Structural breaks and long memory in US inflation rates: Do they matter for forecasting?

Namwon Hyung*

Tinbergen Institute, Erasmus University Rotterdam

Philip Hans Franses

Econometric Institute, Erasmus University Rotterdam

Econometric Institute Research Report EI2001-13

April 10, 2001

Abstract

There is substantial evidence that several economic time series variables experience occasional structural breaks. At the same time, for some of these variables there is evidence of long memory. In particular, it seems that inflation rates have both features. One cause for this finding may be that the two features are difficult to distinguish using currently available

*Corresponding author. Address: P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, E-mail: hyung@few.eur.nl. Details on all empirical results, reported and unreported, can be obtained from the corresponding author.

econometric tools. Indeed, various recent studies show that neglecting occasional breaks may lead to a spurious finding of long-memory properties. In this paper we focus on this issue within the context of out-of-sample forecasting. First, we show that indeed data with breaks can be viewed as long-memory data. Next, we compare time series models with structural breaks, models with long-memory and linear autoregressive models for 23 monthly US inflation rates in terms of out-of-sample forecasting for various horizons. A key finding is that the linear models do not perform as well as the other two, and that the model with breaks and the model with long memory perform about equally well. We also examine their joint performance by combining the forecasts. A by-product of our empirical analysis is that we can relate the value of the long-memory parameter with the number of detected breaks, in which case we find a strong positive relationship.

KEY WORDS: Occasional Breaks; Long Memory; US Inflation Rates; Forecast Performance

1 Introduction

Several economic time series seem to experience occasional structural breaks. Some of these series also display long memory, as evidenced by the empirical suitability of certain long-memory models. Recent studies show that time series processes with occasional breaks may generate a long-memory effect in the autocorrelation function. There are also studies which suggest that data from one model can be captured in a model of the other type, and the other way around. Of course, the key question remains whether it matters for forecasting which model is used? Or otherwise, should we pay effort in making a formal distinction between the two models, or should we perhaps seek for a model that contains the two features? Another related question concerns the best modeling and forecasting strategy for a given forecasting horizon. Indeed, it may be that one model is better for short-term forecasts, while another is better for the longer horizon.

Long memory implies that shocks have a long-lasting effect. A popular application of long-memory time series models concerns inflation. There is substantial evidence that inflation rates have long memory, a feature which can be captured by a fractionally integrated I(d) model, see Hassler and Wolters (1995), Bos, Franses and Ooms (2001) and several papers cited in Baillie (1996). Alternatively, US inflation rates may perhaps show long memory because of the presence of neglected occasional breaks in the series rather than that they are really I(d). In this paper, we examine whether evidence of long memory in various inflation rates has any correlation with occasional breaks in terms of in-sample estimation and out-of-sample forecastability. We show that there might exist a long-memory component in US CPI inflation rates because of occasional breaks. This long-range dependence, however, makes no significant difference between occasional breaks model and I(d) model in terms of out-of sample forecasting performance.

The paper is organized as follows. In Section 2 we discuss the two models, and we show that it may not be easy to make a formal distinction between long memory and breaks in practice. The estimation methods and methods to generate forecasts are presented in Section 3. In Section 4, we fit both models to 23 US inflation series in order to see which does better in terms of out-of-sample forecasting, where we also include a regular autoregressive model for comparison. We find that almost all autoregressive models can be improved, but that there is not much difference in performance between the two models of our interest. This leads to several suggestions for further research, as we outline in Section 5.

2 Long-Memory and Occasional Breaks

In this section we discuss a class of models which can usefully describe long-term persistence of shocks. The ARFIMA model, or simply I(d) model, which will be discussed further below, has been reguarly used to describe the persistence of economic and financial data, see for example articles in the special issue of the *Journal of Econometrics*, 73, 1996. Inference on the dynamics of such a time series is usually based on the autocorrelation function, where its decay pattern measures the persistence of the process. Interestingly, a typical long-memory decay pattern can also be generated by neglected structural breaks. However, the theoretical and empirical econometric literature on long memory and on structural change have evolved independently. Recently, there have appeared studies which explain how infrequent stochastic breaks can create strong persistence in the autocorrelation structure, see Mikosch and Stărică (1999), Granger and Hyung (2000), and Diebold and Inoue (2001). To illustrate, we review the Occasional-Break model and the I(d) model briefly and we show that occasional structural change is approximately observationally equivalent to long memory.

Consider a simple model with R occasional breaks in the mean, that is,

$$
\begin{aligned}
y_t &= \mu_0 + u_t, \quad \text{if } 0 < t \leq k_1 \\
y_t &= \mu_1 + u_t, \quad \text{if } k_1 < t \leq k_2 \\
& \dots \\
y_t &= \mu_R + u_t, \quad \text{if } k_R < t \leq T
\end{aligned} \tag{1}
$$

where u_t follows an ARMA process

$$\Phi(L)\, u_t = \Theta(L)\, \varepsilon_t \tag{2}$$

with ε_t is $i.i.d(0, \sigma_\varepsilon^2)$ for $t = 1, ..., T$, and with $\Phi(L) = \left(1 - \phi_1 L - \cdots - \phi_p L^p\right)$ and $\Theta(L) = (1 + \theta_1 L + \cdots + \theta_q L^q)$ are polynomials of finite orders p and q in the usual lag operator L. We assume all roots of $\Phi(z) = 0$ and $\Theta(z) = 0$ to be outside of the unit circle.

We use this model to examine 'spurious' long-memory effects caused by neglected occasional breaks and its consequence for forecasting. Of course, for our empirical data we need to estimate R, see Section 3 below. To examine the effects of neglecting occasional breaks in a more theoretical way, a slightly different version of (1) - (2) turns out to be more useful. This model is,

$$y_t = m_t + u_t, \tag{3}$$

where u_t is a noise variable, and occasional level shifts, m_t, are controlled by two variables q_t (date of break) and η_t (size of jump), as

$$m_t = m_{t-1} + q_t \eta_t = m_0 + \sum_{i=1}^{t} q_i \eta_i, \tag{4}$$

where η_t is $i.i.d(0, \sigma_\eta^2)$. We assume that q_t follows an $i.i.d.$ binomial distribution, that is,

$$q_t = \begin{cases} 0, \text{ with probability } 1-p \\ 1, \text{ with probability } p \end{cases} \quad (5)$$

Combining (3) with (4) yields occasional level shifts in the mean of y_t, which is thus represented by

$$y_t = \{m_0 + q_1\eta_1 + \cdots + q_t\eta_t\} + u_t \quad (6)$$

The expected number of breaks for a given sample is Tp. Note that two models in (1) and (6) can be equivalent by setting $m_0 = \mu_0$, $Tp = R$, and so on.

