes and unemployment rates. However, in the paper by Stock and Watson (1999), a series of structural break tests for the presence of a single break show strong evidence of instability on the coefficients of lagged inflation rates, but not on the unemployment rate coefficients.

Although autoregressive coefficients of Stock and Watson's models present instability, Stock and Watson (1999) find the shifts are quantitatively small and thus they ignore coefficients instability in their forecasts. In 2006, Stock and Watson revisit the U.S. inflation forecasts by scrutinizing a univariate inflation process. They suggest that the failure to vary the autoregressive coefficients may lead to the breakdown of a recursive autoregressive distributed lagged inflation forecasts. The study of the univariate inflation process help them to revise the Phillips curve model subject to a restriction between the autoregressive coefficients and activity variable coefficients.

The question of whether the NAIRU itself has changed overtime has also attracted policymaker and academics' attention. Staiger et al. (1997) model the U.S. NAIRU using a cubic spline and find statistical evidence of a declining shift from 1980's to 1990's. To estimate the movement of the NAIRU, Gordon (1997) treats the NAIRU as a time-varying variable that follows a stochastic process. His results confirm a lower NAIRU in the end of 1990's. However, Staiger et al. (1997) also report that their NAIRU estimates are very imprecise, which produce little difference between forecasts based on different values of the NAIRU.

The previous literature on the instability of the NAIRU Phillips curve illustrates the nature of break uncertainty in the parameters of this model. We thus implement different forecasting methods that account for break uncertainty to reexamine the inflation forecasting ability of the unemployment-based NAIRU Phillips curve model.

4.3 The NAIRU Forecasts of the U.S. Inflation

4.3.1 Empirical Model and Data

The main forecasting model (shown in equation (19)) in this paper are specified based on equation (18) since we would like to directly show the potential benefits of dealing with structural breaks from comparing with a benchmark forecast.

$$\pi_{t+12}^{12} - \pi_t^{12} = \alpha + \beta(L)x_t + \gamma(L)\Delta\pi_t^{12} + e_{t+12} \tag{19}$$

We measure the annual U.S. inflation at time t by computing the 12-month changes of the U.S. core CPI (CPI less food and energy), given by $\pi_t^{12} = 100 \times (\ln P_t - \ln P_{t-12})$. The activity variable x_t in this paper is mainly the unemployment rates, denoted as u_t, or the changes of unemployment rates, denoted as Δu_t.

All the data from 1959:01 to 2007:06 in a monthly frequency are retrieved from DATASTREAM. We aim to conduct the out-of-sample forecasts of the 12-month ahead inflation changes in the past 10 years (from 1997:07 to 2007:06). Thus the initial estimation starts by employing the information from 1959:01 to 1997:06 for forecasting the inflation change in 1997:07. We then recursively estimate the forecasting models once new information is included.

Setting the annual inflation changes rather than inflation itself as the dependent variable means we treat the inflation as an non-stationary or I(1) variable. This assumption is consistent with our empirical observation[4] across the whole sample.

Most literature includes the level of unemployment rates in the NAIRU Phillips curve[5]. However, if the lags of u_t are not stationary while the dependent variable $\pi_{t+12}^{12} - \pi_t^{12}$ is stationary, the imbalanced model may lack of explanatory power, which ruins the associated forecasting performance. Therefore, we also consider the models with lagged changes of unemployment rates as one of the regressors. Note that since $\Delta u_t = u_t - u_{t-12}$, we rewrite equation (19) as,

$$\pi_{t+12}^{12} - \pi_t^{12} = \alpha + \beta(L)u_t + \beta(L)u_{t-12} + \gamma(L)\Delta\pi_t^{12} + e_{t+12} \tag{20}$$

where longer historic information of unemployment rates is included and the estimation is subject to the restriction that coefficients of lagged u_t and lagged u_{t-12} are the same for the same lag order. In the next section, we report out-of-sample forecasting results generated from both forecasting models.

One minor difference of equation (19) from equation (18) is the second regressor that implies the own dynamics of the inflation changes. Following Stock and Watson's model, Atkeson and Ohanian (2001) define the monthly inflation at annual rates as $\pi_t = 1200 \times (\ln P_t - \ln P_{t-1})$ and use the lags of $\Delta\pi_t$ to predict the 12-month changes of annual inflation $\Delta\pi_{t+12}^{12}$. After examining the movements of both dependent variable and explanatory variables overtime, we find that the monthly changes of 1-month inflation at annual rates are extremely noisy, whereas the 12-month changes of annual inflation moves along a

[4]The conclusion is based on recursive ADF tests and Correlograms of π_t^{12}.

