Optimal Forecasts in the Presence of Discrete Structural Breaks under Long Memory

Mwasi Paza Mboya[a], Philipp Sibbertsen[a,*]

[a] *Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany*

Abstract

We develop methods to obtain optimal forecast under long memory in the presence of a discrete structural break based on different weighting schemes for the observations. We observe significant changes in the forecasts when long-range dependence is taken into account. Using Monte Carlo simulations, we confirm that our methods substantially improve the forecasting performance under long memory. We further present an empirical application to inflation rates that emphasizes the importance of our methods.

Keywords: long memory, forecasting, structural break, optimal weight, ARFIMA model
JEL classification: C12, C22

1. Introduction

Forecasting is among the most prominent areas of time-series analysis. It has drawn particular interest in macroeconomics and finance, although imprecise and unreliable forecasts might be produced in the presence of structural breaks due to instabilities. A reason for this instability is that the usual forecasting strategy when there are structural breaks in the series would be to estimate the break point and use the post-break data for forecasting. This strategy leads on the one hand to only a short time period used for forecasting and on the other hand to neglecting available information given by the dependence structure of the time series. Many studies (see Clements and Hendry (2000); Rossi (2013); Clements and Hendry (2000); Rossi (2013); Giacomini and Rossi (2009); Inoue and Rossi (2011); Stock and Watson (1996); Paye and Timmermann (2006)) provide evidence of such instabilities. However, Bayesian models have been proposed by Pesaran et al. (2006), Koop and Potter (2007), Maheu and Gordon (2008) and Maheu and McCurdy (2009) to address this issue.

In addition to this instability of forecasts structural breaks can also increase estimates of the long-run variance which is used for normalization in tests to evaluate the forecast performance such as the Diebold-Mariano (DM) test. Such an increase in the long-run variance estimate leads to serious power problems for these tests as recently pointed out by Casini (2021) and Casini et al. (2021).

[1] We are grateful to an anonymous referee and Siem Jan Koopman for helpful comments.

[*] Corresponding author.
Email addresses: mboya@statistik.uni-hannover.de (Mwasi Paza Mboya), sibbertsen@statistik.uni-hannover.de (Philipp Sibbertsen)

To overcome the aforementioned instabilities the problem of forecasting under discrete structural breaks can be addressed based on weighted observations to obtain optimal forecasts through minimization of the mean-square forecast error (MSFE). The most prominent element of refining the forecasting performance is the one-step-ahead forecast assumption, which plays an important role in improving the precision of forecasts within a variety of methods that propose different weighting observations. For instance, Pesaran et al. (2013) suggest defining optimal weights for each pre-break and post-break observation. However, Pesaran and Timmermann (2007) propose an optimal window in which equal weights are given to observations within the window and zero weights given to those elsewhere. And, also defining a post-break window allows equal weights to be applied to observations within the window after the break, as the name suggests. Lastly, Pesaran and Timmermann (2007) use average forecasts across estimation windows (AveW) when time and size of the break is uncertain, which as Pesaran and Pick (2011) shows to improve forecasts; this method has the advantage of not relying on estimated break dates and sizes.

There is a growing literature showing that processes with structural breaks can empirically mimic long-memory behaviour in the sense of an observationally equivalent autocovariance function or spectral density. Examples for this literature include among others Granger and Ding (1996), Granger and Hyung (2004), Diebold and Inoue (2001), Mikosch and Stărică (2004) or Casini et al. (2021). Hou and Perron (2014) and Qu (2011) show that the two phenomena are distinct though and lead to different asymptotic behaviours. A test for long memory against structural breaks can be found in Qu (2011) and a multivariate extension is Sibbertsen et al. (2018).

A study by Sibbertsen and Kruse (2009) points out that forecasting precision is substantially reduced if a break in persistence is ignored. Likewise, we might experience the same problem if we apply the theoretical forecasting procedures in Pesaran et al. (2013) under discrete structural breaks, ignoring possible long-range dependencies, to obtain the optimal forecast of a time series exhibiting long memory. In this paper, we adapt the different forecasting methods discussed in Pesaran et al. (2013) by introducing long memory in such a setting. This develops the existence of variance and covariance terms of an error, which depends solely on the long memory parameter d. Involvement of such terms in the theoretical forecasting procedures are substantially important, as they modify the MSFEs, which results in an increase of the pre-break weight, a decrease in the post-break weight and an increase in the optimal window size. Consequently, the approaches in Pesaran et al. (2013) are no longer robust when long memory is present in the time series. The main reason for this is that the optimal forecast error is driven by the autocovariance function of the underlying time series process which is in our case only hyperbolically decaying and dependent on the memory parameter d.

In practice, the dates and size of the break and the memory parameters must be estimated since they are unknown. A method for estimating the break dates under long memory has been considered in Lavielle and Moulines (2000a) extending results of Bai and Perron (1998), and conditional on these estimates, we obtain the break size estimate. We use the modified local Whittle (LW) estimator of Hou and Perron (2014) that accounts for possible low frequency contaminations with bandwidth $m = T^{\delta}$, where $\delta \in (0, 1)$ to estimate the memory parameters. Nevertheless, the problem of imprecise estimates deteriorating the forecasting performance remains.

We conduct Monte Carlo experiments to compare the forecasting performance of the proposed methods with the ones discussed in Pesaran et al. (2013). We generally observe that under discrete breaks, with larger breaks, one can obtain more precisely estimated values and, hence, an improved forecasting performance in terms of optimal weight forecasts, post-break forecasts and optimal window forecasts. Apart from this, we observe that under different estimates of the break size, memory parameters and break dates, the MSFE is in all cases much lower under the proposed methods than those discussed in Pesaran et al. (2013). However, the elements of the proposed methods displaying the most significant changes in the MSFE are the estimated optimal weights and estimated optimal window, while the rest of the elements show no change. We further analyse the performance by providing the DM test statistics that compares the forecast accuracy of our proposed methods to the ones discussed in Pesaran et al. (2013).

We apply different forecasting methods, to both proposed methods in this paper and the ones discussed in Pesaran et al. (2013) for comparison, to forecast the real inflation rates for 10 countries covering the period from January 1967 to December 2017. The general findings, similar to the Monte Carlo results, are that the methods proposed in this paper outperform the ones discussed in Pesaran et al. (2013).

A related though somehow different problem is the question of the out-of-sample stability of forecasts. This problem is discussed in Casini (2018) and Perron and Yamamoto (2021). However, this problem needs a different methodology and is therefore not discussed in this paper.

The rest of the paper is organized as follows. Section 2 sets up the model and derives the forecasting procedures of the proposed methods, with the error assumed to be an innovation process with long memory parameter d. Section 3 conduct Monte Carlo experiments that compares the forecasting performance of different proposed methods with the ones discussed in Pesaran et al. (2013). The results and discussion of the empirical application of our findings are presented in section 4. Section 5 concludes. All detailed proofs are gathered in the online appendix.

2. A Single, Discrete Break in a Simple Regression Model

Consider the linear regression model:

$$y_t = \beta_t + \sigma_\varepsilon \varepsilon_t, \qquad t = 1, \ldots, T+1 \tag{1}$$

where β_t describes the mean or slope parameter, σ_ε^2 describes the scalar error variance subject to a single break, and ε_t is the innovation process associated with long memory.

Now, we assume that β_t is subject to a single, discrete break at T_b, $1 < T_b < T$:

$$\beta_t = \begin{cases} \beta_1 & \text{for} \quad t \le T_b \\ \beta_2 & \text{for} \quad T_b < t \le T+1 \end{cases} \tag{2}$$

Let ε_t be a long memory process generated according to the ARFIMA(p, d, q) model as proposed by Granger and Joyeux (1980):

$$\Phi(L)\,(1-L)^d\,\varepsilon_t \;=\; \Psi(L)\,\eta_t, \qquad \text{as} \quad t = 1, \ldots, T,$$

where η_t is i.i.d. white noise with mean 0, variance $\sigma_\eta^2 \;=\; 1$ and $E|\eta_t|^{2+\delta} < \infty$ for some $\delta > 0$. The AR and MA polynomials, *i.e.*, $\Phi(L)$ and $\Psi(L)$, respectively, are assumed to have all roots outside the unit circle.