Now we turn to a description of a long-memory time series model. The ARFIMA (p, d, q) model is widely used for series with long memory, and usually it is referred to as the I(d) model. In this paper, we consider the following specification, that is,

$$\Phi'(L)(1-L)^d(y_t - \mu) = \Theta'(L)\varepsilon_t, \quad (7)$$

with similar conditions for $\Phi'(L)$ and $\Theta'(L)$ as (2). For any real $d > -1$, the fractional difference operator, $(1-L)^d$, is defined by its Maclaurin series, that is,

$$(1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j, \ \pi_j = \frac{j-1-d}{j}\pi_{j-1}, \ \pi_0 = 1. \quad (8)$$

From this expression we can infer already a potential slow decay in the autocorrelation function. The ARFIMA model is a flexible statistical tool to describe the behavior of the autocorrelations at high lags, that is, the long-term persistence. The long-term decay is solely determined by d, which is also called the memory parameter. It can be shown that the autocorrelation function ρ_k, is proportional to k^{2d-1} as $k \to \infty$. Consequently, the autocorrelations of this model decay hyperbolically to zero as $k \to \infty$, in contrast to the exponential decay in a stationary ARMA model.

One way to estimate the value of d is to use the Geweke and Porter-Hudak (1983) (henceforth GPH) method, which is based on the following regression, that is,

$$\ln\{I(\omega_j)\} = c - d\ln\left\{4\sin^2(\omega_j/2)\right\} + u_j,\ j = 1, ..., g(T), \tag{9}$$

where $I(\omega_j) = \frac{1}{2\pi}\left|\sum_{i=1}^{T} y_t \exp(i\omega_j t)\right|^2$ is the periodogram at frequency $\omega_j = 2\pi j/T$, which depends on the sample size T. Often, $g(T) = T^{1/2}$ is used.

It can be proved that it is not easy to make a formal distinction between the break model and I(d) model in practice. The formal proof of "observational equivalence" concerns rare breaks, achieved by letting the break probability to shrink with T, which is similar in spirit to a "local to unity" asymptotic analysis. The key idea is to let p decrease with the sample size, so that regardless of the sample size, realizations tend to have the same number of breaks. For any fixed sample size T, realizations tend to have just a few breaks.

Assume that $p \to 0$ as $T \to \infty$, and that $lim_{T\to\infty}Tp = R$, where R is non-zero finite constant. Let y_t be a series from (3), (4) and (5) and we assume u_t in (3) is $i.i.d(0, \sigma_u^2)$ for simplicity, then the estimated value of d using GPH method with $g(T) = T^{1/2}$, $\hat{d}_{GPH} \xrightarrow{p} d^*$such that $0 < \left(1 + 4\pi^2\frac{\sigma_u^2}{R\sigma_\eta^2}\right)^{-1} < d^* < 1$ as $T \to \infty$, see Granger and Hyung (2000) for details. It is straightforward to show that this also holds for other estimation methods, like for example the Kim and Phillips (2000) modified log periodogram (henceforth MLP) regression estimator. When the values of R and σ_η^2 increase, the estimated memory parameter d deviates further from zero. More breaks or a larger size of the break cause more persistent memory of the series. For example, increases in R or in σ_η^2 make this process to become more similar to an I(1)process. In the empirical section below, we will examine if the supposed link between R and d holds in practice.

The reverse direction of "observational equivalence" can be understood as well. The long memory of a time series causes spurious breaks, at least, when relying on standard estimation methods of breaks. Unlike I(0) processes, I(d) or I(1) processes have different effects on the estimated number of breaks. When the DGP is an integrated or fractionally integrated series without breaks, spuriously many breaks (but with different numbers depending on the value of d) are inferred, see Nunes, Kuan and Newbold (1996) for the I(1) case. Suppose the DGP of y_t is given by (7) without break, one finds a break point near the middle of the time series when $d > 0$ by the Schwarz-Bayesian criterion (SBC). Simulations in Granger and Hyung(2000) suggest a positive relation between the number of breaks and the value of d in a finite sample.

In sum, it seems that long-memory models and models with occasional breaks can capture the same kind of phenomena, even though their mathematical expressions are rather different.

3 Estimation and Forecasting methods

As we aim to compare the two models of interest in practical situations, we now turn to a discussion of estimation and forecasting methods in this section.

3.1 The Model with Breaks

In this section we focus on how to identify break points with dynamic components in a parametric model, for example by introducing lagged dependent variables so as to have an autoregressive model. Next we discuss how to select the order of the autoregressive model. As indicated in Pesaran and Timmermann (2000), little is known about the properties and optimality of alternative forecasting methods under breaks. They discuss how to optimally combine pre-break and post-break data in the mean squared error sense, and find that forecasting accuracy can be improved by pre-testing for a structural break. In the present paper, we abstain from a selection of an optimal sample size to forecast out of sample when structural breaks have in fact occurred.

We use the following ARMA representation.

$$y_t = \mu_t + u_t, u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{10}$$

and we assume there are R breaks in μ_t as in (1). We can transform this model to

$$\phi(L) y_t = \phi(L) \mu_t + \varepsilon_t, \tag{11}$$

where $\phi(L) = \psi(L)^{-1}$. For practical purposes, we restrict this infinite order model to

$$y_t = \mu_t + \phi_1(y_{t-1} - \mu_{t-1}) + \cdots + \phi_p(y_{t-p} - \mu_{t-p}) + \varepsilon_t, \tag{12}$$

which we will call as the BREAK model. We use Bai and Perron (1998)'s sequential procedure to estimate the breaks. To select a proper lag in (12), we use the following algorithm.

(a) We estimate the number of breaks by Bai and Perron (1998)'s sequential procedure, while assuming certain properties of the residuals, see Assumptions A4 in Bai and Perron (1998). The dynamic effects are incorporated in an indirect nonparametric approach, which leaves the dynamics in the disturbances and applies a nonparametric correction for serial correlation in the residuals.

