Essays on Fractional Cointegration and Long Memory Time Series

Der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

— Doctor rerum politicarum —

genehmigte Dissertation

von

M.Sc. Mwasi Paza Mboya

geboren am 24.09.1994 in Dar-es-Salaam (Tansania)

2022

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Korreferent: Prof. Dr. Philipp Otto, Leibniz Universität Hannover

Tag der Promotion: 16.06.2022

Acknowledgments

First and foremost, praises and thanks to the ALMIGHTY GOD for his gift of life, health, love, protection, knowledge, wisdom and guidance which have been the spear point towards the completion of the research successfully.

I would like to express my deep and sincere gratitude to my supervisor and co-author Prof. Dr. Philipp Sibbertsen for giving me the opportunity to do the research and providing invaluable guidance throughout the research. His liveliness, honesty, patience and immense knowledge has deeply inspired me. His guidance and advice carried me through all stages of the research. It was a great privilege and honor to work and learn under his guidance. I could not have imagined having a better mentor for my Ph.D study.

A debt of gratitude is also owed to my dissertation committee. Prof. Dr. Philipp Otto, my second reviewer, Prof. Dr. Annika Herr, my chair, and Dr. Steven Gronau, my advisor, went above and beyond to help me reach my goal.

My appreciation also extends to my colleagues and coworkers who have willingly helped me out with my accomplishments.

Last but not the least, I am grateful to my parents for their love, sacrifices and prayers. For bringing me up, educating me and preparing me for the future. I would also like to give special thanks to my brother and sister for their continuous support and understanding. Thank you for your unending inspiration and prayers.

Abstract

This dissertation contains three essays on distinguishing between structural breaks under long memory, testing for fractional cointegration relationship between the financial markets and developing optimal forecast methods under long memory in the presence of a discrete structural break. Chapter 1 introduces the concepts of long memory, fractional cointegration and briefly describes the rest of the chapters.

Chapter 2 suggests a testing procedure to discriminate between stationarity, a break in the mean and a break in persistence in a time series that may exhibit long memory is introduced. The asymptotic properties of test statistics based on the CUSUM statistic are studied. In a Monte Carlo study we further analyze the finite sample properties of the procedure. An application to inflation rates shows the potential of our procedure for future research.

Chapter 3 revisits the question whether volatilities of different markets and trading zones have a long-run equilibrium in the sense that they are fractionally cointegrated. We consider the U.S., Japanese and German stock, bond and foreign exchange markets to see whether there is fractional cointegration between the markets in one trading zone or for one market across trading zones. Also the other combinations of different markets in different trading zones are considered. Applying a purely semiparametric approach through the whole analysis shows fractional cointegration can only be found for a small minority of different cases. Investigating further we find that all volatility series show persistence breaks during the observation period which may be a reason for different findings in previous studies.

Finally, we develop methods in Chapter 4 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, Changing Persistence, High-frequency Data, Semiparametric Estimation, Fractional Cointegration, Realized Volatility, Structural Break, ARFIMA Model

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The concept of long memory plays a vital role in modelling and forecasting of macroeconomic variables. In many fields of study such as econometrics and statistics, it is well known that time series exhibit long memory, for instance inflation rates (Hsu 2005), interest rates (Tsay 2000), volatilities (Lu and Perron 2010) or trading volumes (Fleming and Kirby 2011). Long memory can be described by an hyperbolic decay of the autocorrelation function or spectral density that is unbounded and characterised by a pole in the periodogram at Fourier frequencies close to zero. Loosely speaking, the time series is said to possess long-range dependence if the level of statistical dependence remains significant between very distant points. Moreover, long memory time series usually coupled with fractional integration was first introduced by C. W. Granger 1980 and C. W. Granger and Joyeux 1980 to provide a theoretical explanation for the hyperbolic decay of sample correlograms in certain empirical contexts.