Now, we simply write $\varepsilon_t \sim$ ARFIMA$(0, d, 0)$ because of the power-like behavior of its covariance function, where ε_t has mean $E[\varepsilon_t] \;=\; 0$, the covariance is given by:

$$Cov[\varepsilon_t, \varepsilon_{t+k}] \;=\; E[\varepsilon_t, \varepsilon_{t+k}] \;=\; \gamma(k) \;=\; \sigma_\varepsilon^2\,\frac{(-1)^k\,\Gamma(1-2d)}{\Gamma(1+k-d)\,\Gamma(1-k-d)}, \qquad t = 1, \ldots, T, \tag{3}$$

and the variance as:

$$Var[\varepsilon_t] \;=\; E[\varepsilon_t^2] \;=\; \gamma(0) \;=\; \sigma_\varepsilon^2\,\frac{\Gamma(1-2d)}{\Gamma^2(1-d)}, \qquad t = 1, \ldots, T, \tag{4}$$

as defined by Beran et al. (2016), where $\Gamma(.)$ denotes the gamma function. The above assumption is chosen only to simplify the derivations mechanism, but does not affect the validity of the proofs in general.

The basic concept of this section is to first derive a general expression for the mean squared forecasting error in our model and derive as a baseline the MSFE if the forecasting weights are assumed to be equal. This simple model serves as a competitor for comparison with a choice of weights taking the long-memory structure of the underlying process into account. We then in a next step derive the MSFE with constant breaks before and different constant weights after the break. Afterwards we introduce optimal forecasting windows and derive first post break window forecasts, afterwards forecasts when the window contains the break. Last, an average across the estimation windows is considered.

Now we turn to considering different methods for weighting past observations w_t, when estimating the regression coefficient. In this case $\hat{\beta}_T(w)$ as suggested by Pesaran et al. (2013) is given by:

$$\hat{\beta}_T(w) \;=\; \sum_{t=1}^{T} w_t\,y_t, \tag{5}$$

subject to the restriction $\sum_{t=1}^{T} w_t \;=\; 1$, such that the resulting MSFE of the one-step-ahead forecast, $\hat{y}_{T+1} \;=\; \hat{\beta}_T(w)$, is minimized.

As we state in the following theorem, we consider the weights of past observations to be used in the estimation, $\hat{\beta}_T(w)$, and thereby obtain the resulting general MSFE of the one-step-ahead forecast.

Theorem 1. In the linear regression model (1), the scaled MSFE of the one-step-ahead forecast is generally computed as

$$E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2(w)\right] = A + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + A\sum_{t=1}^{T} w_t^2 + 2\sum_{s=2}^{T}\sum_{t=s}^{T} w_{s-1}\, w_t\, \gamma(t-s+1), \tag{6}$$

where $k = t-s+1$, $A = \sigma_\varepsilon^2 \dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}$, $\gamma(k) = \dfrac{(-1)^k\,\Gamma(1-2d)}{\Gamma(1+k-d)\,\Gamma(1-k-d)}$, $\lambda = (\beta_1-\beta_2)/\sigma_\varepsilon$ and $e_{T+1}(w) = y_{T+1} - \hat{y}_{T+1}$ describes the forecast error.

The above result is derived by using equations (1), (2) and (5) to obtain the expression of the forecast error, and then the error is squared, divided by σ_ε^2 and the expected value is applied to obtain the derivation of the MSFE scaled by the error variance.

By using equations (1), (2) and (5), we obtain the simplified expression as

$$\hat{\beta}_T(w) - \beta_T = (\beta_1-\beta_2)\sum_{t=1}^{T_b} w_t + \sigma_\varepsilon \sum_{t=1}^{T} w_t\,\varepsilon_t,$$

Then, the expression of forecast error is given by

$$\begin{aligned} e_{T+1}(w) &= y_{T+1} - \hat{y}_{T+1}, \\ &= y_{T+1} - \hat{\beta}_T(w), \\ &= \sigma_\varepsilon\,\varepsilon_{T+1} - (\beta_1-\beta_2)\sum_{t=1}^{T_b} w_t - \sigma_\varepsilon \sum_{t=1}^{T} w_t\,\varepsilon_t, \end{aligned}$$

and lastly the MSFE scaled by the error variance is

$$E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2(w)\right] = A + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + A\sum_{t=1}^{T} w_t^2 + 2\sum_{s=2}^{T}\sum_{t=s}^{T} w_{s-1}\, w_t\, \gamma(t-s+1),$$

Next, we construct the baseline against which all other forecasting methods are compared by suggesting equal weights to be used in the estimation $\hat{\beta}_T(w)$, yielding the MSFE of the one-step-ahead forecast, which is taken as a reference.

Theorem 2. Under the conditions of Theorem 1, where the equal weights $w_t^{\text{equal}} = 1/T$ is suggested, then the scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2} | w_t^{\text{equal}}\right] = A + \lambda^2\, b^2 + \frac{A}{T} + \frac{2}{T^2} \sum_{s=2}^{T} \sum_{t=s}^{T} \gamma(t-s+1), \quad (7)$$

where $b = T_b/T$, $k = t-s+1$, $A = \sigma_{\varepsilon}^{2} \dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}$, $\gamma(k) = \dfrac{(-1)^k\, \Gamma(1-2d)}{\Gamma(1+k-d)\, \Gamma(1-k-d)}$ and $\lambda = (\beta_1 - \beta_2)/\sigma_{\varepsilon}$.

Using equation (6), we replace the weights by $w_t = 1/T$, and we obtain the scaled MSFE for the equal weights.

Remark 1. It is obvious that forecasts using equal weights to observations will have the largest MSFEs among all forecasting methods. This is why we need different methods for weighting observations while minimizing the MSFE of the one-step-ahead forecast.

2.1. Optimal weights in a model with a single, discrete break

Now, we derive the optimal weights to be used in the estimation of the regression parameter to minimize the MSFE of the one-step-ahead forecast.

By using equation (6), we obtain the optimal weights by minimizing the equation subject to $\sum_{t=1}^{T} w_t = 1$. The first derivatives are:

For $t \leq T_b$

$$2\,\lambda^2 \sum_{t=1}^{T_b} w_t + 2\,A\,w_t + 2\sum_{s=2}^{t} w_{s-1}\,\gamma(t-s+1) + 2\sum_{s=t+2}^{T} w_{s-1}\,\gamma(s-t-1) + \theta\, = 0,$$

For $T_b < t \leq T$

$$2\,A\,w_t + 2\sum_{s=2}^{t} w_{s-1}\,\gamma(t-s+1) + 2\sum_{s=t+2}^{T} w_{s-1}\,\gamma(s-t-1) + \theta\, = 0,$$

where θ is the Lagrange multiplier associated with $\sum_{t=1}^{T} w_t$.

Hence, as the weights for each pre-break and post-break observation, we obtain:

$$w_t = \begin{cases} w_1 \;=\; \frac{-\lambda^2}{A} \sum_{t=1}^{T_b} w_t - \frac{1}{A}\left[\sum_{s=2}^{t} w_{s-1}\,\gamma(t-s+1) + \sum_{s=t+2}^{T} w_{s-1}\,\gamma(s-t-1)\right] - \frac{\theta}{2\,A} \\ \quad \text{for} \quad 1 < t \leq T_b \\ \\ w_2 \;=\; \frac{-1}{A}\left[\sum_{s=2}^{t} w_{s-1}\,\gamma(t-s+1) + \sum_{s=t+2}^{T} w_{s-1}\,\gamma(s-t-1)\right] - \frac{\theta}{2\,A} \qquad \text{for} \quad T_b < t \leq T+1 \end{cases}$$

and $w_2 - w_1 = \frac{\lambda^2}{A} \sum_{t=1}^{T_b} w_t = \frac{\lambda^2}{A} T_b \, w_1$. Then, we substitute $\sum_{t=1}^{T} w_t = T_b \, w_1 + (T - T_b) \, w_2 = 1$ to yield the optimal weights:

For $t \leq T_b$

$$w_1 = \frac{1}{T} \frac{A}{T \, b \, (1-b) \, \lambda^2 + A}, \tag{8}$$

For $T_b < t \leq T$

$$w_2 = \frac{1}{T} \frac{T \, b \, \lambda^2 + A}{T \, b \, (1-b) \, \lambda^2 + A}, \tag{9}$$

Remark 2. In comparison to Pesaran et al. (2013), we introduce the variance and covariance terms of an error that depends on the long memory parameter d, which results to the equation (6) and through minimization leads to an increase in the prebreak weight and decrease in the postbreak weight obtained in equation (8) and equation (9), respectively. Intuitively, this is due to the strong correlation structure of the long-memory process and the slowly decaying correlation function leading to higher weights for observations further in the past.

The following theorem is obtained by using equations (8) and (9) in equation (6), in which the reduced form of the scaled MSFE for the optimal weights is obtained.