(b) After filtering out the breaks from the series, we can estimate the ARMA structure in the residual series $\hat{u}_t = \hat{y}_t - \hat{\mu}_t$ in (10) and we choose the number of lags p by minimizing AIC.

(c) Once p has been determined, we estimate the number of breaks using the sequential procedure again, without a serial correlation assumption in the errors, but where lagged dependent variables are allowed as regressors in (12).

It seems useful at this stage to consider Bai (1997)'s repartition estimation procedure. When the first break point is identified at k, the whole sample is divided into two subsamples with the first subsample consisting of $k \in (1, T)$ observations and the second subsample consisting of the rest of the observations $(T - k)$. One then estimates a break point for the subsample where a hypothesis test of parameter constancy fails. Divide the corresponding subsample further into subsamples at the newly estimated break point, and perform parameter constancy tests for the hierarchically obtained subsamples. This procedure is repeated until parameter constancy is not rejected for all subsamples. Bai (1997) shows how the repartition procedure coupled with hypothesis testing can yield a consistent estimate for the true number of breaks. The details of the test statistic and methods are given in Bai (1997).

The number of estimated breaks can be different across the Bai (1997) and Bai and Perron (1998) methods. Some of our unreported empirical results show that Bai (1997)'s repartition method often gives a much larger number of breaks as compared to Bai and Perron's sequential method. Also the estimation results under different assumptions of dynamic effects (parametric or nonparametric assumptions in (a) and (c)) are a little different. However, methods of endogenous determination of breaks are prone to over-estimate the number of breaks in the presence of nonlinearity, such as smooth transition, nonlinear trend, and so on. There is another possibility of overfitting under the existence of heteroskedasticity. Therefore, we use the estimation results from algorithm (c), which is giving a conservative estimate of the number of breaks.

We must mention here that we also considered in our empirical analysis the so-called STOPBREAK model, developed in Engle and Smith (1999). This model includes an endogenous smooth transition function to indicate structural breaks in (3), that is,

$$q_t(\gamma) = \frac{(u_t + \cdots + u_{t-s})^2}{\gamma + (u_t + \cdots + u_{t-s})^2}, \gamma > 0, s > 0 \tag{13}$$

with $m_t = m_{t-1} + q_{t-1}u_{t-1}$. Smith (1999) generalizes this process by allowing past deviations from m_t to have an effect on short-horizon forecasts. Since m_t represents the long-run forecast of y_t, given information up to time $t-1$, these past deviations take the form $y_{t-i} - m_{t-i+1}$. Specifically, the model is

$$y_t = m_t + \alpha(L)(y_{t-1} - m_t) + u_t \tag{14}$$

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_p L^p$ and $\alpha(1) < 1$. As this model can be seen as a serious contender to the BREAK model in (12), we tried to estimate the parameters for the empirical data to be introduced above. However, we found that the available estimation method gave very unstable and inaccurate results, and in fact did so far a range of specifications of (13). Hence, we further abstained from considering this model.

3.2 The I(d) Model

Geweke and Porter-Hudak (1983) show that, when attention is confined to frequencies near zero, the differencing parameter can be estimated consistently from the least-squares regression in (9). However this method is only valid when $d < \frac{1}{2}$. In our empirical examples of inflation rates, there is no reason to believe $|d| < \frac{1}{2}$, and hence we have to allow for nonstationarity of this series by allowing that $d > \frac{1}{2}$ is also possible. Similar to the BREAK model we adopt a two-step estimation method. We use the following procedure to estimate d and ARMA parameters.

(a) Estimate $\hat{d}$ by using Kim and Phillips (2000)'s **modified log periodogram regression**, which is useful in cases where nonstationarity is suspected.

(b) Calculate $\tilde{y}_t = \sum_{j=0}^{t-1} \tilde{\pi}_j (y_{t-j} - \hat{\mu})$, $t = 1, ..., T$, where $\tilde{\pi}_j$ is calculated from (8) using $\hat{d}$.

(c) Estimate the ARMA parameters of $\Phi'(L)\tilde{y}_t = \Theta'(L)\varepsilon_t$ for the time series $\tilde{y}_1, \ldots, \tilde{y}_T$. This can be done by using MLE or least-squares regression.

The consistency and the central limit theorem of this numerical algorithm can be proven by using Beran (1995)'s Theorem 1 and the results in Kim and Phillips (2000). The estimated value of d is robust against serial correlation which is shown in Kim and Phillips (2000). Upon using the estimated value $\hat{d}$, we can estimate the ARMA coefficients consistently, which is due to the fact that the ultimate decay of the autocorrelations and the coefficients of the AR(∞) representation is exponential.

For comparison we also used the frequency domain MLE (see Beran (1994) and Breidt, Crato and de Lima (1998)) and Beran (1995)'s approximate maximum likelihood method. Note that the methods of Geweke and Porter-Hudak (1983) and frequency domain MLE are based on assumption of $|d| < \frac{1}{2}$, but approximate MLE of Beran (1995) and the MLP regression estimator are consistent for the nonstationary region too ($d > \frac{1}{2}$).

The assessment of the goodness of fit of a fractionally differenced ARMA model is not straightforward, as pointed out by Hosking (1984). The ARFIMA model allows for jointly modeling the short-term and long-term behavior of a time series. The AIC, which concentrates on the short-term forecasting ability of the fitted model, may not be the best criterion for all applications involving fractional differencing, but for the present purpose we do implement it as

$$AIC = \ln \hat{\sigma}_{\varepsilon}^2 + 2(p + q + \delta_d)/T$$

where δ_d is 1 if the d parameter is estimated and 0 if it is fixed at zero.

For out-of-sample forecasting, we consider the "naive" method. We compute h-step ahead forecasts recursively, using $\hat{\pi}(L)$ found from:

$$\hat{\pi}(L) = \hat{\Theta}(L)^{-1} \hat{\Phi}(L) (1 - L)^{\hat{d}}$$

where $\hat{\pi}(L) = \sum_{j=0}^{\infty} \hat{\pi}_j L^j$, and hence

$$\hat{y}_{T+h} = \sum_{j=0}^{T-1} \hat{\pi}_{j+h} \left(y_{T-j} - \hat{\mu} \right), \; h = 1, 2, ...$$

This naive method can also be applied to nonstationary series with $d > \frac{1}{2}$.