In the multivariate case, the natural extension of fractional integration is the concept of fractional cointegration. Fractional cointegration is a generalization of standard cointegration, which allows the order of integration to take fractional values. Engle and C. W. J. Granger 1987 suggested that the two series are fractionally cointegrated if both processes are fractional integrated and there exists a linear combination of them such that the cointegrating residual is fractional integrated with a lower order.

The main contribution in this dissertation is to introduce the new procedure in Chapter 2 that discriminate the two forms of a break and later emphasizes the importance of detecting the structural breaks under long memory in the last two chapters. First, in Chapter 3, we consider an intensive investigation of fractional cointegration relationships between different financial markets and asset classes, where the estimation methods adopted through the whole analysis applies a purely semiparametric approach, in addition changes in persistence is taken into account in the analysis to avoid spurious fractional cointegration. Then, in Chapter 4, we develop methods to obtain optimal forecast under long memory in the presence of a discrete structural break and we confirm that the proposed methods substantially improve the forecasting performance. This dissertation presents the three self-contained chapters that appear to be connected.

Chapter 2 introduces a novel procedure to discriminate between no structural break, a break in the mean or a break in persistence. To detect structural breaks in time series, we devise to use the most popular CUSUM-based test statistics as inspired by (Aue, Hörmann, et al. 2009; Aue and Horváth 2013; Shao and Zhang 2010). These test statistics are based on functionals of the partial sums of fractionally integrated processes that satisfy the functional central limit theorem. The procedure consists of two steps: First, we detect whether or not the structural break exists at all. After having detected a break in mean or a break in persistence in the first step. Afterwards in a second step we distinguish these two forms of a break. The test procedure in this essay is inspired by Aue, Horváth, et al. 2009. We examine the finite sample properties of the test in a Monte Carlo study. We further illustrate the potential of our procedure with an application to inflation rates.

Chapter 3 reexamines the paper by A. Clements et al. 2016 to determine whether the volatilities of different markets and trading zones have a long-run equilibrium in the sense that they are fractionally cointegrated. This essay believes that the results by A. Clements et al. 2016 are driven by the combination of a semiparametric estimation of the memory parameters and a parametric modelling of the fractional cointegration relation. For this reason, we apply a purely semiparametric approach throughout the whole analysis. First, we start our analysis by estimating the order of fractional integration for each series using the exact local Whittle (ELW) estimator of Shimotsu 2010 with bandwidth $m = T^{\delta}$, where $\delta = 0.75$ that allow consistent estimation in presence of low frequency contaminations. We find that the estimated memory parameters of all series is in stationary region which makes the concept of fractional cointegration questionable for these asset classes. This is in line with the findings in Wenger et al. 2018. This contradicts the results by A. Clements et al. 2016, may be due to accounting for low frequency contaminations which leads to an upward bias in the memory estimates. In addition, Nguyen et al. 2020 found different memory parameters in different markets. We therefore test for the equality of the memory parameter between the series as suggested by Robinson and Yajima 2002 and thus rejects the null hypothesis of a common memory parameter in most cases. This will exclude fractional cointegration, since the equality of the memory parameters is the fundamental assumption of fractional cointegration.

Then, we proceed to testing for fractional cointegration on the remaining cases using the semiparametric tests for fractional cointegration, namely (Chen and Hurvich 2006; Souza et al. 2018; Wang et al. 2015) and find that the null hypothesis of no fractional cointegration is rejected in few cases. These findings are in contrast with A. Clements et al. 2016, who found fractional cointegration between different markets and trading zones in all cases. Finally, we extend further our analysis of fractional cointegration by investigating the existence of breaks in persistence since many studies have focused on the impact of financial crises on the volatility spillover. We therefore test for structural breaks by performing the regression-based Lagrange Multiplier test introduced by Martins and Rodrigues 2014 and find evidence of a break in persistence for all series during the global financial crisis and European debt crisis. The majority of the shifts suggest a decrease in persistence following the breakpoint. We, then, apply rolling window regressions to gain further insights into the dynamics of volatility persistence among the financial series and the respective trading zones. Therefore, we re-apply the persistence testing according to the estimated breakpoints and observe shifts in the order of integration during different periods. A possible reason for different findings in previous studies are presence of persistence breaks around global financial crisis. We therefore commit to taking into account the frequent changes in persistence over time to avoid spurious fractional cointegration in the analysis.