[5]See Stock and Watson (1999), Staiger et al. (1997), Atkeson and Ohanian (2001) and Fisher et al. (2002).

relatively smooth path. Therefore, we replace the noisy regressor in equation (18) with the lagged dependent variable (see equation (19)), resulting in a standard autoregressive distributed lagged model (ADL).

The lag structures of u_t and $\Delta\pi_t^{12}$ are time-variant and selected by both Akaike information criterion (AIC) and Schwartz information criterion (BIC) for every forecast. Generally speaking, BIC penalizes more for a long lag structure than AIC. Thus the models selected by BIC may avoid estimation errors of coefficients caused by the loss of degrees of freedom. However, when we have a relatively long historical data set or set the minimum estimation window size to be a function of the number of estimated coefficients, the advantage of BIC becomes rather inconclusive. Therefore, in this empirical forecasting exercise, we use AIC to allow for a long lag structural and let the number of lags to vary from 1 to 12 for both regressors.

To obtain preliminary knowledge about structural breaks in the evolution of the U.S. 12-month changes of annual inflation, we show the time series plot of $\Delta\pi_{t+12}^{12}$ from 1959:01 to 2007:06. Figure 1 clearly displays a dramatic volatility change around 1984. However, to know the parameter stability in the NAIRU Phillips curve, as the main focus of this paper, we need further structural break tests. Although the year 1984 could be a candidate breakpoint, we are still unsure about the number of breaks and whether December 1983 could be the last structural break. Therefore, we are facing a break uncertainty problem when forecasting the U.S. inflation changes based on the NAIRU Phillips curve models.

[FIGURE 1 ABOUT HERE]

4.3.2 Results

Tables 3 to 6 summarizes the out-of-sample forecasting results through July 1997 to June 2007 when the possible presence of structural breaks is considered. We report the means of squared forecast errors (MSFE) relative to the benchmark forecasts generated from equation (17). MSFEs that are smaller than one imply that there is an advantage of using the NAIRU Phillips curve model over the random walk model for inflation forecasts. Each table contains two panels. Panel A presents single forecasting results that are based on expanding-window estimations when possible breaks are totally ignored as well as a variety of break-dealing methods that choose only one estimation window and thus generate a single forecast. Under break uncertainty, we can treat each date as a possible last-breakpoint and obtain a series of post-possible-last-break forecasts. We apply five different weighting schemes to combine those forecasts and report the averaged forecasting results in panel B.

Table 3 gives the results based on the model (19) with lags of u_t. The lag length is chosen using AIC, BIC or simply set to 1, and we examine all given historical information for forecasting the annual inflation changes. If we totally ignore the possibility of break occurrence in the history, we use the expanding windows to estimate forecasting models. The MSFEs are around two times larger than a forecast of zero changes of annual inflation, which is consistent with the finding of poor forecasting ability from the unemployment-based NAIRU model in the other literature. The second row shows the results from rolling window estimations that researchers commonly use in forecasting financial series under the uncertainty of structural breaks. This method has a "moderate adaptivity" with "small coefficient revolution" (see Stock and Watson, 1996) since each time when we update the forecasting model, we include the most current data into the estimation window and remove the oldest. We obtained similar evidence as in Stock and Watson (1996) that rolling regression performs worse than expanding window when forecasting inflation[6]. However, the sizes of MSFEs drop dramatically after we employ the three methods to choose the "optimal" estimation window and produce single forecasts. In particular, with the lag structure selected by AIC and BIC for every estimation, the NAIRU model that incorporates break uncertainty beats the random walk model.

[TABLE 3 ABOUT HERE]

The difference of forecasting results between three methods lies in different recursive estimates of the last break or the "optimal" estimation window. Figure 2 shows the sequence of the recursive estimates of the last break from the Bai-Perron method and the 2-stage ROC. Assuming that there can be 5 breaks at maximum through the full sample, the Bai-Perron method constantly chooses 5 breaks and recursively estimates the last breakpoint at October 1983, regardless of the lag specifications. In contrast to the Bai-Perron estimates, the two-stage ROC recursively estimates the last breakpoint in a near monotonically increasing trend as we move towards the last forecasting point, and it picks much later time points as the most-recent breakpoints compared with the Bai-Perron estimates.