4 Forecasting US Inflation Rates

In this section we use the BREAK model and the I(d) model for forecasting inflation rates. We consider 23 monthly US Consumer Price Index series which are randomly chosen from the U.S. City Average data set. The sample period covers from 1967:01 - 2000:08 and the base years are 1982 - 1984. All series are considered in seasonally adjusted format. Inflation rates are constructed from the price indices by taking 100 times the first differences of the logarithmic transformed series. In Table 1 we present the list of variables, where we approximately arrange the series from highly aggregated to less aggregated series.

First, we estimate the parameters and we use the estimated values for all out-of-sample forecasts. We call this the static method. We choose 1967:01-1990:12 as the in-sample period, which has 288 observations, and hence the out-of-sample period is 1991:01-2000:08 with 116 observations. Second, we re-estimate the model parameters and generates forecasts when a new observation is added to series. We call this the recursive method. For the recursive method, we estimate the parameters of the three models sequentially, starting with sample 1967:01-1990:12 and ending with sample 1967:01-2000:07. We choose lag p of each model using AIC for the short-sample period and we fix this value in the subsequent analysis. For Series 18 (Alcoholic beverages), results are based on after removing 3 outliers.

4.1 Empirical Results from In-sample Estimation

The number of breaks in each series is estimated using Bai and Perron (1998)'s sequential method. First, we estimate breaks by the sequential method using non-parametric serial correlation correction. After filtering out break components, we fixed the AR orders in the filtered residuals by minimizing AIC. Next, using the pre-determined AR order to correct serial correlation, we apply Bai and Perron's method again. We allow up to 10 breaks, and the minimum length between breaks is set at 12 observations (1 year) to reduce the possibility that any heteroskedasticity in a series is mistaken for a break.

For an example of break analysis, let us examine the results for Series 1 (All Items). There are four breaks in the series for the full sample period (three breaks are obtained for the shorter sample period), see the third column of Table 2. Interestingly, two out of four breaks (1973:July and 1982:July) are similar to the break dates of Bos, Franses and Ooms (2001), in which they only provide a historical explanation of their choice of two break dates. Our analysis additionally detects one more regime between the two oil shocks from October 1974 to September 1978 (see Figure 1).

As we expand the size of sample, we estimate more breaks except for Series 13. From this result we expect that recursive estimation of the occasional breaks model would perhaps provide better forecasts. If there are more breaks after the in-sample period, forecasts from the occasional breaks model will not be accurate when using the static method. In the next subsection we will elaborate on the difference between the two approaches.

For the estimation of the long-memory parameter, we used four methods: the method of Geweke and Porter-Hudak (1983), of Kim and Phillips (2000), the frequency domain MLE and Beran (1995)'s method. GPH, Kim and Phillips (2000) and Beran (1995)'s methods provide similar results. In theory, GPH is valid for $|d| < \frac{1}{2}$, but it appears to be quite robust for the non-stationary region too, at least in our empirical analysis. Recall that the frequency domain MLE uses the spectral density and is defined for the range $-\frac{1}{2} < d < \frac{1}{2}$ only.

Although this method is an efficient estimation method, it is not recommended to use in this context as we cannot impose an a prior plausible restriction on the memory parameter of inflation rate.

We estimate d for the shorter sample period and compare it with the estimated values for the full-sample period, see Table 2 for the estimated values of d using Kim and Philips (2000)'s method, which turns out to be the most reliable method. The estimated memory parameter of "(1) All Items" has the highest value of d among CPIs (regardless of estimation methods). CPIs from the service sector and the energy sector show a higher memory parameter as compared to the estimated memory parameters of food, agricultural or industrial products.

Series 15 - 23 have a relatively small estimated number of breaks, which suggests a possible relationship between breaks and the degree of aggregation. A by-product of our empirical analysis is now that we can correlate the value of the long-memory parameter with the degree of aggregation. Although we do not use a rigorous classification of aggregation, we could however assume that the number of the time series in Table 1 correlates strongly with its aggregation level. When we regress $\hat{d}$ of series i on an intercept and the series number i, we get a t-statistic of -3.22 for the slope coefficient. This suggests that the memory of a series is getting more persistent for higher aggregated series.

Furthermore, Table 2 also suggests that there is perhaps a link between the estimated number of breaks R and d. Indeed, when we regress $\hat{d}$ on an intercept and this number $\hat{R}$, we get a t-value of 3.11, which seems to agree with the conjecture in Granger and Hyung (2000). If we further consider the size of break by taking the average of squared jump size of breaks, and regress $\hat{d}$ on an intercept and the number of breaks times the size of break, we obtain a positive slope coefficient but it is not significant (t-value is 1.29).

4.2 Out-of-sample Forecasts

We now turn to an evaluation of the forecasting performance of various models. We compare predictability of the BREAK model, I(d) model and a linear AR model for an I(0) series for the 23 US inflation series, where we set the forecast horizon h at 1,3,12 and 24. Tables 3 and 4 present the results of point forecasts and cumulative predictions when we use the static method. Tables 5 and 6 present the results for the recursive method. Finally, in Table 7 we present encompassing test results of one-step ahead forecasts for the static and recursive cases

We compute the root mean squared of forecast errors (RMSFE). For comparison, we report the ratios of the RMSFEs for the BREAK and I(d) model over the RMSFE of linear AR models. We find that the occasional breaks model and fractionally integrated model perform better than almost all AR models, which can be noticed from the fact that the ratios are almost always below 1. There is however not much difference in performance between the two models of interest, I(d) model and Break model. The results in Tables 3 to 6 show that one model is not dominated by the other model for all series. By the way, we also estimated an I(1) specification for some series, but we find that forecasts are not better compared with the I(0) specification.

Typically, the long-run forecasts of linear AR models are changing little when new observations are added, due to the fact that these forecasts converge to the unconditional mean. If there is a changing mean, forecasts of linear models suffer from bias and consequently would produce poor forecasts. Given the presence of structural breaks, and assuming it is easy to detect, the occasional break model can give better forecasts than a linear AR model. The break model can adjust its forecasts immediately when structural changes are detected. However, the occasional breaks model can be worse than the linear AR model when there exist the nonlinear trending component, the changing variance or a recent break (until we accumulate enough information, we do not know if break occurred or not), see for example Series 2, 9 or 16. Typically, after a structural break, the forecasts from a BREAK model are not accurate for a while, simply because the break is not evident with only a few observations, see Figure 2, where we display the forecast errors for Series 2.