Chapter 4 develops methods to obtain optimal forecast under long memory in the presence of a discrete structural break based on different weighing schemes for the observations. This essay adapts different forecasting methods discussed in M. H. Pesaran, Pick, et al. 2013 by introducing long memory in such a setting. This develops the existence of variance and covariance terms of an error that 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, where through minimization an increase of the pre-break weight, a decrease in the post-break weight and an increase in the optimal window size is observed. We, then, confirm that our proposed methods substantially improve the forecasting performance under long memory by using Monte Carlo simulations. We further present an empirical application to inflation rates that emphasizes the importance of our methods.

CHAPTER 2

Distinguishing between Breaks in the Mean and Breaks in Persistence under Long Memory

Co-authored with Simon Wingert and Philipp Sibbertsen.

Published in Economics Letters.

https://doi.org/10.1016/j.econlet.2020.109338

CHAPTER 3

Volatility transmission across financial markets: A semiparametric analysis

Co-authored with Theoplasti Kolaiti and Philipp Sibbertsen.
Published in Journal of Risk and Financial Management.

https://doi.org/10.3390/jrfm13080160

CHAPTER 4

Optimal forecasts in the presence of discrete structural breaks under long memory

Co-authored with Philipp Sibbertsen.

Under revision in Journal of Forecasting.

4.1 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 M. P. Clements and Hendry 2000; Rossi 2013; M. P. 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 M. H. Pesaran, Pettenuzzo, 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 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.

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, M. H. Pesaran, Pick, et al. 2013 suggest defining optimal weights for each pre-break and post-break observation. However, M. H. 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, M. H. Pesaran and Timmermann 2007 use average forecasts across estimation windows (AveW) when time and size of the break is uncertain, which as H. 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 C. W. Granger and Ding 1996, C. W. 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 or a multivariate extension is Sibbertsen, Leschinski, 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 M. H. Pesaran, Pick, 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 M. H. Pesaran, Pick, 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 M. H. Pesaran, Pick, 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 2000 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 M. H. Pesaran, Pick, 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 many cases much lower under the proposed methods than those discussed in M. H. Pesaran, Pick, 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 apply different forecasting methods, to both proposed methods in this paper and the ones discussed in M. H. Pesaran, Pick, et al. 2013 for comparison, to forecast the real inflation rates for Germany and Australia 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 M. H. Pesaran, Pick, et al. 2013 in most cases.

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 4.2 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 4.3 conduct Monte Carlo experiments that compares the forecasting performance of different proposed methods with the ones discussed in M. H. Pesaran, Pick, et al. 2013. The results and discussion of the empirical application of our findings are presented in section 4.4. Section 4.5 concludes. All proofs are gathered in the appendix.

4.2 A Single, Discrete Break in a Simple Regression Model

Consider the linear regression model:

$$y_t = \beta_t + \sigma_\varepsilon \, \varepsilon_t, \qquad t = 1, \ldots, T+1 \tag{4.1}$$

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, $1 < T_b < T$:

$$\beta_t = \begin{cases} \beta_1 & \text{for} \quad t \leq T_b \\ \beta_2 & \text{for} \quad T_b < t \leq T+1 \end{cases} \tag{4.2}$$