Theorem 3. Under the conditions of Theorem 1, we assume that the weights are constant for each pre-break observation as w_1 and those for each post-break observation as w_2, then the scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_\varepsilon^{-2} \, e_{T+1}^2 | w_1, w_2\right] = A + (T_b \, \lambda \, w_1)^2 + T_b \, A \, w_1^2 + (T - T_b) \, A \, w_2^2 + 2 \sum_{s=2}^{T} \sum_{t=s}^{T} w_{s-1} \, w_t \, \gamma(t - s + 1).$$

Using equations (8) and (9), we obtain the reduced form of scaled MSFE for the optimal weights:

$$\begin{aligned}
E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2}|w_1, w_2\right] &= A + (T_b^2\,\lambda^2 + T_b\,A)\,w_1^2 + (T - T_b)\,A\,w_2^2 + 2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b} w_{s-1}\,w_t\,\gamma(t-s+1) \\
&+ 2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T} w_{s-1}\,w_t\,\gamma(t-s+1), \\
&= A\left[1 + \frac{1}{T}\,\frac{T\,b\,\lambda^2 + A}{T\,b\,(1-b)\,\lambda^2 + A}\right] + 2\,w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1) \\
&+ 2\,w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
&= A\,(1+w_2) + 2\left[w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1) + w_2^2\sum_{s=Tb+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right], \qquad (10)
\end{aligned}$$

Now, we compare the MSFEs of the forecasts from the equal weights to that of the optimal weights. So, we compute the difference between equations (7) and (10) as:

$$\begin{aligned}
&E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2}|w_t^{\text{equal}}\right] - E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2}|w_1, w_2\right] \\
&= A + \lambda^2\,b^2 + \frac{A}{T} + \frac{2}{T^2}\sum_{s=2}^{T}\sum_{t=s}^{T}\gamma(t-s+1) - A - \frac{A}{T}\,\frac{T\,b\,\lambda^2 + A}{T\,b\,(1-b)\,\lambda^2 + A} \\
&- \frac{2}{T^2}\,\frac{A^2}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1) - \frac{2}{T^2}\,\frac{\left[T\,b\,\lambda^2 + A\right]^2}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
&= \lambda^2\,b^2 - \frac{A\,b^2\,\lambda^2}{T\,b\,(1-b)\,\lambda^2 + A} + \frac{2}{T^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T}\gamma(t-s+1), \\
&+ \frac{2}{T^2}\,\frac{T\,b^2\,\lambda^2\left[T\,b^2\,\lambda^2 - 2\left(T\,b\,\lambda^2 + A\right)\right]}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1) - \frac{2}{T^2}\,\frac{A^2}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1),
\end{aligned}$$

First, we consider

$$\begin{aligned}
\lambda^2\,b^2 - \frac{A\,b^2\,\lambda^2}{T\,b\,(1-b)\,\lambda^2 + A} &= \frac{T\,b\,(1-b)\,b^2\,\lambda^4 + A\,b^2\,\lambda^2 - A\,b^2\,\lambda^2}{T\,b\,(1-b)\,\lambda^2 + A}, \\
&= \frac{T\,b^3\,(1-b)\,\lambda^4}{T\,b\,(1-b)\,\lambda^2 + A} \geq 0\,,
\end{aligned}$$

Next, we have

$$\frac{2}{T^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T}\gamma(t-s+1) \geq \frac{2}{T^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)\,,$$

and

$$\begin{aligned}\sum_{s=T_b+1}^{T}\ \sum_{t=s}^{T}\gamma(t-s+1)\ &\geq \sum_{s=T_b+1}^{T}\ \sum_{t=s}^{T_b}\gamma(t-s+1)\,,\\ &=\sum_{s=2}^{T}\sum_{t=s}^{T_b}\gamma(t-s+1)-\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1),\\ &\geq -\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1),\end{aligned}$$

Thus

$$\begin{aligned}&\frac{2}{T^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T}\gamma(t-s+1)+\frac{2}{T^2}\frac{T\,b^2\,\lambda^2\left[T\,b^2\,\lambda^2-2\,\left(T\,b\,\lambda^2+A\right)\right]}{\left[T\,b\,(1-b)\,\lambda^2+A\right]^2}\sum_{s=T_b+1}^{T}\ \sum_{t=s}^{T}\gamma(t-s+1)\\ &-\frac{2}{T^2}\,\frac{A^2}{\left[T\,b\,(1-b)\,\lambda^2+A\right]^2}\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1),\\ &\geq\frac{2}{T^2}\left[1-\frac{T\,b^2\,\lambda^2\left[T\,b^2\,\lambda^2-2\,\left(T\,b\,\lambda^2+A\right)\right]}{\left[T\,b\,(1-b)\,\lambda^2+A\right]^2}-\frac{A^2}{\left[T\,b\,(1-b)\,\lambda^2+A\right]^2}\right]\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)\geq 0\,,\end{aligned}$$

because

$$\begin{aligned}&\frac{T\,b^2\,\lambda^2\left[T\,b^2\,\lambda^2-2\,\left(T\,b\,\lambda^2+A\right)\right]+A^2}{\left[T\,b\,(1-b)\,\lambda^2+A\right]^2}\\ &=\frac{T^2\,b^4\,\lambda^4-2\,T^2\,b^3\,\lambda^4-2\,T\,b^2\,\lambda^2\,A+A^2}{T^2\,b^4\,\lambda^4-2\,T^2\,b^3\,\lambda^4-2\,T\,b^2\,\lambda^2\,A+A^2+T^2\,b^2\,\lambda^4+2\,T\,b\,\lambda^2\,A}\leq 1\end{aligned}$$

For this reason, we have

$$
\begin{aligned}
& E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2 | w_t^{\text{equal}}\right] - E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2 | w_1, w_2\right] \\
&= \lambda^2\, b^2 - \frac{A\, b^2\, \lambda^2}{T\, b\,(1-b)\,\lambda^2 + A} + \frac{2}{T^2} \sum_{s=2}^{T_b} \sum_{t=s}^{T} \gamma(t-s+1), \\
&+ \frac{2}{T^2} \frac{T\, b^2\, \lambda^2 \left[T\, b^2\, \lambda^2 - 2\left(T\, b\, \lambda^2 + A\right)\right]}{\left[T\, b\,(1-b)\,\lambda^2 + A\right]^2} \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1), \\
&- \frac{2}{T^2} \frac{A^2}{\left[T\, b\,(1-b)\,\lambda^2 + A\right]^2} \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) \geq 0\,,
\end{aligned}
$$

Remark 3. It can be seen that the forecasts based on optimal weights have a lower MSFE than that applying equal weights to the observations.

2.2. Optimal window and post break window forecasts

As proposed in Pesaran and Timmermann (2007), an optimal window is chosen in which equal weights are used for the observations within the window and zero weights to the remaining observations.

$$
w_t = \begin{cases} 0, & \text{for} \quad t < T_v \\ \dfrac{1}{T - T_v + 1}, & \text{for} \quad T_v \leq t < T + 1. \end{cases} \tag{11}
$$

Suppose that the optimal window size, v, contains observations from T_v to T (inclusive), where $v = (T - T_v + 1)/T$ such that $T_v = T(1-v) + 1$.

Now, we consider the model (1), where β_t is subject to a single, discrete break at T_b,

$$
\beta_t = \begin{cases} \beta_1 & \text{for} \quad T_v \leq t \leq T_b \\ \beta_2 & \text{for} \quad T_b < t \leq T + 1. \end{cases} \tag{12}
$$

Based on the above considerations, we now mainly focus on the choice of the window size rather than the weighting of observations. Henceforth, the following theorem is obtained by using equations (5), (11) and (12), in which the general scaled MSFE is derived.

Theorem 4. In the linear regression model (1), we assume that there is equal weights within the window and zero weights to preceding observations according to equation (11), then the general scaled MSFE of the one-step-ahead forecast is computed as

$$
E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2\right] = A + \lambda^2 \left[1 - \frac{(1-b)}{v}\right]^2 \mathbb{I}(v - (1-b)) + \frac{A}{T\, v} + \frac{2}{T^2\, v^2} \sum_{s=T_v+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1), \tag{13}
$$

where $\lambda = (\beta_2 - \beta_1)/\sigma_\varepsilon$, $b = T_b/T$ and $\mathbb{I}(v-(1-b))$ is an indicator function introduced to allow flexibility in cases whether the window contain a break or not, and equals to 1 if $v > (1-b)$ and 0 otherwise.