[FIGURE 2 ABOUT HERE]

Moreover, as shown in table 3 panel B, the combination methods that average forecasts across different estimation windows also reduce the MSFEs compared with the expanding

[6]Due to this evidence , Stock and Watson (1999) believe little gain can be achieved from incorporating model instability and thus ignore structural breaks when they forecast inflation changes using the NAIRU Phillips curve.

window forecasts. The MSFE-weighted forecasts and the location-weighted forecasts are the best two among the five averaged forecasts. The ROC weights (with the heading ROC-W1) reduce MSFE close to one and the adjust-ROC weights (with the heading ROC-W2) achieve a MSFE less than one when the lag structure is chosen using information criteria.

One can argue that it is not appropriate to assume structural break uncertainty when forecasting the U.S. inflation since at least we can observe less-volatile inflation changes after 1984[7]. Therefore, knowing a possible break in 1984, we can short the estimation sample to only include post-1984 data. Thus facing unknown knowledge of breaks after January 1984, we forecast the U.S. inflation changes through July 1997 to June 2007. Table 4 shows that ignoring break possibility after January 1984 does not damage the forecasting ability of the NAIRU Phillips curve very much any more. Especially the MSFE from the model selected by AIC is about 18% smaller than the random walk forecast. The results that the the end of year 1983 is one of structural breaks that shift the parameters in the NAIRU Phillips curve model. Recall that in Figure 2 the Bai-Perron method persistently estimates October 1983 as the last breakpoint, which is the reason that we observe similar MSFEs between applying the Bai-Perron method in table 3 and the expanding window estimations in table 4.

[**TABLE 4 ABOUT HERE**]

Despite the fact that the forecasting performance has been improved after we use post-1984 data to estimate forecasting models, dealing with possible structural breaks after 1984 can still improve the forecasting performance of the NAIRU Phillips curve model to some extent. The most attractive result, that the MSFE becomes about 40% smaller than 1, comes from the single forecasts based on the ADL(1,1) models when the most recent break is estimated by the Bai-Perron method. Figure 3 plots these recursive estimates of the most recent break assume there are maximum 3 breaks in the past[8]. It shows that until the mid of year 2002, the last break is estimated around the end of 1992. Following the last break identified at late 1996 for about 2 years, march 2004 is chosen until the end of the forecasting time. Among the averaged forecasts, the location-weighted forecasts is the best. The simple average as well as our proposed ROC and adjust-ROC weighting schemes achieve the similar forecasting results. The average forecasts based on historical MSFE cannot perform better than the forecasts when possible in-sample breaks are ignored.

[7] Many literature refers this change in the mid of 1980's as a result of monetary policy shift. For instance, Clarida et al. (2000) show that the U.S. macroeconomy is more stabilized after the appointment of Paul Volker as the Fed chairman in 1979.

[8] We set a smaller maximum number of breaks since we have a shorter in-sample period.

[FIGURE 3 ABOUT HERE]

We redo the forecasting exercise, replacing the level of unemployment rates with the 12-month changes of unemployment rates. In table 5, the results from the BIC specification and ADL(1,1) are the same, which indicates that BIC constantly favors the ADL(1,1) model. Moreover, models with short lag structure now generate better forecasts than models with long lag structure, regardless of forecasting methods. It seems that when we include the lagged u_{t-12} in the model and thus can use more past information for forecasts, a model with a shorter lag length is preferred. Further, with a short lag structure, the NARIU Phillips curve outperforms the random walk model when possible breaks are ignored. When models are specified using AIC, the forecasting methods that account for structural breaks have reduced the MSFEs successfully, except for the rolling regression method. The 2-stage ROC method generates the best single forecast and the location-weight that make the average forecast heavily rely on the more current information performs the best among the five weighting schemes.

[TABLE 5 ABOUT HERE]

Table 6 records MSFEs of the forecasts based on equation (20) that is estimated employing post-1984 data. Even if the possibility of the presence of the post-1984 breaks is ignored, all of the relative MSFEs from the NAIRU Phillips curve are below 1 regardless of the lag structure. Under AIC, considering break uncertainty can hardly increase the forecasting ability, except for taking simple average of forecasts from different estimation windows or weight them according to the location of the beginning observation in each estimation window. However, with a shorter lag structure, accounting for the post-1984 break possibility can make the NAIRU Phillips curve model outperform the random walk model much more than when such a possibility is ignored. For example, with a BIC specification, the combination forecasting methods increase the forecasting accuracy of the NAIRU Phillips curve by at least 20% comparing with the expanding window forecasts. In particular, the location-weighted forecasts achieve almost 40% reduction of MSFE.