Let ε_t be a long memory process generated according to the ARFIMA(p,d,q) model as proposed by C. W. 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 $\varepsilon_t \sim \text{ARFIMA}(0,d,0)$ because of the power-like behavior of its covariance function, where ε_t has mean $E[\varepsilon_t] = 0$, the covariance is given by:

$$Cov[\varepsilon_t, \varepsilon_{t+k}] = E[\varepsilon_t, \varepsilon_{t+k}] = \gamma(k) = \sigma_\varepsilon^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(1+k-d)\Gamma(1-k-d)}, \qquad t = 1, \ldots, T, \quad (4.3)$$

and the variance as:

$$Var[\varepsilon_t] = E[\varepsilon_t^2] = \gamma(0) = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}, \qquad t = 1, \ldots, T, \quad (4.4)$$

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 M. H. Pesaran, Pick, et al. 2013 is given by:

$$\hat{\beta}_T(w) = \sum_{t=1}^{T} w_t y_t, \tag{4.5}$$

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 (4.1), the scaled MSFE of the one-step-ahead forecast is generally computed as

$$E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\right] = A + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + A \sum_{t=1}^{T} w_t^2 + 2 \sum_{s=2}^{T} \sum_{t=s}^{T} w_{s-1} w_t \gamma(t-s+1), \tag{4.6}$$

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 (4.1), (4.2) and (4.5) 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 (4.1), (4.2) and (4.5), 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

$$E\left[\sigma_\varepsilon^{-2}\, e_{T+1}^2 | w_t^{\text{equal}}\right] = A + \lambda^2 b^2 + \frac{A}{T} + \frac{2}{T^2}\sum_{s=2}^{T}\sum_{t=s}^{T} \gamma(t-s+1), \tag{4.7}$$

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 (4.6), 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.

4.2.1 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 (4.6), we obtain the optimal weights by minimizing the equation subject to $\sum_{t=1}^{T} w_t = 1$. The first derivatives are:

For $t \leq T_b$

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

For $T_b < t \leq T$

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

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:

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

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 $t \leq T_b$

$$w_1 = \frac{1}{T}\frac{A}{T\,b\,(1-b)\,\lambda^2 + A}, \tag{4.8}$$

For $T_b < t \leq T$

$$w_2 = \frac{1}{T}\frac{T\,b\,\lambda^2 + A}{T\,b\,(1-b)\,\lambda^2 + A}, \tag{4.9}$$

Remark 2. In comparison to M. H. Pesaran, Pick, 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 (4.6) and through minimization leads to an increase in the prebreak weight and decrease in the postbreak weight obtained in equation (4.8) and equation (4.9), 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 (4.8) and (4.9) in equation (4.6), 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 (4.8) and (4.9), we obtain the reduced form of scaled MSFE for the optimal weights:

$$\begin{aligned} E\left[\sigma_\varepsilon^{-2} e_{T+1}^2 | w_1, w_2\right] =& A + (T_b^2 \lambda^2 + T_b A) w_1^2 + (T - T_b) A w_2^2 + 2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} w_{s-1} w_t \gamma(t-s+1) \\ &+ 2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} w_{s-1} w_t \gamma(t-s+1), \\ =& A\left[1 + \frac{1}{T} \frac{T b \lambda^2 + A}{T b (1-b) \lambda^2 + A}\right] + 2 w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) \\ &+ 2 w_2^2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1), \\ =& A(1 + w_2) + 2\left[w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) + w_2^2 \sum_{s=Tb+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1)\right], \end{aligned} \tag{4.10}$$

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 (4.7) and (4.10) 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.

4.2.2 Optimal window and post break window forecasts

As proposed in M. H. 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.

$$
w_t = \begin{cases} 0, & \text{for} \quad t < T_v \\ \dfrac{1}{T - T_v + 1}, & \text{for} \quad T_v \leq t < T+1. \end{cases} \tag{4.11}
$$

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 (4.1), where β_t is subject to a single, discrete break at T_b,

$$\beta_t = \begin{cases} \beta_1 & \text{for} \quad T_v \leq t \leq T_b \\ \beta_2 & \text{for} \quad T_b < t \leq T+1. \end{cases} \tag{4.12}$$

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 (4.5), (4.11) and (4.12), in which the general scaled MSFE is derived.