First, we obtain the simplified form of one-step-ahead forecast as

$$\begin{aligned}\hat{y}_{T+1} &= \hat{\beta}_T(w),\\ &= \beta_2\{1-\mathbb{I}(v-(1-b))\} + \mathbb{I}(v-(1-b))\left[\frac{\beta_2\,(1-b)+\beta_1\,(v-(1-b)}{v}\right] + \frac{\sigma_\varepsilon}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t,\end{aligned}$$

Next, the expression of forecast error is given by

$$\begin{aligned}e_{T+1} &= y_{T+1} - \hat{y}_{T+1},\\ &= \mathbb{I}(v-(1-b))(\beta_2-\beta_1)\left[1-\frac{(1-b)}{v}\right] + \sigma_\varepsilon\,\varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t,\end{aligned}$$

and finally the MSFE scaled by the error variance is

$$E\left[\sigma_\varepsilon^{-2}\,e_{T+1}^2\right] = A + \lambda^2\left[1-\frac{(1-b)}{v}\right]^2\mathbb{I}(v-(1-b)) + \frac{A}{T\,v} + \frac{2}{T^2\,v^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

If we consider the window that contains the break so that $\mathbb{I}(v-(1-b)) = 1$ and minimize the MSFE obtained in equation (13), the optimal window size, v^0, is:

$$v^0 = \begin{cases}\dfrac{(1-b) + \frac{4}{2\,\lambda^2\,(1-b)\,T^2}\sum\limits_{s=T_v+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}{1-\frac{A}{2\,\lambda^2\,(1-b)\,T}}, & \text{if} \quad \lambda^2 \geq \frac{A\,T}{2\,(T-T_b)\,T_b}\\ 1, \quad \text{if} \quad \lambda^2 < \frac{A\,T}{2\,(T-T_b)\,T_b}. & \end{cases} \tag{14}$$

Remark 4. Again compared with Pesaran et al. (2013), we introduce the variance and covariance terms of an error that depends on the long memory parameter d, which results to the equation (13) and through minimization leads to an increase in the optimal window size obtained in equation (14). Again this is intuitively due to the stronger correlation structure using more information from observations further in the past.

We now consider the window that contains the break, so we substitute equation (14) into equation (13), and henceforth, the resulting MSFE for the optimal window observations is stated in the theorem below.

Theorem 5. In the linear regression model (1), we assume that there is equal weights within the window and zero weights to preceding observations but now the window contains the break, then the scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2}|v_{v>(1-b)}^{0}\right] = A + \frac{A}{T\,(1-b)} - \frac{A^2}{4\,\lambda^2\,(1-b)^2\,T^2}\left[\frac{1+\dfrac{4}{\lambda^2\,(1-b)\,T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{\left(1+\dfrac{4}{2\,\lambda^2\,(1-b)\,T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right)^2}\right], \tag{15}$$

where $\lambda = (\beta_2 - \beta_1)/\sigma_\varepsilon$.

Next, we consider the windows that contain no break ($\mathbb{I}(v-(1-b)) = 0$), so we substitute the size of the windows with no break, $v_{v\leq(1-b)}^{0} = (1-b)$, into equation (13), and henceforth, the resulting MSFE for the post-break window observations is stated in the below theorem.

Theorem 6. In the linear regression model (1), we assume that there is equal weights within the window and zero weights to preceding observations but now the window contains no break, then the scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2}|v = (1-b)\right] = A\left[1+\frac{1}{T\,(1-b)}\right] + \frac{2}{T^2\,(1-b)^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \tag{16}$$

where $b = T_b/T$.

2.3. Averaging across estimation windows

The theoretical properties of the average across estimation windows (AveW) are discussed in Pesaran and Pick (2011). Using the model (1), the one-step-ahead forecast for the AveW is:

$$\hat{y}_{T+1} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(v_{(i)}), \tag{17}$$

where

$$\hat{y}_{T+1}(v_{(i)}) = \frac{1}{T-T_{v_{(i)}}+1}\sum_{t=T_{v_{(i)}}}^{T} y_t.$$

Here, we take the average over m different estimation windows containing breaks so that $\mathbb{I}(v-(1-b)) = 1$, while given uncertainty over the break dates, we begin with the minimum window, $v_{min} = 0.05$. Then, we set $v_{(i)} = (T-T_{v_{(i)}}+1)/T$ such that $T_{v_{(i)}} = T(1-v_{(i)})+1$, and using equations (1), (5), (11), (12) and

(17), the resulting MSFE for the AveW forecast is stated in the following theorem.

Theorem 7. In the linear regression model (1), we assume the average over m different estimation windows containing breaks, then the scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2} | v_{\min}\right] = A + \left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\,\mathbb{I}(v_{(i)}-(1-b))\right]^{2} + \frac{A}{m^{2}}\sum_{i=1}^{m}\frac{1+2(i-1)}{T\,v_{(i)}}$$

$$+\frac{2}{m^{2}}\sum_{i=1}^{m}\frac{1}{T^{2}\,v_{(i)}^{2}}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \tag{18}$$

where $\lambda = (\beta_2-\beta_1)/\sigma_{\varepsilon}$, $b = T_b/T$, $v_{\min} = 0.05$ and m is the number of windows.

First, we proceed with the one-step-ahead forecast for AveW

$$\begin{aligned}\hat{y}_{T+1} &= \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(v_{(i)}),\\ &= \beta_2 + \frac{\beta_2\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\,\mathbb{I}(v_{(i)}-(1-b)) - \frac{\beta_1\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\,\mathbb{I}(v_{(i)}-(1-b))\\ &+\frac{\beta_1}{m}\sum_{i=1}^{m}\mathbb{I}(v_{(i)}-(1-b)) - \frac{\beta_2}{m}\sum_{i=1}^{m}\mathbb{I}(v_{(i)}-(1-b)) + \frac{\sigma_{\varepsilon}}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\end{aligned}$$

Using the result above, the one-step-ahead forecast error for AveW is

$$\begin{aligned}e_{T+1} &= y_{T+1} - \hat{y}_{T+1},\\ &= \sigma_{\varepsilon}\,\varepsilon_{T+1} + \frac{(\beta_2-\beta_1)}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\,\mathbb{I}(v_{(i)}-(1-b)) - \frac{\sigma_{\varepsilon}}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\end{aligned}$$

and finally the MSFE for AveW forecast is

$$E\left[\sigma_{\varepsilon}^{-2}\, e_{T+1}^{2} | v_{\min}\right] = A + \left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\,\mathbb{I}(v_{(i)}-(1-b))\right]^{2} + \frac{A}{m^{2}}\sum_{i=1}^{m}\frac{1+2(i-1)}{T\,v_{(i)}}$$

$$+\frac{2}{m^{2}}\sum_{i=1}^{m}\frac{1}{T^{2}\,v_{(i)}^{2}}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

3. Simulation Results

In this section we provide a Monte Carlo simulation study of the forecasting performance of the different optimal methods proposed in this paper and compare them to the ones discussed in Pesaran et al. (2013). We examine the simulation results for a long memory time series with a single, discrete break based on the simple linear regression model (1) applied to different forecasting methods.

Initially, we simulate a fractionally integrated time series and choose stationary long memory parameters $d \in \{0.1\,;\,0.2\,;\,0.3\,;\,0.4\}$, standard break dates $b \in \{0.2\,;\,0.9\,;\,0.95\}$, break sizes $\lambda \in \{0.5\,;\,1\,;\,2\}$ and sample sizes $T \in \{250\,;\,500\,;\,1000\}$. Next, we use the simulated fractionally integrated time series to obtain the modified LW estimator of the memory parameters $\hat{d}$ as in Hou and Perron (2014) with bandwidth $m = T^{\delta}$, where $\delta = 0.8$. We report the chosen bandwidth that is said to be MSE-optimal in estimating the long memory parameters although the results are robust to other smaller bandwidths, e.g. $\delta = 0.75$. We also estimate the break dates $\hat{b}$ as suggested by Lavielle and Moulines (2000a), and conditional on these estimates, we obtain the break size estimates $\hat{\lambda}$.

Then, we use these estimates in a simple linear regression model (1) with $\hat{d}$, $\hat{b}$ and $\hat{\lambda}$ in place of d, b and λ to compute feasible forecasts and report the MSFE results for $N = 10,000$ replications.

As expected, these forecasts provide large improvements in MSFE relative to the equal weights forecasts.

We also generally observe that in all cases, the forecasting performance of the estimated optimal weight and estimated optimal window methods proposed in this paper (II) outperform the ones discussed in Pesaran et al. (2013) (I).