In contrast, the I(d) model allows for a long-term deviation from the long-run mean, due to the fact that the I(d) model uses all past information. It has slow hyperbolic autocorrelation and impulse response decay, compared to the faster geometric decay of ARMA processes. Hence, the I(d) model is supposed to have quite a long-run predictability because it can capture long-run components from the series. Indeed we often find that the I(d) model performs well in Tables 3 and 4, in fact often better than the BREAK model.

Given the different break numbers in Table 2 for the short-sample and full-sample data, the recursive estimation should improve the forecastability of occasional breaks model. It turns out that the occasional breaks model can detect short-run components well, and that this gets bettter if the model is updated using recent information. Hence, the model performs relatively well in a recursive setting but not in a static setting (compare Tables 3 and 5 or compare Tables 4 and 6). For example, we can observe dramatic improvements in forecasts of Series 7, 9 and 11 by using the recursive method.

Overall the results in Tables 3 to 6 show that the I(d) model performs constantly, but the forecastability of the BREAK model gets improved substantially using the recursive method. Of course, the break model has similar memory properties as the I(d) process, but it fails to capture the dynamics of break components by itself if we use the static method. It may now be, though, that each model alone does not capture all of the persistence in US inflation rates. Therefore, a more complete analysis would allow for both representations. For example, one can consider to combine forecasts, $\hat{y}_t = \alpha\hat{y}_{1t} + (1-\alpha)\hat{y}_{2t}$, $0 \leq \alpha \leq 1$, where $\hat{y}_{1t}$ and $\hat{y}_{2t}$ are the forecasts from occasional breaks model and $I(d)$ model, respectively. Combining forecasts would be useful if a test (Harvey, Leybourne and Newbold, 1998) would indicate so. The test is based on the regression where

$$e_{1t} = \alpha \left(e_{1t} - e_{2t}\right) + u_t, \tag{15}$$

where $e_{it} = y_t - \hat{y}_{it}$ is the forecast error of model i. If a composite predictor formed as a weighted average of two individual forecasts is considered, then the forecast of benchmark model is said to encompass the alternative forecast if the inferior forecast's optimal weight in the composite predictor is equal to zero. Our case is the issue of testing for forecast encompassing when two forecasts of the same quantity are available. This analysis provides an easy-to-compute statistical measure of the relative forecasting performance of the models under scrutiny. If one forecast incorporates all the information, nothing can be gained by combining forecasts.

From Table 7 we see that in most cases, the combination of the two forecasts does not improve predictability. In terms of lower MSFE, we choose a benchmark and alternative model of each series and we estimate α and its standard error in (15). Only for 5 out of 23 series, forecasts can be improved by combining inferior forecasts for both the static and recursive case. Hence, the BREAK or the I(d) model can forecast about equally well and nothing seems to be left to be predicted by the other model. This result implies that even in terms of out-of-sample forecasts, it is not easy to distinguish between the I(d) and the occasional breaks model.

5 Conclusion and Extensions

In this paper we compared time series models with structural breaks, models with long memory and linear autoregressive models for 23 monthly US inflation rates in terms of out-of-sample forecasting. Linear models did not perform as well as the other two. The model with breaks and the model with long-memory performed about equally well. We also examined their joint performance by combining the forecasts, which did not change the overall conclusion that these two models are difficult to distinguish.

The interesting topic for further research is now given by exploiting the possibility that the BREAK and I(d) model can be summarized into one single model. One motivation for this may be that both individual models capture a long-memory component to some extent, but that a joint model would be able to capture all long memory. To construct such a joint model, one can think of

Model A: $y_t = m_t + u_t$, $(1-L)^d m_t = q_t \eta_t$,

Model B: $y_t = m_t + u_t$, $m_t = m_{t-1} + q_t \eta_t$, $(1-L)^d u_t = \varepsilon_t$, or

Model C: $(1-L)^d y_t = m_t + u_t$, $m_t = m_{t-1} + q_t \eta_t$,

where each builds on an individual model considered in this paper by adding features of the other model. Our subsequent work will be to develop estimation and inference techniques for these models.

References

[1] Bai, J. (1997) Estimating Multiple Breaks One at a Time, Econometric Theory, 13, 315-352.

[2] Bai, J. & P. Perron (1998) Estimating and Testing Linear Models with Multiple Structural Changes, Econometrica, 66, 47-78.

[3] Baillie, R.T. (1996) Long Memory Processes and Fractional Integration in Econometrics, Journal of Econometrics, 73, 5-59.

[4] Beran, J. (1994) Statistics for Long-Memory Processes, Chapman & Hall, New York.

[5] Beran, J. (1995) Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, Journal of Royal Statistical Society B, 57, 659-672.

[6] Bos, C., P.H. Franses & M. Ooms (2001) Inflation, Forecast Intervals and Long Memory Regression Models, International Journal of Forecasting, to appear.

[7] Breidt, F.J., N. Crato & P. de Lima (1998) The Detection and Estimation of Long Memory in Stochastic Volatility, Journal of Econometrics, 83, 325-348.

[8] Diebold, F.X. & A. Inoue, (2001) Long Memory and Structural Change, Journal of Econometrics, to appear.

[9] Engle, R.F. & A.D. Smith (1999) Stochastic Permanent Breaks, Review of Economics and Statistics, 81, 553-574.

[10] Geweke, J. & S. Porter-Hudak (1983) The Estimation and Application of Long Memory Time Series Models, Journal of Time Series Analysis, 4, 15-39.

[11] Granger, C.W.J. & N. Hyung (2000) Occasional Structural Breaks and Long Memory, revised version of *UCSD Working Paper* 99-14.

[12] Harvey, D.I., S.J. Leybourne & P. Newbold (1998) Tests for Forecast Encompassing, *Journal of Business and Economic Statistics*, 16, 254-259.

[13] Hassler,V. & J. Wolters (1995) Long memory in inflation rates: International evidence, *Journal of Business & Economic Statistics*, 13, 37-45.