[TABLE 6 ABOUT HERE]

We conclude the followings from the empirical findings. First, unlike the small effect from incorporating model instability that Stock and Watson (1999) claim in forecasting inflation changes, we find that accounting for break occurrence in the past can improve the forecasting ability of the unemployment-based NAIRU Phillips curve model. Second, it appears that the coefficients instability is one of the reasons for the random walk model to outperform the unemployment-based NAIRU Phillips curve model. After we solve for the in-sample structural break problem, the NAIRU forecasts have achieved smaller MSFEs than the random walk forecasts. Moreover, the ROC and adjusted-ROC weighting schemes are able to reduce the MSFEs and their associated forecasts are comparable with the MSFE-weighted forecasts and the "superior" equally weighted forecasts. The adjust-ROC averaged forecasts are constantly more accurate than the ROC-weighted forecasts since it benefits from the prior belief that more current information is more useful for forecasts. The power of this prior belief has also been shown from the fact that location-based weighting scheme beats all the other sophisticated weights determined by historical forecasting performance or the ROC structural break tests.

5 Conclusion

Financial and macroeconomic time series are often found to be subject to parameter instability reflecting policy changes or regime switches. Ignoring the presence of breaks may produce biased forecasts that are generated based on parameter estimations. However, the information of parameter shifts in the past sample is unlikely to be known to forecasters. Although a large number of papers provide techniques for testing for structural breaks, far less discusses the use of test results in the context of forecasting. If a break test rejects the parameter constancy, forecasters may simply estimate the forecasting model using the data subsequent to the estimated last break; otherwise, they conduct a full sample estimation. Instead of following the traditional approach, this paper has proposed a new forecasting combination method that utilizes break test statistics. Under structural break uncertainty, we consider each in-sample time as the possible most-recent breakpoint and estimate forecasting models excluding data prior to each time point. The weights depend on the probability of each time point being the most-recent break date and its proxy is determined by the reversed ordered Cusum squared statistics with a prior probability of the break location.

The other weighting scheme proposed in this paper simply assumes that the probability for each time point being the most-recent break is proportional to the location of the time point in the full sample period. Under this weighting scheme, we essentially places more weights on the forecasts using more recent information. Through the Monte Carlo simulations, we have examined the forecasting performance of two new combination methods, as well as a range of alternative forecasting techniques that account for break uncertainty. It includes the two-stage ROC, the cross-validation and the Bai-Perron method that produce single forecasts as well as the forecasts averaged by equal weights and historical accuracy-based weights. Our results support the new weighting schemes base on the ROC statistics, particularly the adjusted-ROC that puts heavier weights on the forecasts using the samples that ROC statistics indicate to be subsequent to the most-recent break and include more recent information. The forecasting performance of the location-weighted forecasts is also shown to be outstanding.

We provide empirical evidence of the benefits obtained from using the proposed combination methods as well as the other alternative methods to reexamine the U.S. inflation forecasting ability of the NAIRU Phillips curve. The inflation forecasts after taking in-sample structural break into account are shown to be more accurate than the random walk forecasts that Atkeson and Ohanian (2001) prefer. Our results also indicate the necessity of considering the presence of structural breaks when we forecast the U.S. inflation changes using the NAIRU Phillips curve.

It is interesting to see that both our Monte Carlo and empirical experiments show a powerful forecasting ability of the location-based combination method that imposes more weights on the forecasts using more recent sample periods. This weighting scheme not only beats the other more sophasticated weighting methods that utilize information from the structural break test or the past forecasting performance, but also outperforms the simple average method that has been commonly used and found hard to beat. A detailed investigation on this location-based weighting technique can be expected in the future research.

References

Atkeson, A. and L. Ohanian (2001): "Are Phillips Curve Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

Bai, J. and P. Perron (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1–22.

Brown, R. L., J. Durbin, and J. M. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society, Series B (Methodological)*, 37, 149–192.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

Clark, T. E. and M. W. McCracken (2005): "The Power of Tests of Predictive Ability in the Presence of Structural Breaks," *Journal of Econometrics*, 124, 1–31.

——— (2006): "Averaging Forecasts from VARs with Uncertain Instabilities," Conference Paper.