In Table A.2, Table A.3, Table A.4 and Table A.5, we observe that the forecasts based on optimal weights, post-break window and optimal window have the highest MSFE when the break size, λ, is small; their performance improves dramatically when λ is large. This shows that the methods are dependent on the ability to detect the break accurately, which is easier when the break size, λ, is large otherwise the detection of the break is difficult.

Additionally, we may conclude that for the breaks that are large and easily identified (i.e. large λ), the methods offer substantial improvements in their forecasting performance, except for AveW. This implies that the incorrect estimation of the break dates can markedly affect the forecast results. Therefore, accurate estimation of the break point $\hat{b}$ is extremely necessary to obtain more precise forecast results.

Lastly, we observe that as T increases the forecasts of these methods improve considerably except for AveW, whereas forecasts appear to perform relatively well for smaller T in most cases.

Furthermore, we present the power results of the DM test statistics as suggested by Diebold and Mariano (1995), this test is used for comparison of one-step-ahead forecast errors of different estimated optimal forecasting methods; both, those obtained under the methods proposed in this paper (II) and the ones discussed in Pesaran et al. (2013) (I).

The simulation studies concentrate on forecast horizon, $h = 1$ and a squared errors (SE) loss function for the DM test.

In Table A.6 and Table A.7, we observe that power results appear quite high to reject the null hypothesis of equal forecast accuracy against the alternative hypothesis that states the methods proposed in this paper (II) are more accurate than the ones discussed in Pesaran et al. (2013) (I); mainly the estimated optimal weight and estimated optimal window in our case. Power results of the DM test for other sample sizes and break dates are not shown here, but yield similar results.

4. Inflation Rate Forecasts

In this section, the performance of inflation rate forecasting is considered. Hyung et al. (2006) and Bos et al. (2002) investigate out-of-sample forecasting of US inflation rates and find evidence of long memory; their findings are the inspiration for this study. Moreover, these authors explore the possibility of developing a single model that captures both occasional structural breaks and all long memory components. Likewise, Hassler and Wolters (1995) use a model with fractional integration allowing for long memory and show evidence of long memory in monthly inflation rates across all countries. Additionally, Gadea et al. (2004) and Hsu (2005) illustrate the risks of neglecting the presence of structural breaks in the modeling of inflation rates.

We collect data from the OECD[1] and use the monthly CPI series for 10 countries namely Germany, France, Italy, Belgium, Finland, USA, Korea, GBR, India and OECDE covering the period from January 1967 to December 2017. First, we deseasonalize the data and then transform the inflation rates to π_t by taking their log differences, i.e., $\pi_t = \log(\text{CPI}_t) - \log(\text{CPI}_{t-1})$, which is common in the literature.

In our case, we observe a single break in the mean for all countries after obtaining the residual sum of squares estimator considered in Lavielle and Moulines (2000a), and we apply the modified LW estimator of Hou and Perron (2014) with bandwidth $m = T^{4/5}$ to estimate the memory parameter of the inflation series.

In Table A.8, we show that stationary long memory exists in all series under consideration.

We apply the estimated values of break date, break size and memory parameter to obtain the MSFE results under different optimal forecast methods; clearly, those obtained under the methods proposed in this paper (II) outperform the ones discussed in Pesaran et al. (2013) (I) in most cases. In this paper, we observe that the estimated AveW methods provide the best forecasts of the inflation rates in most cases. In contrast, the estimated post-break window performs poorly, displaying the highest MSFEs among all methods in all cases.

[1]Dataset from https://data.oecd.org/price/inflation-cpi.htm.

Moreover, Table A.9 presents the p-values of less than 5% in all countries; this implies that our proposed methods provide the best forecasts in comparison to those in Pesaran et al. (2013).

Figure B.1 presents the series of inflation rates for 10 countries, where the red vertical lines represent their corresponding estimated break dates. As before, we obtain the memory parameter estimate $\hat{d}$ as in Hou and Perron (2014) based on the bandwidth parameter $\delta = 0.8$. We also obtain the break date estimate $\hat{b}$ as suggested by Lavielle and Moulines (2000a), and conditional on these estimates, we obtain the break size estimate $\hat{\lambda}$.

5. Conclusion

This paper shows the advantages of incorporating long-range dependencies to obtain optimal forecasts, whenever long memory is present in the time series. In addition to Pesaran et al. (2013), the methods proposed in this paper incorporate the variance and covariance terms of an error, where the error term is the innovation process associated with long memory. This results to some improvements in the MSFEs, where through minimization an increase in the pre-break weight, a decrease in the post-break weight and an increase in the optimal window size is obtained. For that reason, there are changes in the optimal weight and optimal window methods in comparison to Pesaran et al. (2013), while the rest of the methods seem to yield no changes.

Our methods, in comparison to the ones discussed in Pesaran et al. (2013), provide superior inflation rate forecasts by incorporating adjustments based on long memory. The findings are interesting because they reveal important improvements in the minimization of the MSFE, which is our ultimate goal.

References

Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1):47–78.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2016). *Long-Memory Processes.* Springer.

Bos, C. S., Franses, P. H., and Ooms, M. (2002). Inflation, forecast intervals and long memory regression models. *International Journal of Forecasting*, 18(2):243–264.

Casini, A. (2018). Tests for forecast instability and forecast failure under a continuous record asymptotic framework. *arXiv:1803.10883.*

Casini, A. (2021). Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *arXiv:2103.02981.*

Casini, A., Deng, T., and Perron, P. (2021). Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference.

Clements, M. P. and Hendry, D. F. (2000). *Forecasting non-stationary economic time series.* mit Press.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3):253–63.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Gadea, M. D., Sabaté, M., and Serrano, J. M. (2004). Structural breaks and their trace in the memory: Inflation rate series in the long-run. *Journal of International Financial Markets, Institutions and Money*, 14(2):117–134.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, 76(2):669–705.

Granger, C. W. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1):37–45.

Hou, J. and Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182(2):309–328.

Hsu, C.-C. (2005). Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters*, 88(2):289–294.

Hyung, N., Franses, P. H., and Penm, J. (2006). Structural Breaks and Long Memory in US Inflation Rates: Do They Matter for Forecasting? *Research in International Business and Finance*, 20(1):95–110.

Inoue, A. and Rossi, B. (2011). Identifying the sources of instabilities in macroeconomic fluctuations. *Review of Economics and statistics*, 93(4):1186–1204.

Koop, G. and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, 74(3):763–789.

Lavielle, M. and Moulines, E. (2000a). Least-Squares Estimation of An Unknown Number of Shifts in a Time Series. *Journal of Time Series Analysis*, 21(1):33–59.

Lavielle, M. and Moulines, E. (2000b). Least-squares estimation of an unknown number of shifts in time series. *Journal of Time Series Analysis*, 21:33 – 59.

Maheu, J. M. and Gordon, S. (2008). Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, 23(5):553–583.

Maheu, J. M. and McCurdy, T. H. (2009). How useful are historical data for forecasting the long-run equity return distribution? *Journal of Business & Economic Statistics*, 27(1):95–112.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1):378–390.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Perron, P. and Yamamoto, Y. (2021). Testing for changes in forecast performance. *Journal of Business and Economic Statistics*, 39(1):148–165.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4):1057–1084.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29(2):307–318.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Rossi, B. (2013). Advances in forecasting under instability. In *Handbook of economic forecasting*, volume 2, pages 1203–1324. Elsevier.

Sibbertsen, P. and Kruse, R. (2009). Forecasting long memory time series under a break in persistence. *Econometrics: Single Equation Models eJournal*.

Sibbertsen, P., Leschinski, C., and Busch, M. (2018). A Multivariate Test Against Spurious Long Memory. *Journal of Econometrics*, 203(1):33–49.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Appendix A. Tables

These tables present the forecasting performance of different estimated optimal forecasting methods; both, those obtained under the methods proposed in this paper (II) and the ones discussed in Pesaran et al. (2013) (I). We use the simulated fractionally integrated time series to estimate the memory parameters with the modified LW estimator of Hou and Perron (2014) with bandwidth $m = T^{\delta}$, where $\delta = 0.8$. We also estimate the break dates $\hat{b}$ as suggested by Lavielle and Moulines (2000b), and conditional on these estimates, we obtain the break size estimates $\hat{\lambda}$. We report the relative MSFE results under different optimal forecasting methods with different estimated values of the break date, break size and memory parameter. We generally observe that the forecast of the estimated equal weight method provides the largest MSFEs among all forecasting methods. This is always true for most cases.

In Table A.5, Table A.4 and Table A.3, we observe that in most cases, there is an decrease in efficiency of the proposed methods due to the decrease in the sample size T.