[14] Hosking, J.R.M. (1984) Modeling Persistence In Hydrological Time Series Using Fractional Differencing, *Water Resources Research*, 20, 1898-1908.

[15] Kim, C.S. & Phillips, P.C.H. (2000) Modified Log Periodogram Regression, *Yale University Working Paper*.

[16] Mikosch, T. and C. Stărică, 1999, Change of Structure in Financial Time Series, Long Range Dependence and the GARCH Model, *Working Paper*, The Wharton School, Chalmers University of Technology.

[17] Nunes, L.C., P. Newbold & C.-M., Kuan (1996) Spurious number of breaks, *Economics Letters*, 50, 175-178.

[18] Pesaran, M.H. & A. Timmermann (2000) Model Instability and Choice of Observation Window, *UCSD Working Paper*.

[19] Smith, A.D. (1999) Forecasting Inflation in the Presence of Structural Breaks, *mimeo* University of Virginia.

Table 1: Descriptions of U.S. City Average CPI

Series	Items in CPI
1	All items
2	Durables
3	Commodities
4	Energy commodities
5	Commodities less food
6	Commodities less food, energy, energy commodities
7	Commodities less food, energy, and used cars and trucks
8	Services
9	Medical care services
10	Transportation services
11	Transportation
12	Housing
13	Electricity
14	Fuels
15	New vehicles
16	Men's and boys' apparel
17	Footwear
18	Alcoholic beverages
19	Eggs
20	Beef and veal
21	Fish and seafood
22	Fruits and vegetables
23	Potatoes

Source: Economagic.com: Economic Time Series Page .

Note: All of series are seasonally adjusted and 1982 - 84 =100. Sample period, 1967:01-2000:08 with 404 observations.

Table 2: Long Memory Parameters of CPI Series

	Break		d (s.e.)	
Series	Short-sample	Full-sample	Short-sample	Full-sample
1	3	4	.973 (.235)	.912 (.195)
2	3	4	.635 (.214)	.541 (.205)
3	3	3	.742 (.245)	.607 (.175)
4	0	0	.617 (.281)	.626 (.159)
5	2	2	.633 (.387)	.560 (.140)
6	3	6	.647 (.133)	.550 (.155)
7	5	6	.657 (.206)	.672 (.150)
8	2	3	.759 (.214)	.698 (.159)
9	3	4	.725 (.148)	.689 (.108)
10	0	6	.504 (.243)	.712 (.212)
11	0	2	.467 (.282)	.460 (.252)
12	6	6	.886 (.163)	.829 (.139)
13	4	3	.778 (.243)	.582 (.110)
14	3	3	.639 (.181)	.765 (.215)
15	0	5	.471 (.152)	.364 (.149)
16	0	1	.278 (.153)	.155 (.139)
17	1	2	.376 (.176)	.702 (.115)
18	0	5	.550 (.143)	.477 (.114)
19	0	0	.142 (.200)	.060 (.222)
20	0	0	.226 (.135)	.199 (.093)
21	2	2	.700 (.301)	.573 (.141)
22	0	0	.348 (.173)	.517 (.229)
23	0	0	-.341 (.227)	-.365 (.167)

Note: Short-sample period, 1967:01-1990:12 with 288 observations and Full-sample period 1967:01-2000:08 with 404 observations. The numbers d with estimated standard error (s.e.) of the CPI series are estimated using Kim and Philips (2000)'s modified log periodogram regression. The number of breaks is estimated by Bai and Perron (1998)'s sequential method using a parametric dynamic autoregressive specification.

Note: Series 18 has 3 outliers in the short-sample period, which we delete prior to estimation.

Table 3: Root Mean Squared Forecast Errors relative to a linear AR model (where the parameters have been estimated only once for the in-sample data)

	BREAK				I(d)			
Series	h =1	3	12	24	1	3	12	24
1	0.971	0.934	0.806	0.691	0.977	0.923	0.850	0.720
2	1.012	1.002	0.787	0.648	0.950	0.872	0.740	0.705
3	0.962	0.977	0.869	0.807	0.980	0.961	0.924	0.943
4^+	1.000	1.001	1.001	1.001	1.064	1.072	1.114	1.058
5	0.946	0.892	0.831	0.930	0.984	0.941	0.898	0.878
6	0.946	0.908	0.734	0.713	0.910	0.770	0.598	0.622
7	1.156	1.230	0.909	0.771	0.958	0.887	0.732	0.684
8	1.008	1.005	0.900	0.774	0.884	0.725	0.512	0.455
9	1.558	1.489	0.903	0.778	0.715	0.614	0.506	0.583
10^+	0.999	0.997	0.994	0.994	0.931	0.920	0.853	0.869
11^+	1.001	1.004	1.005	1.005	1.009	0.985	0.962	0.947
12	0.944	0.930	0.648	0.506	0.935	0.821	0.597	0.489
13	0.907	0.887	0.850	0.849	0.929	0.868	0.854	0.861
14	0.911	0.913	0.906	0.865	0.977	0.975	0.971	0.930
15^+	0.996	0.994	0.994	0.994	0.812	0.719	0.705	0.807
16^+	1.000	1.001	1.001	1.001	0.970	0.955	0.961	0.971
17	0.965	0.967	0.935	0.925	0.954	0.959	0.936	0.944
18^+	0.997	0.995	0.988	0.986	0.929	0.895	0.805	0.812
19^+	0.995	0.996	0.997	0.998	1.003	1.001	1.004	0.998
20^+	1.001	1.002	1.002	1.002	1.001	0.958	0.942	0.953
21	0.974	0.986	0.953	0.919	0.996	0.986	0.941	0.923
22^+	0.999	0.999	0.999	0.999	1.035	1.014	0.993	1.001
23^+	0.996	0.994	0.995	0.998	1.043	1.064	1.049	0.999
#	12	7	9	10	11	16	14	13

Note: In-sample period, 1967:01-1990:12 with 288 observations and out-of-sample period, 1991:01-2000:08 with 116 observations. Lags are selected by AIC. $^+$ denotes no break is detected in the in-sample period. We estimate breaks by Bai and Perron (1998)'s sequential method. # denotes the number of times one of the two models has a lower RMSFE.