Theorem 4. In the linear regression model (4.1), we assume that there is equal weights within the window and zero weights to preceding observations according to equation (4.11), then the general scaled MSFE of the one-step-ahead forecast is computed as

$$E\left[\sigma_\varepsilon^{-2} e_{T+1}^2\right] = A + \lambda^2 \left[1 - \frac{(1-b)}{v}\right]^2 \mathbb{I}(v-(1-b)) + \frac{A}{T\,v} + \frac{2}{T^2 v^2} \sum_{s=T_v+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1), \tag{4.13}$$

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 (4.13), the optimal window size, v^0, is:

$$v^0 = \begin{cases} \dfrac{(1-b) + \frac{4}{2\lambda^2(1-b)T^2}\sum\limits_{s=T_v+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}{1-\frac{A}{2\lambda^2(1-b)T}}, & \text{if} \quad \lambda^2 \geq \frac{A\,T}{2(T-T_b)\,T_b} \\ 1, \quad \text{if} \quad \lambda^2 < \frac{A\,T}{2(T-T_b)\,T_b}. \end{cases} \tag{4.14}$$

Remark 4. Again compared with M. H. Pesaran, Pick, 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 (4.13) and through minimization leads to an increase in the optimal window size obtained in equation (4.14). 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 (4.14) into equation (4.13), and henceforth, the resulting MSFE for the optimal window observations is stated in the theorem below.

Theorem 5. In the linear regression model (4.1), 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

$$E\left[\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | v_{v>(1-b)}^{0}\right] = A + \frac{A}{T(1-b)} - \frac{A^2}{4\lambda^2(1-b)^2 T^2}\left[\frac{1+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{\left(1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right)^2}\right], \tag{4.15}$$

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 (4.13), and henceforth, the resulting MSFE for the post-break window observations is stated in the below theorem.

Theorem 6. In the linear regression model (4.1), 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

$$E\left[\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | v = (1-b)\right] = A\left[1+\frac{1}{T(1-b)}\right] + \frac{2}{T^2(1-b)^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \tag{4.16}$$

where $b = T_b/T$.

4.2.3 Averaging across estimation windows

The theoretical properties of the average across estimation windows (AveW) are discussed in H. Pesaran and Pick 2011. Using the model (4.1), the one-step-ahead forecast for the AveW is:

$$\hat{y}_{T+1} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(v_{(i)}), \tag{4.17}$$

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 (4.1), (4.5), (4.11), (4.12) and (4.17), the resulting MSFE for the AveW forecast is stated in the following theorem.

Theorem 7. In the linear regression model (4.1), we assume the average over m different estimation windows containing breaks, then the scaled MSFE of the one-step-ahead forecast is computed as

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

$$+\frac{2}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \tag{4.18}$$

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

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

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

4.3 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 M. H. Pesaran, Pick, 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 (4.1) 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 $b \in \{0.1; 0.2\}$, break sizes $\lambda \in \{0.5; 1; 2\}$ and sample sizes $T \in \{250; 300; 500; 1000\}$. 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 $\delta = 0.75$. 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 2000, and conditional on these estimates, we obtain the break size estimates $\hat{\lambda}$.

Then, we use these estimates in a simple linear regression model (4.1) 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.

In Table 4.1, we 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 M. H. Pesaran, Pick, et al. 2013 (I), with the exception of a few cases. However, in Table 4.2, Table 4.3 and Table 4.4, we observe a decrease in the efficiency of the proposed methods due to the decrease in sample size T.

Moreover, we observe a significant decrease in the efficiency of the proposed methods due to the increase in the modified LW estimates, $\hat{d}$. In conclusion, accurate estimation of the long memory parameter $\hat{d}$ should be taken into account to obtain more precise forecast results.

Incorrect estimation of the break dates can markedly affect the forecast results. We observe in Table 4.5 that the proposed methods perform well due to the increase in the actual break date, $b = 0.2$. Therefore, accurate estimation of the break point $\hat{b}$ is extremely necessary to obtain more precise forecast results.