Morever, in Table A.2, we observe that in most cases, the forecasting performance of the optimal proposed methods perform better than those in Table A.1, due to the increase in the actual break date, $b = 0.9$ for the time period $T = 500$.

λ		0.5	—				1		—					2		
T = 500																
b = 0.2																
Estimated Optimal weight																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.09	0.59	0.19	0.5897	0.5876	—	0.17	0.97	0.14	0.5944	0.5923	—	0.06	2.07	0.20	0.5850	0.5829
0.24	0.74	0.16	0.5926	0.5904	—	0.13	1.10	0.17	0.5890	0.5869	—	0.12	1.68	0.20	0.5865	0.5844
0.34	0.49	0.71	0.5892	0.5870	—	0.18	1.13	0.22	0.5893	0.5872	—	0.19	2.29	0.20	0.5839	0.5818
0.35	1.26	0.94	0.5662	0.5643	—	0.38	0.57	0.73	0.5887	0.5873	—	0.34	1.84	0.19	0.5863	0.5842
Estimated Post-break window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.09	0.59	0.19	0.5889	0.5889	—	0.17	0.97	0.14	0.5935	0.5935	—	0.06	2.07	0.20	0.5842	0.5842
0.24	0.74	0.16	0.5918	0.5918	—	0.13	1.10	0.17	0.5882	0.5882	—	0.12	1.68	0.20	0.5857	0.5857
0.34	0.49	0.71	0.5884	0.5884	—	0.18	1.13	0.22	0.5885	0.5885	—	0.19	2.29	0.20	0.5832	0.5832
0.35	1.26	0.94	0.5652	0.5652	—	0.38	0.57	0.73	0.5873	0.5873	—	0.34	1.84	0.19	0.5855	0.5855
Estimated Optimal window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.09	0.59	0.19	0.5904	0.5895	—	0.17	0.97	0.14	0.5950	0.5942	—	0.06	2.07	0.20	0.5856	0.5848
0.24	0.74	0.16	0.5932	0.5924	—	0.13	1.10	0.17	0.5897	0.5889	—	0.12	1.68	0.20	0.5871	0.5863
0.34	0.49	0.71	0.5898	0.5889	—	0.18	1.13	0.22	0.5899	0.5891	—	0.19	2.29	0.20	0.5845	0.5837
0.35	1.26	0.94	0.5666	0.5660	—	0.38	0.57	0.73	0.5894	0.5889	—	0.34	1.84	0.19	0.5868	0.5861
Estimated AveW																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.09	0.59	0.19	0.6005	0.6005	—	0.17	0.97	0.14	0.6039	0.6039	—	0.06	2.07	0.20	0.5957	0.5957
0.24	0.74	0.16	0.6030	0.6030	—	0.13	1.10	0.17	0.5995	0.5995	—	0.12	1.68	0.20	0.5975	0.5975
0.34	0.49	0.71	0.6002	0.6002	—	0.18	1.13	0.22	0.5997	0.5997	—	0.19	2.29	0.20	0.5960	0.5960
0.35	1.26	0.94	0.6099	0.6099	—	0.38	0.57	0.73	0.6920	0.6920	—	0.34	1.84	0.19	0.5957	0.5957

Table A.1: Relative simulation results of the MSFEs i.e. $\text{MSFE}_i/\text{MSFE}_{\text{equal}}$ of each method applied on fractionally integrated time series for the time period $T = 500$ and break date $b = 0.2$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a single, discrete break in a simple regression model.

λ	0.5		—				1			—			2			
									T = 500							
									b = 0.9							
									Estimated Optimal weight							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.06	1.49	0.98	0.5623	0.5618	—	0.08	1.09	0.88	0.2279	0.2271	—	0.34	1.82	0.90	0.0742	0.0735
0.22	0.70	0.09	0.6894	0.6883	—	0.19	0.54	0.71	0.2010	0.2003	—	0.21	2.11	0.90	0.0859	0.0845
0.29	0.77	0.65	0.5312	0.5301	—	0.29	0.92	0.79	0.3329	0.3316	—	0.29	2.41	0.91	0.0929	0.0917
0.41	1.28	0.80	0.4863	0.4850	—	0.36	1.24	0.76	0.2860	0.2851	—	0.47	2.74	0.90	0.1113	0.1085
									Estimated Post-break window							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.06	1.49	0.98	0.6104	0.6104	—	0.08	1.09	0.88	0.2286	0.2286	—	0.34	1.82	0.90	0.0737	0.0737
0.22	0.70	0.09	0.7053	0.7053	—	0.19	0.54	0.71	0.1974	0.1974	—	0.21	2.11	0.90	0.0866	0.0866
0.29	0.77	0.65	0.5333	0.5333	—	0.29	0.92	0.79	0.3354	0.3354	—	0.29	2.41	0.91	0.0927	0.0927
0.41	1.28	0.80	0.4888	0.4888	—	0.36	1.24	0.76	0.2872	0.2872	—	0.47	2.74	0.90	0.1081	0.1081
									Estimated Optimal window							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.06	1.49	0.98	0.5892	0.5840	—	0.08	1.09	0.88	0.2279	0.2275	—	0.34	1.82	0.90	0.0734	0.0730
0.22	0.70	0.09	0.6859	0.6847	—	0.19	0.54	0.71	0.2010	0.1998	—	0.21	2.11	0.90	0.0867	0.0861
0.29	0.77	0.65	0.5327	0.5320	—	0.29	0.92	0.79	0.3351	0.3339	—	0.29	2.41	0.91	0.0938	0.0930
0.41	1.28	0.80	0.4886	0.4873	—	0.36	1.24	0.76	0.2859	0.2855	—	0.47	2.74	0.90	0.1101	0.1086
									Estimated AveW							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.06	1.49	0.98	0.8286	0.8286	—	0.08	1.09	0.88	0.6749	0.6749	—	0.34	1.82	0.90	0.6352	0.6352
0.22	0.70	0.09	0.8583	0.8583	—	0.19	0.54	0.71	0.7013	0.7013	—	0.21	2.11	0.90	0.6181	0.6181
0.29	0.77	0.65	0.8096	0.8096	—	0.29	0.92	0.79	0.6926	0.6926	—	0.29	2.41	0.91	0.6332	0.6332
0.41	1.28	0.80	0.7098	0.7098	—	0.36	1.24	0.76	0.6487	0.6487	—	0.47	2.74	0.90	0.6517	0.6517

Table A.2: Relative simulation results of the MSFEs i.e. $\text{MSFE}_\text{i}/\text{MSFE}_\text{equal}$ of each method applied on fractionally integrated time series for the time period $T = 500$ and break date $b = 0.9$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a single, discrete break in a simple regression model.

λ	0.5		—				1			—			2			
									T = 250							
									b = 0.95							
									Estimated Optimal weight							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.02	1.10	0.96	0.6145	0.6120	—	0.11	1.25	0.96	0.2311	0.2307	—	0.21	2.40	0.94	0.0742	0.0715
0.28	1.69	0.94	0.5927	0.5908	—	0.15	0.80	0.89	0.2446	0.2431	—	0.37	2.39	0.95	0.0733	0.0724
0.27	1.07	0.10	0.4387	0.4374	—	0.22	0.78	0.16	0.2320	0.2304	—	0.44	2.52	0.95	0.1297	0.1288
0.35	0.96	0.80	0.4861	0.4797	—	0.46	1.19	0.44	0.4117	0.4089	—	0.45	3.21	0.95	0.1784	0.1766
									Estimated Post-break window							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.02	1.10	0.96	0.6433	0.6433	—	0.11	1.25	0.96	0.2461	0.2461	—	0.21	2.40	0.94	0.0749	0.0749
0.28	1.69	0.94	0.6297	0.6297	—	0.15	0.80	0.89	0.2475	0.2475	—	0.37	2.39	0.95	0.0739	0.0739
0.27	1.07	0.10	0.4456	0.4456	—	0.22	0.78	0.16	0.2318	0.2318	—	0.44	2.52	0.95	0.1282	0.1282
0.35	0.96	0.80	0.4942	0.4942	—	0.46	1.19	0.44	0.4115	0.4115	—	0.45	3.21	0.95	0.1792	0.1792
									Estimated Optimal window							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.02	1.10	0.96	0.6151	0.6126	—	0.11	1.25	0.96	0.2522	0.2495	—	0.21	2.40	0.94	0.0738	0.0729
0.28	1.69	0.94	0.6102	0.6040	—	0.15	0.80	0.89	0.2481	0.2465	—	0.37	2.39	0.95	0.0747	0.0734
0.27	1.07	0.10	0.4449	0.4425	—	0.22	0.78	0.16	0.2323	0.2309	—	0.44	2.52	0.95	0.1304	0.1286
0.35	0.96	0.80	0.4896	0.4836	—	0.46	1.19	0.44	0.1801	0.1787	—	0.45	3.21	0.95	0.1801	0.1786
									Estimated AveW							
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.02	1.10	0.96	0.9002	0.9002	—	0.11	1.25	0.96	0.7886	0.7886	—	0.21	2.40	0.94	0.7821	0.7821
0.28	1.69	0.94	0.8448	0.8448	—	0.15	0.80	0.89	0.8270	0.8270	—	0.37	2.39	0.95	0.7921	0.7921
0.27	1.07	0.10	0.7477	0.7477	—	0.22	0.78	0.16	0.7668	0.7668	—	0.44	2.52	0.95	0.7865	0.7865
0.35	0.96	0.80	0.7092	0.7092	—	0.46	1.19	0.44	0.7379	0.7379	—	0.45	3.21	0.95	0.7337	0.7337