Table 4: Root Mean Squared Cumulative Forecast Errors relative to a linear AR model (where the parameters have been estimated only once for the in-sample data)

	BREAK				I(d)			
Series	h =1	3	12	24	1	3	12	24
1	0.971	0.865	0.715	0.613	0.977	0.832	0.724	0.592
2	1.012	1.132	0.860	0.620	0.950	0.781	0.616	0.583
3	0.962	0.964	0.863	0.687	0.980	0.905	0.808	0.655
4^+	1.000	1.003	1.009	1.027	1.064	1.045	1.405	1.751
5	0.946	0.775	0.557	0.474	0.984	0.851	0.645	0.573
6	0.946	0.954	0.711	0.625	0.910	0.618	0.370	0.365
7	1.156	1.599	1.154	0.861	0.958	0.781	0.542	0.500
8	1.008	1.015	0.960	0.848	0.884	0.569	0.372	0.342
9	1.558	1.684	1.191	0.958	0.715	0.479	0.408	0.467
10^+	0.999	0.993	0.986	0.984	0.931	0.836	0.577	0.546
11^+	1.001	1.010	1.017	1.022	1.009	0.939	0.782	0.693
12	0.944	1.018	0.689	0.498	0.935	0.659	0.403	0.323
13	0.907	0.855	0.560	0.487	0.929	0.731	0.453	0.413
14	0.911	0.907	0.671	0.400	0.977	0.948	0.861	0.691
15^+	0.996	0.992	0.991	0.990	0.812	0.628	0.581	0.629
16^+	1.000	1.002	1.007	1.009	0.970	0.909	0.781	0.770
17	0.965	0.935	0.723	0.613	0.954	0.911	0.703	0.673
18^+	0.997	0.989	0.979	0.977	0.929	0.798	0.587	0.588
19^+	0.995	0.989	0.953	0.929	1.003	0.983	0.953	0.929
20^+	1.001	1.004	1.007	1.008	1.001	0.881	0.762	0.753
21	0.974	0.974	0.783	0.590	0.996	0.938	0.632	0.564
22^+	0.999	0.996	0.982	0.973	1.035	0.963	0.924	0.894
23^+	0.996	0.985	0.961	0.941	1.043	1.176	1.073	0.968
#	12	4	5	4	11	19	18	19

Note: In-sample period, 1967:01-1990:12 with 288 observations and out-of-sample period, 1991:01-2000:08 with 116 observations. Lags are selected by AIC. $^+$ denotes no break is detected in the in-sample period. We estimate breaks by Bai and Perron (1998)'s sequential method. # denotes the number of times one of the two models has a lower RMSFE.

Table 5: Root Mean Squared Forecast Errors relative to a linear AR model (where the parameters have been estimated recursively using expanding samples)

	BREAK				I(d)			
Series	h =1	3	12	24	1	3	12	24
1	0.970	0.976	0.878	0.789	0.996	0.957	0.926	0.778
2	1.039	1.075	0.900	0.723	0.974	0.939	0.857	0.813
3	0.959	0.986	0.893	0.872	1.003	0.979	0.959	0.890
4^+	1.000	1.000	1.000	1.000	1.101	1.068	1.124	1.008
5	0.945	0.919	0.861	0.858	0.990	0.929	0.930	0.901
6	0.905	0.864	0.716	0.713	0.907	0.818	0.685	0.709
7	0.948	0.984	0.814	0.716	0.976	0.943	0.835	0.770
8	0.999	0.991	0.917	0.790	0.935	0.839	0.621	0.519
9	1.098	1.086	0.750	0.705	0.873	0.796	0.644	0.689
10	0.976	0.979	0.954	0.972	1.001	0.942	0.890	0.885
11	0.966	0.964	0.944	0.946	1.054	0.945	1.019	0.979
12	0.919	0.931	0.713	0.568	0.958	0.890	0.680	0.555
13	0.901	0.895	0.863	0.869	0.943	0.903	0.886	0.901
14	0.930	0.937	0.930	0.905	0.991	0.992	1.024	1.016
15	0.976	0.967	0.991	0.997	0.858	0.779	0.766	0.871
16	1.008	1.014	0.994	1.006	1.092	1.018	1.047	1.006
17	0.995	0.966	0.948	0.940	0.951	0.970	0.954	0.955
18	0.963	0.976	0.939	0.928	0.948	0.931	0.854	0.855
19^+	0.996	0.997	0.999	0.999	1.022	0.981	1.001	1.001
20^+	1.000	1.001	1.001	1.001	1.054	1.072	1.012	0.989
21	0.968	0.984	0.962	0.941	1.004	1.001	0.966	0.945
22^+	0.999	0.999	0.999	1.000	1.145	1.027	1.048	1.073
23^+	0.998	0.997	0.998	0.999	1.030	1.003	1.012	1.003
#	17	11	15	13	6	12	8	10

Note: We estimate the parameters recursively, starting with sample 1967:01-1990:12. Lags are selected by AIC for the short-sample period and we fix this value in the subsequent analysis. $^+$ denotes no break is detected in the full-sample period. We estimate breaks by Bai and Perron (1998)'s sequential method. # denotes the number of times one of the two models has a lower RMSFE.

Table 6: Root Mean Squared Cumulative Forecast Errors relative to a linear AR model (where the parameters have been estimated recursively using expanding samples)

	BREAK				I(d)			
Series	h =1	3	12	24	1	3	12	24
1	0.970	0.934	0.791	0.676	0.996	0.959	0.801	0.608
2	1.039	1.109	0.989	0.754	0.974	0.926	0.818	0.772
3	0.959	0.955	0.837	0.677	1.003	1.003	0.902	0.725
4^+	1.000	0.999	1.001	1.001	1.101	1.266	1.323	1.188
5	0.945	0.826	0.575	0.441	0.990	0.929	0.666	0.563
6	0.905	0.778	0.606	0.575	0.907	0.763	0.483	0.486
7	0.948	0.897	0.728	0.600	0.976	0.913	0.699	0.630
8	0.999	0.985	0.958	0.855	0.935	0.818	0.503	0.424
9	1.098	1.125	0.859	0.746	0.873	0.764	0.606	0.613
10	0.976	0.944	0.864	0.878	1.001	0.985	0.793	0.645
11	0.966	0.919	0.777	0.646	1.054	1.043	0.839	0.748
12	0.919	0.870	0.680	0.505	0.958	0.854	0.488	0.358
13	0.901	0.726	0.507	0.450	0.943	0.795	0.567	0.541
14	0.930	0.891	0.748	0.483	0.991	1.004	1.018	0.947
15	0.976	0.961	0.992	0.996	0.858	0.726	0.652	0.723
16	1.008	1.059	1.090	1.105	1.092	1.276	1.350	1.194
17	0.995	0.912	0.757	0.656	0.951	0.894	0.729	0.696
18	0.963	0.931	0.887	0.870	0.948	0.886	0.706	0.673
19^+	0.996	0.988	0.968	0.960	1.022	1.007	1.024	1.084
20^+	1.000	1.001	1.003	1.004	1.054	1.107	1.059	1.011
21	0.968	0.908	0.696	0.528	1.004	0.987	0.797	0.660
22^+	0.999	0.996	0.984	0.975	1.145	1.233	1.684	1.897
23^+	0.998	0.994	0.975	0.956	1.030	1.039	1.026	1.027
#	17	15	14	15	6	8	9	8