Fisher, J. D. M., C. T. Liu, and R. Zhou (2002): "When Can We Forecast Inflation," *Economic Perspectives*, 30–42.

Gordon, R. J. (1997): "The Time-Varying NAIRU and Its Implications for Economic Policy," *The Journal of Economic Perspectives*, 11, 11–32.

Pesaran, M. H. and A. Timmermann (2002): "Market Timing and Return Prediction Under Model Instability," *Journal of Empirical Finance*, 9, 495–510.

——— (2007): "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, 137, 134–161.

Staiger, D., J. H. Stock, and M. W. Watson (1997): "The NAIRU, Unemployment and Monetary Policy," *The Journal of Economic Perspectives*, 11, 33–49.

Stock, J. H. and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

——— (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

Timmermann, A. (2005): "Forecast Combinations," Forthcoming in Handbook of Economic Forecasting.

Tobin, J. (1972): "Inflation and Unemployment," *The American Economic Review*, 62, 1–18.

Figure 1: 12-month changes of U.S. annual core CPI inflation

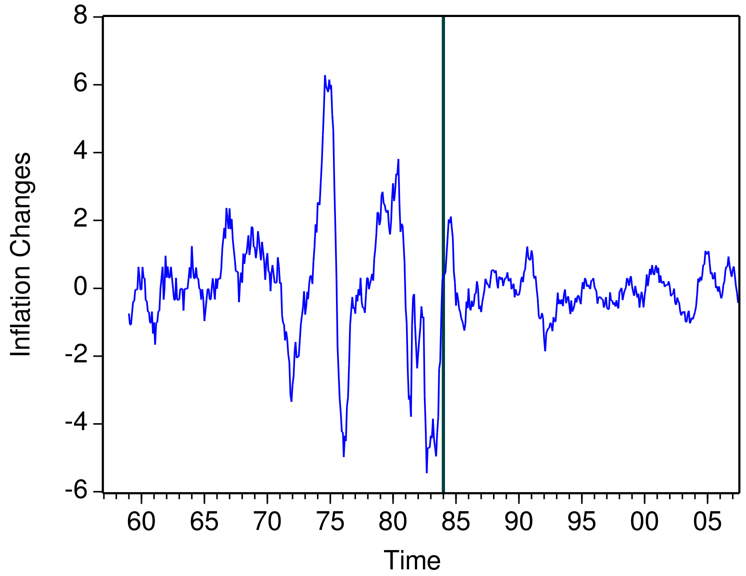

Figure 2: The recursively estimated last breakpoints by the Bai-Perron method and the 2-Stage ROC