Table A.3: Relative simulation results of the MSFEs i.e. $\text{MSFE}_\text{i}/\text{MSFE}_\text{equal}$ of each method applied on fractionally integrated time series for the time period $T = 250$ and break date $b = 0.95$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a single, discrete break in a simple regression model.

λ	0.5		—				1			—				2		
T = 500																
b = 0.95																
Estimated Optimal weight																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.08	1.52	0.98	0.5329	0.5323	—	0.15	1.19	0.95	0.2155	0.2152	—	0.26	1.90	0.96	0.0662	0.0659
0.21	0.67	0.09	0.6888	0.6876	—	0.17	0.36	0.71	0.2877	0.2872	—	0.31	2.08	0.95	0.0736	0.0732
0.28	0.70	0.65	0.5996	0.5977	—	0.28	0.57	0.58	0.3529	0.3507	—	0.34	2.22	0.95	0.1006	0.0999
0.38	1.40	0.80	0.4468	0.4459	—	0.34	1.04	0.76	0.3892	0.3884	—	0.47	2.89	0.95	0.1424	0.1232
Estimated Post-break window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.08	1.52	0.98	0.5806	0.5806	—	0.15	1.19	0.95	0.2200	0.2200	—	0.26	1.90	0.96	0.0663	0.0663
0.21	0.67	0.09	0.7054	0.7054	—	0.17	0.36	0.71	0.2883	0.2883	—	0.31	2.08	0.95	0.0751	0.0751
0.28	0.70	0.65	0.6083	0.6083	—	0.28	0.57	0.58	0.3581	0.3581	—	0.34	2.22	0.95	0.0995	0.0995
0.38	1.40	0.80	0.4490	0.4490	—	0.34	1.04	0.76	0.3914	0.3914	—	0.47	2.89	0.95	0.1232	0.1232
Estimated Optimal window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.08	1.52	0.98	0.5598	0.5555	—	0.15	1.19	0.95	0.2150	0.2148	—	0.26	1.90	0.96	0.0666	0.0659
0.21	0.67	0.09	0.6910	0.6897	—	0.17	0.36	0.71	0.2878	0.2873	—	0.31	2.08	0.95	0.0746	0.0742
0.28	0.70	0.65	0.6012	0.5998	—	0.28	0.57	0.58	0.3567	0.3550	—	0.34	2.22	0.95	0.1008	0.1006
0.38	1.40	0.80	0.4491	0.4477	—	0.34	1.04	0.76	0.3932	0.3894	—	0.47	2.89	0.95	0.1240	0.1234
Estimated AveW																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.08	1.52	0.98	0.8995	0.8995	—	0.15	1.19	0.95	0.8124	0.8124	—	0.26	1.90	0.96	0.7993	0.7993
0.21	0.67	0.09	0.9027	0.9027	—	0.17	0.36	0.71	0.8251	0.8251	—	0.31	2.08	0.95	0.7866	0.7866
0.28	0.70	0.65	0.8643	0.8643	—	0.28	0.57	0.58	0.8008	0.8008	—	0.34	2.22	0.95	0.7869	0.7869
0.38	1.40	0.80	0.7637	0.7637	—	0.34	1.04	0.76	0.7148	0.7148	—	0.47	2.89	0.95	0.7975	0.7975

Table A.4: Relative simulation results of the MSFEs i.e. $\mathrm{MSFE_i}/\mathrm{MSFE_{equal}}$ of each method applied on fractionally integrated time series for the time period $T = 500$ and break date $b = 0.95$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a single, discrete break in a simple regression model.

λ	0.5		—				1			—				2		
T = 1000																
b = 0.95																
Estimated Optimal weight																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.14	0.41	0.80	0.5103	0.5092	—	0.17	0.89	0.94	0.2377	0.2371	—	0.11	2.05	0.95	0.0602	0.0596
0.22	0.68	0.94	0.6586	0.6572	—	0.25	1.13	0.95	0.1995	0.1989	—	0.32	2.03	0.96	0.0582	0.0577
0.25	0.56	0.21	0.5260	0.5246	—	0.28	1.24	0.96	0.2303	0.2292	—	0.28	1.92	0.95	0.0658	0.0649
0.27	1.31	0.30	0.3751	0.3744	—	0.21	2.24	0.93	0.4073	0.4058	—	0.51	3.85	0.95	0.1155	0.1147
Estimated Post-break window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.14	0.41	0.80	0.5161	0.5161	—	0.17	0.89	0.94	0.2392	0.2392	—	0.11	2.05	0.95	0.0599	0.0599
0.22	0.68	0.94	0.6719	0.6719	—	0.25	1.13	0.95	0.1995	0.1995	—	0.32	2.03	0.96	0.0580	0.0580
0.25	0.56	0.21	0.5275	0.5275	—	0.28	1.24	0.96	0.2301	0.2301	—	0.28	1.92	0.95	0.0658	0.0658
0.27	1.31	0.30	0.3739	0.3739	—	0.21	2.24	0.93	0.4091	0.4091	—	0.51	3.85	0.95	0.1154	0.1154
Estimated Optimal window																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.14	0.41	0.80	0.5144	0.5137	—	0.17	0.89	0.94	0.2382	0.2377	—	0.11	2.05	0.95	0.0599	0.0586
0.22	0.68	0.94	0.6597	0.6586	—	0.25	1.13	0.95	0.1988	0.1982	—	0.32	2.03	0.96	0.0579	0.0566
0.25	0.56	0.21	0.5259	0.5251	—	0.28	1.24	0.96	0.2304	0.2295	—	0.28	1.92	0.95	0.0661	0.0653
0.27	1.31	0.30	0.3758	0.3749	—	0.21	2.24	0.93	0.4086	0.4074	—	0.51	3.85	0.95	0.1156	0.1148
Estimated AveW																
$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	—	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II
0.14	0.41	0.80	0.8947	0.8947	—	0.17	0.89	0.94	0.8325	0.8325	—	0.11	2.05	0.95	0.8018	0.8018
0.22	0.68	0.94	0.8599	0.8599	—	0.25	1.13	0.95	0.8274	0.8274	—	0.32	2.03	0.96	0.8019	0.8019
0.25	0.56	0.21	0.7863	0.7863	—	0.28	1.24	0.96	0.8131	0.8131	—	0.28	1.92	0.95	0.7833	0.7833
0.27	1.31	0.30	0.7854	0.7854	—	0.21	2.24	0.93	0.7883	0.7883	—	0.51	3.85	0.95	0.7784	0.7784

Table A.5: Relative simulation results of the MSFEs i.e. $\mathrm{MSFE_i}/\mathrm{MSFE_{equal}}$ of each method applied on fractionally integrated time series for the time period $T = 1000$ and break date $b = 0.95$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a single, discrete break in a simple regression model.

Table A.6 and Table A.7 present the power results of the Diebold-Mariano (DM) test statistics as suggested by Diebold and Mariano (1995), this test is used for comparison of one-step-ahead forecast errors of different estimated optimal forecasting methods; both, those obtained under the methods proposed in this paper (II) and the ones discussed in Pesaran et al. (2013) (I). The simulation studies concentrate on forecast horizon, $h = 1$ and a squared errors (SE) loss function for the DM test. The power results appear quite high to reject the null hypothesis of equal forecast accuracy against the alternative hypothesis that states the methods proposed in this paper (II) are more accurate than the ones discussed in Pesaran et al. (2013) (I); mainly the estimated optimal weight and estimated optimal window in our case. Power results of the DM test for other sample sizes and break dates are not shown here, but yield similar results.