Note: We estimate the parameters recursively, starting with sample 1967:01-1990:12. Lags are selected by AIC for the short-sample period and we fix this value in the subsequent analysis. $^+$ denotes no break is detected in the full-sample period. We estimate breaks by Bai and Perron (1998)'s sequential method. # denotes the number of times one of the two models has a lower RMSFE.

Table 7: Tests for one-step ahead Forecast Encompassing

	Static method		Recursive method	
Series	Break vs I(d)	I(d) vs Break	Break vs I(d)	I(d) vs Break
1	0.35 (0.48)	-	-0.37 (0.53)	-
2	-	0.17 (0.21)	-	-0.06 (0.27)
3	0.29 (0.31)	-	-0.15 (0.35)	-
4	-1.11 (0.45)	-	-1.41 (0.37)	-
5	-0.66 (0.50)	-	-1.09 (0.53)	-
6	-	0.29 (0.21)	0.46 (0.37)	-
7	-	0.08 (0.13)	0.22 (0.28)	-
8	-	-0.13 (0.19)	-	-0.27 (0.31)
9	-	-0.04 (0.05)	-	-0.24 (0.15)
10	-	-0.88 (0.39)	0.04 (0.40)	-
11	0.35 (0.41)	-	-0.75 (0.33)	-
12	-	0.47 (0.16)	0.24 (0.23)	-
13	-	0.69 (0.26)	-0.21 (0.36)	-
14	-0.65 (0.36)	-	-0.43 (0.34)	-
15	-	-0.55 (0.18)	-	-0.56 (0.25)
16	-	-0.24 (0.45)	-0.20 (0.27)	-
17	-	-0.52 (0.89)	-	-0.16 (0.35)
18	-	-0.45 (0.33)	-	-0.04 (0.55)
19	-1.29 (1.43)	-	-1.32 (0.78)	-
20	0.49 (0.44)	-	-1.56 (0.55)	-
21	0.26 (0.30)	-	-0.05 (0.36)	-
22	-0.21 (0.41)	-	-0.63 (0.24)	-
23	-0.23 (0.36)	-	-0.51 (0.52)	-

Note: The values in the parenthesis are standard errors. Results compare one-step ahead forecasts between Break and I(d) models. In terms of lower MSFE, we choose a benchmark and alternative model of each series and estimate α and its standard error in (15). - denotes that we choose the other model as a benchamrk model.

Figure 1: Inflation of CPI (Series 1: All Items)

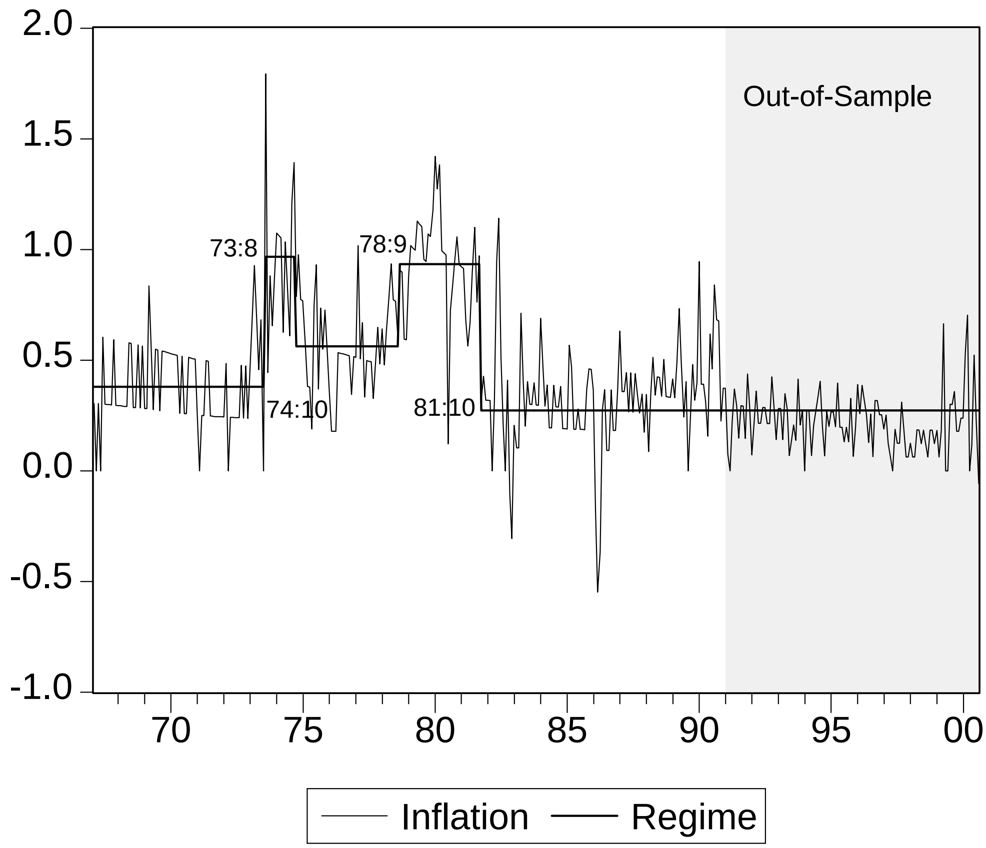

Figure 2: Forecasts Errors of Series 2 (Durables)