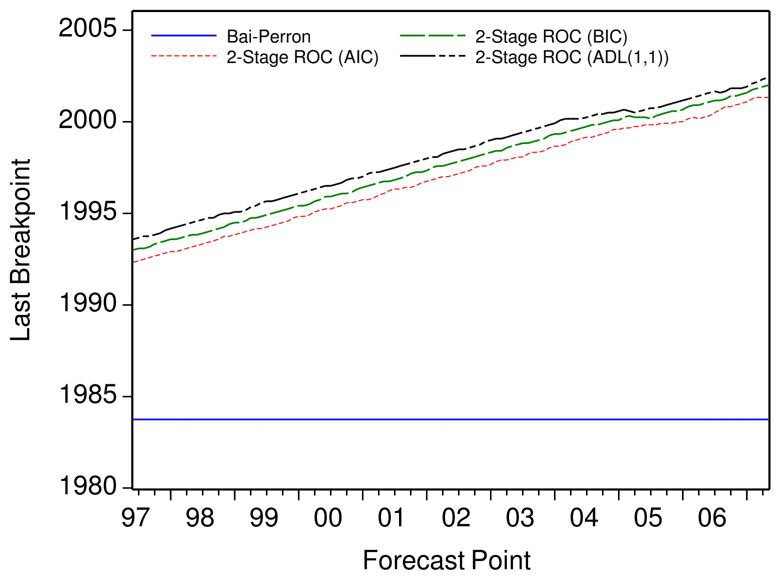

Figure 3: The recursively estimated last breakpoints by the Bai-Perron method

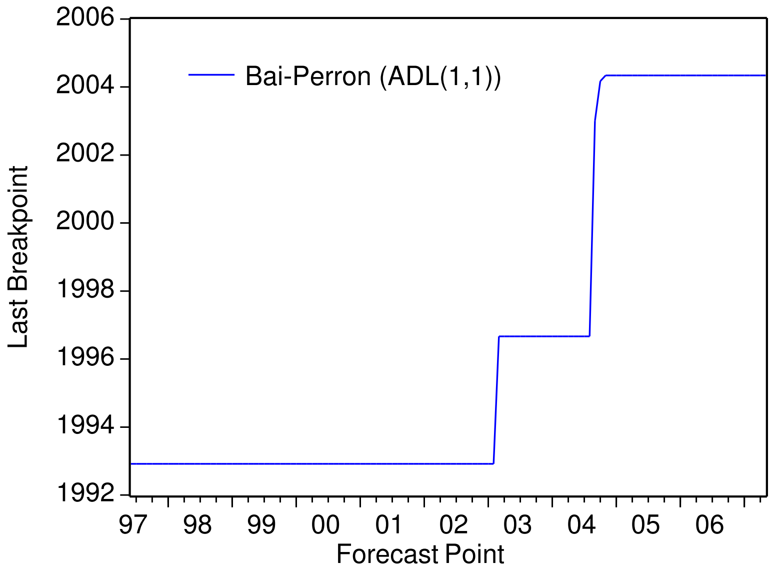

Table 1: MSFE values under multiple structural breaks[a]

						Single forecast			Forecast combination			
d_{11}	d_{12}	d_{22}	d_{11}^*	d_{12}^*	d_{22}^*	Cross-validation[b]	2-stage ROC	Bai-Perron	Equal-W	MSFE-W[b]	ROC-W1	ROC-W2
-0.4	0	0	-0.4	0	0	0.5100	0.7978	0.4811	0.5160	0.5965	0.5188	0.4752
-0.2	0	0	-0.4	0	0	0.6115	0.9148	0.4894	0.5891	0.7134	0.6029	0.5424
-0.4	0	0	-0.2	0	0	0.5138	0.7822	0.5528	0.5216	0.5550	0.5129	0.4934
-0.4	0	0	0.2	0	0	0.8134	0.9210	0.9030	0.7578	0.7851	0.7825	0.7751
-0.4	0	0	0.4	0	0	0.8588	0.9625	0.7422	0.8396	0.9689	0.8775	0.8012
0	1	0	0	1	0	0.3521	0.8053	0.2434	0.3712	0.4720	0.3635	0.3005
0	-0.4	0	0	-0.4	0	0.7480	0.8775	0.8123	0.7107	0.7753	0.7178	0.6951
0	0	0	0	0	0	1.0270	1.0007	1.3624	1.0321	1.0100	1.0349	1.0624

[a] We set the values of parameters to be $a_{11} = 0.9, a_{12} = 1, a_{22} = 0.9$; The in-sample observation is $T = 100$. The true locations of the two structural breaks are $T_1 = p1 \times T$, where $p1 = 0.25$ and $T_2 = p2 \times T$ where $p2 = 0.75$. We set that the minimum estimation window to be $\bar{w} = 10$.

[b] We set the test sample size $\tilde{w} = 25$.

[c] The maximum number of breaks is set to be 5.

Table 2: MSFE values of the location-weighted forecasts under multiple structural breaks[a]

						Forecast Combination
d_{11}	d_{12}	d_{22}	d^*_{11}	d^*_{12}	d^*_{22}	Location-Weighted
-0.4	0	0	-0.4	0	0	0.4530
-0.2	0	0	-0.4	0	0	0.5072
-0.4	0	0	-0.2	0	0	0.4898
-0.4	0	0	0.2	0	0	0.7608
-0.4	0	0	0.4	0	0	0.7527
0	1	0	0	1	0	0.2818
0	-0.4	0	0	-0.4	0	0.6871
0	0	0	0	0	0	1.0747

[a] The basic setup is the same as in table 1.

Table 3: MSFEs of Phillips-curve based inflation forecasts using full sample data[a]

Panel A: Single Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Expanding Window	1.7855	1.8698	2.3060
Rolling Window	1.9539	2.0957	2.5374
Cross-validation[b]	0.7656	0.8098	0.9453
2-stage ROC	0.7640	0.7728	1.0008
Bai-Perron Method[c]	0.8883	0.9128	1.0139
Panel B: Combined Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Simple average	1.0876	1.1683	1.4168
MSFE-weighted[b]	0.8735	0.9409	1.0439
ROC-W1	1.0032	1.0992	1.3736
ROC-W2	0.9183	0.9798	1.1357
Location-weighted	0.8806	0.9068	1.0199

[a] All of the MSFEs are relative to the benchmark forecasts where the expected changes of 12-month inflations are zero. Information from 1959:01 to 1997:06 are used for the initial forecasting model estimation. We use the level of unemployment rates as one of the regressors in forecasting models, following Stock and Watson (1999), Atkeson and Ohanian (2001) and Fisher et al. (2002). The minimum estimation window is twice as the number of estimated coefficients.