	Estimated Optimal weight			Estimated Optimal window		
λ	0.5	1	2	0.5	1	2
$d = 0.1$	0.9404	0.9997	1.0000	0.9342	0.9878	0.9865
$d = 0.2$	0.8679	0.9744	1.0000	0.8612	0.9686	0.9963
$d = 0.3$	0.8191	0.9000	0.9955	0.8054	0.8974	0.9957
$d = 0.4$	0.7375	0.7928	0.8996	0.7451	0.8008	0.9104

Table A.6: Power results of the DM test statistics. We set $T = 500$, $b = 0.2$, $h = 1$ and SE loss function.

	Estimated Optimal weight			Estimated Optimal window		
λ	0.5	1	2	0.5	1	2
$d = 0.1$	0.9861	1.0000	1.0000	0.9606	0.9838	0.9831
$d = 0.2$	0.9392	0.9979	1.0000	0.9144	0.9864	0.9970
$d = 0.3$	0.8741	0.9641	0.9993	0.8526	0.9484	0.9990
$d = 0.4$	0.7635	0.8533	0.9550	0.7690	0.8531	0.9562

Table A.7: Power results of the DM test statistics. We set $T = 500$, $b = 0.9$, $h = 1$ and SE loss function.

This table presents the empirical application to inflation rates forecasting performance for 10 countries. We observe that the estimated AveW methods provide the best forecasts of the inflation rates in most cases. In contrast, the estimated post-break window performs poorly, displaying the highest MSFEs among all methods in all cases.

Methods	Estimate Values			Optimal Weight		Postbreak Window		Optimal Window		AveW	
	$\hat{d}$	$\hat{\lambda}$	$\hat{b}$	I	II	I	II	I	II	I	II
DEU	0.0175	0.2222	0.3033	2.4413	2.4411	2.7254	2.7254	2.5129	2.5128	2.6021	2.6021
FRA	0.0685	0.5583	0.3607	0.4886	0.4880	0.5398	0.5398	0.5142	0.5134	0.1933	0.1933
ITA	0.1704	0.6808	0.3623	0.1898	0.1886	0.2077	0.2077	0.1973	0.1967	0.0804	0.0804
BEL	0.0109	0.3679	0.3590	0.0024	0.0016	0.0109	0.0109	0.0044	0.0028	0.0147	0.0147
FIN	0.0237	0.5601	0.3590	0.1301	0.1285	0.1197	0.1197	0.1259	0.1243	0.1915	0.1915
USA	0.2219	0.3652	0.3049	0.3391	0.3357	0.3080	0.3080	0.3273	0.3251	0.3391	0.3391
KOR	0.1306	0.9355	0.2885	0.0138	0.0123	0.0154	0.0154	0.0143	0.0138	0.0014	0.0014
GBR	0.1284	0.6534	0.3016	0.0100	0.0089	0.0145	0.0145	0.0098	0.0083	0.0003	0.0003
IND	0.3125	0.3840	0.0852	1.0922	1.0881	1.1056	1.1056	1.0954	1.0885	1.0146	1.0146
OECDE	0.0865	0.4942	0.5836	0.0256	0.0249	0.0376	0.0376	0.0292	0.0287	0.0354	0.0354

Table A.8: Relative MSFE results i.e. $\mathrm{MSFE_i}/\mathrm{MSFE_{equal}}$ for inflation rates under different forecast optimal methods, with a single, discrete break and sample size $T = 610$, break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$, and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{4/5}$ in a simple regression model.

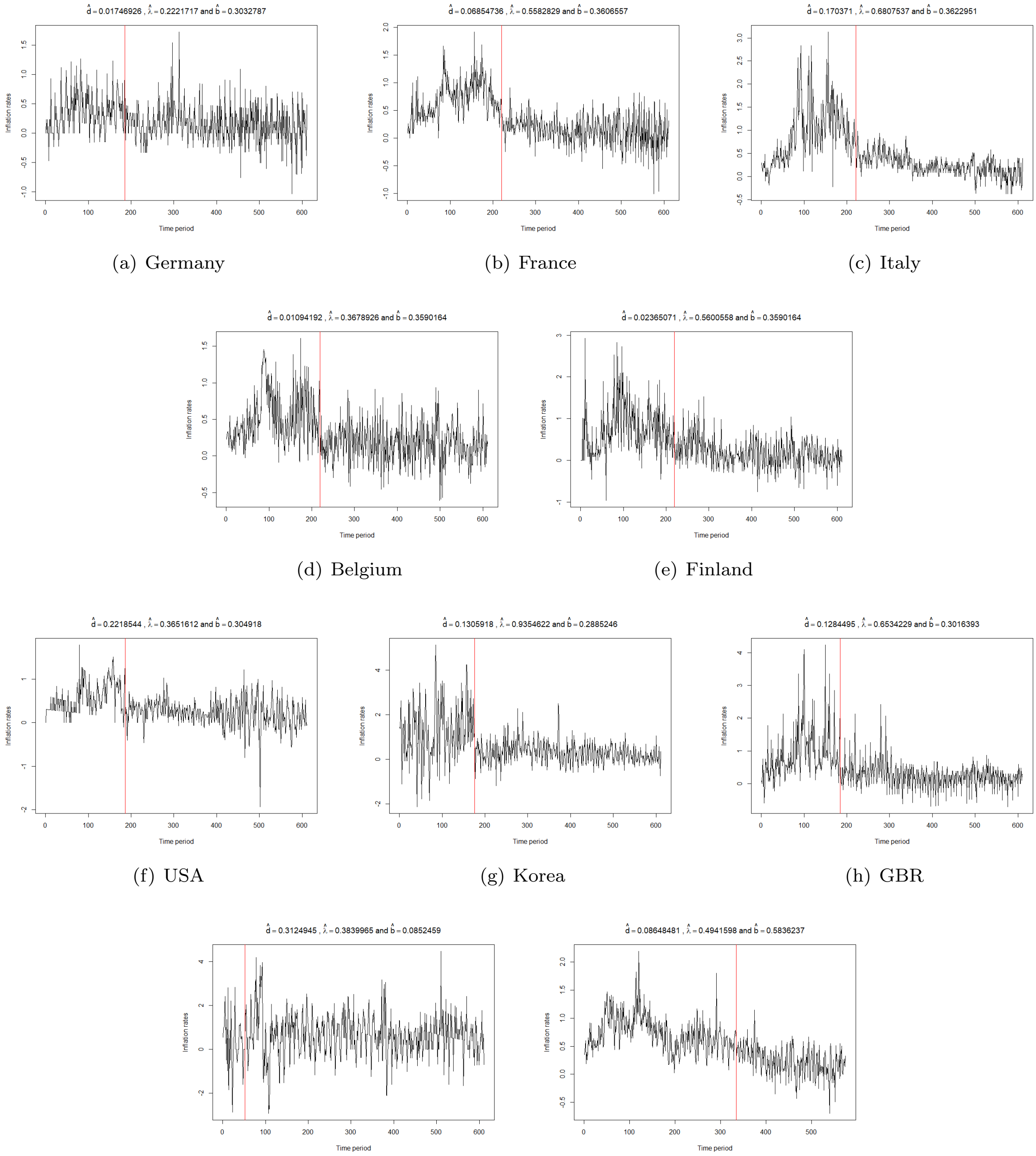
In the table below, we observe that all countries have p-values less than 5%; this implies that our proposed methods provide the best forecasts in comparison to those in Pesaran et al. (2013).

Country	DEU	FRA	ITA	BEL	FIN	USA	KOR	GBR	IND	OECDE
Estimated Optimal weight	0.0422	0.0272	0.0036	0.0413	0.0155	0.0157	0.0145	0.0102	0.0203	0.0392
Estimated Optimal window	0.0480	0.0316	0.0045	0.0282	0.0124	0.0238	0.0179	0.0156	0.0091	0.0421

Table A.9: P-values for the DM test statistic (P-value) for different countries.

Appendix B. Figures

Figure B.1 present the series of inflation rates for 10 countries, where the red vertical lines represent their corresponding estimated break dates.



(a) Germany (b) France (c) Italy

(d) Belgium (e) Finland

(f) USA (g) Korea (h) GBR

(i) India (j) OECDE

Figure B.1: Inflation Rates for 10 countries with their respective memory estimate $\hat{d}$, break size estimate $\hat{\lambda}$ and break date estimate $\hat{b}$. The red vertical line indicate the break point estimate.