[b] We include 50 observations in the test sample.

[c] When applying Bai-Perron Method, we assume 5 maximum breaks during the estimation sample.

Table 4: MSFEs of Phillips-curve based inflation forecasts using subsample data after 1984[a]

Panel A: Single Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Expanding Window	0.8279	0.9203	1.0237
Rolling Window	1.0232	1.0191	1.0963
Cross-validation[b]	0.7429	0.8049	0.9375
2-stage ROC	0.7805	0.9294	0.9247
Bai-Perron Method[c]	0.9116	0.8423	0.5838

Panel B: Combined Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Simple average	0.7636	0.8177	0.8517
MSFE-weighted[b]	0.8596	0.9058	0.9998
ROC-W1	0.7785	0.8392	0.8934
ROC-W2	0.7591	0.8304	0.8841
Location-weighted	0.7348	0.7893	0.7705

[a] All of the MSFEs are relative to the benchmark forecasts where the expected changes of 12-month inflations are zero. Information from 1984:01 to 1997:06 are used for the initial forecasting model estimation. We use the level of unemployment rates as one of the regressors in forecasting models, following Stock and Watson (1999), Atkeson and Ohanian (2001) and Fisher et al. (2002). The minimum estimation window is twice as the number of estimated coefficients.

[b] We include 50 observations in the test sample.

[c] When applying Bai-Perron Method, we assume 3 maximum breaks during the estimation sample.

Table 5: MSFEs of Phillips-curve based inflation forecasts using full sample data[a]

Panel A: Single Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Expanding Window	1.2803	0.8104	0.8104
Rolling Window	1.4112	0.9253	0.9253
Cross-validation[b]	0.8451	0.6790	0.6790
2-stage ROC	0.7049	0.6797	0.6797
Bai-Perron Method[c]	1.0054	0.9528	0.9528

Panel B: Combined Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Simple average	0.8672	0.7133	0.7133
MSFE-weighted[b]	0.8770	0.7314	0.7314
ROC-W1	0.9295	0.8139	0.8139
ROC-W2	0.8980	0.8110	0.8110
Location-weighted	0.7638	0.6952	0.6952

[a] All of the MSFEs are relative to the benchmark forecasts where the expected changes of 12-month inflations are zero. Information from 1959:01 to 1997:06 are used for the initial forecasting model estimation. We use the monthly changes of unemployment rates as one of the regressors in forecasting models. The minimum estimation window is twice as the number of estimated coefficients.

[b] We include 50 observations in the test sample.

[c] When applying Bai-Perron Method, we assume 5 maximum breaks during the estimation sample.

Table 6: MSFEs of Phillips-curve based inflation forecasts using subsample data after 1984[a]

Panel A: Single Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Expanding Window	0.8576	0.9415	0.9424
Rolling Window	0.9642	0.8524	0.8694
Cross-validation[b]	0.8971	0.6297	0.6550
2-stage ROC	1.1371	0.8017	0.8448
Bai-Perron Method [c]	0.7688	0.5288	0.6187

Panel B: Combined Forecasts			
Forecasting methods	The Choice of Lag Structure		
	AIC	BIC	1 Lag
Simple average	0.7953	0.6245	0.6813
MSFE-weighted [b]	0.8934	0.7131	0.7649
ROC-W1	0.8994	0.7211	0.7799
ROC-W2	0.8748	0.6924	0.7498
Location-weighted	0.7684	0.5620	0.6160

[a] All of the MSFEs are relative to the benchmark forecasts where the expected changes of 12-month inflations are zero. Information from 1984:01 to 1997:06 are used for the initial forecasting model estimation. We use the changes of unemployment rates as one of the regressors in forecasting models. The minimum estimation window is twice as the number of estimated coefficients.

[b] We include 50 observations in the test sample.

[c] When applying Bai-Perron Method, we assume 3 maximum breaks during the estimation sample.