4.4 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 Germany and Australia 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., $\pi_t = \log(\mathrm{CPI}_t) - \log(\mathrm{CPI}_{t-1})$, which is common in the literature.

In our case, we observe a single break in the mean for both countries after obtaining the residual sum of squares estimator considered in Lavielle and Moulines 2000, and we apply the modified LW estimator of Hou and Perron 2014 with bandwidth $m = T^{3/4}$ to estimate the memory parameter of the inflation series.

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

In Table 4.6, we show that stationary long memory exists in both 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 M. H. Pesaran, Pick, et al. 2013 (I) in most cases. In this paper, we observe that the estimated optimal window and estimated AveW methods provide the best forecasts of the inflation rates of Germany and Australia, respectively. In contrast, the estimated post-break window performs poorly, displaying the highest MSFEs among all methods in both cases.

Figure 4.1 presents the series of inflation rates for Germany and Australia, respectively, where the red vertical lines represent their corresponding estimated break points, $\hat{T_b^G} = 185$ and $\hat{T_b^A} = 206$. As before, we obtain the memory parameter estimate $\hat{d}$ as in Hou and Perron 2014 based on the bandwidth parameter $\delta = 0.75$. We also obtain the break date estimate $\hat{b}$ as suggested by Lavielle and Moulines 2000, and conditional on these estimates, we obtain the break size estimate $\hat{\lambda}$.

4.5 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 M. H. Pesaran, Pick, 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 M. H. Pesaran, Pick, et al. 2013, while the rest of the methods seem to yield no changes.

Our methods, in comparison to the ones discussed in M. H. Pesaran, Pick, 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.

4.6 Appendix

Appendix A - Proofs

Proof of Theorem 1

Proof. In this case, the forecast is $\hat{y}_{T+1} = \hat{\beta}_T(w)$ where $\hat{\beta}_T(w) = \sum_{t=1}^{T} w_t\, y_t$ then we obtain the derivation of the simplified expression below

$$
\begin{aligned}
\hat{\beta}_T(w) - \beta_T &= \hat{\beta}_T(w) - \beta_2, \\
&= \sum_{t=1}^{T} w_t\, y_t - \beta_2 \sum_{t=1}^{T} w_t, \\
&= \sum_{t=1}^{T} w_t\, (\beta_t + \sigma_\varepsilon\, \varepsilon_t) - \beta_2 \sum_{t=1}^{T} w_t, \\
&= \sum_{t=1}^{T} w_t\, \beta_t + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t - \beta_2 \sum_{t=1}^{T} w_t, \\
&= \sum_{t=1}^{T_b} w_t\, \beta_1 + \sum_{t=T_b+1}^{T} w_t\, \beta_2 + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t - \beta_2 \sum_{t=1}^{T} w_t, \\
&= \beta_1 \sum_{t=1}^{T_b} w_t + \beta_2 \sum_{t=T_b+1}^{T} w_t - \beta_2 \sum_{t=1}^{T} w_t + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t, \\
&= \beta_1 \sum_{t=1}^{T_b} w_t - \beta_2 \left(\sum_{t=1}^{T} w_t - \sum_{t=T_b+1}^{T} w_t \right) + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t, \\
&= \beta_1 \sum_{t=1}^{T_b} w_t - \beta_2 \sum_{t=1}^{T_b} w_t + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t, \\
&= (\beta_1 - \beta_2) \sum_{t=1}^{T_b} w_t + \sigma_\varepsilon \sum_{t=1}^{T} w_t\, \varepsilon_t,
\end{aligned}
$$

Using the result above, we simply obtain the expression of the forecast error as follows:

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

Thereafter, square the forecast error above and divide the result by σ_ε^2 and apply the expected value then we obtain the derivation of the MSFE scaled by the error variance:

$$
\begin{aligned}
E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\right] &= E\left[\varepsilon_{T+1}^2\right] + E\left[\left(\frac{\beta_1 - \beta_2}{\sigma_\varepsilon}\right)^2 \left(\sum_{t=1}^{T_b} w_t\right)^2\right] + E\left[\left(\sum_{t=1}^{T} w_t \, \varepsilon_t\right)^2\right], \\
&= A + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + A \sum_{t=1}^{T} w_t^2 + 2 \sum_{s=2}^{T} \sum_{t=s}^{T} w_{s-1} \, w_t \, E(\varepsilon_{s-1} \, \varepsilon_t), \\
&= A + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + A \sum_{t=1}^{T} w_t^2 + 2 \sum_{s=2}^{T} \sum_{t=s}^{T} w_{s-1} \, w_t \, \gamma(t - s + 1),
\end{aligned}
\tag{4.19}
$$

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

□

Proof of Theorem 2

Proof. Using equation 4.19, we replace the weights by $w_t^{\text{equal}} = \frac{1}{T}$, then we obtain the derivation of the scaled MSFE for the equal weights:

$$
\begin{aligned}
E\left[\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | w_t^{\text{equal}}\right] =& A+\lambda^2\left(\sum_{t=1}^{T_b}\frac{1}{T}\right)^2 + A\sum_{t=1}^{T}\frac{1}{T^2} + 2\sum_{s=2}^{T}\sum_{t=s}^{T}\frac{1}{T^2}\gamma(t-s+1), \\
=& A+\lambda^2\left(\frac{T_b}{T}\right)^2 + A\frac{T}{T^2} + \frac{2}{T^2}\sum_{s=2}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\lambda^2\left(\frac{T_b}{T}\right)^2 + A\frac{T}{T^2} + \frac{2}{T^2}\sum_{s=2}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\lambda^2 b^2 + \frac{A}{T} + \frac{2}{T^2}\sum_{s=2}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \qquad (4.20)
\end{aligned}
$$

where $b = T_b/T$. □

Proof of Theorem 3

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

For $t \leq T_b$

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

For $T_b < t \leq T$

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

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:

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

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 $t \leq T_b$

$$w_1 = \frac{1}{T}\frac{A}{T\,b\,(1-b)\,\lambda^2 + A}, \tag{4.21}$$

For $T_b < t \leq T$

$$w_2 = \frac{1}{T}\frac{T\,b\,\lambda^2 + A}{T\,b\,(1-b)\,\lambda^2 + A}, \tag{4.22}$$

Proof. Here, we assume that the weights are constant within both pre-break observations and post-break observations in equation (4.19) to obtain the scaled MSFE:

$$E\left[\sigma_\varepsilon^{-2} e_{T+1}^2 | w_1, w_2\right] = A + (T_b\,\lambda\,w_1)^2 + T_b\,w_1^2 A + (T - T_b)\,w_2^2 A + 2\sum_{s=2}^{T}\sum_{t=s}^{T} w_{s-1}\,w_t\,\gamma(t-s+1). \tag{4.23}$$

We substitute equations (4.21) and (4.22) into equation (4.23) to obtain the derivation of the reduced form of scaled MSFE for the optimal weights:

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

$$= A + (T_b^2 \lambda^2 + T_b A) w_1^2 + (T - T_b) A w_2^2 + 2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} w_{s-1} w_t \gamma(t-s+1)$$

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

$$= A + \frac{(T_b^2 \lambda^2 + T_b A) A^2}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2} + \frac{(T A - T b A) \left(T b \lambda^2 + A\right)^2}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2}$$

$$+ 2 w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) + 2 w_2^2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1),$$

$$= A \left[1 + \frac{(T^2 b^2 \lambda^2 + T b A) A}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2} + \frac{(T - T b) \left(T b \lambda^2 + A\right)^2}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2}\right]$$

$$+ 2 w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) + 2 w_2^2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1),$$

$$= A \left[1 + \frac{(T^2 b^2 \lambda^2 + T b A) A + (T - T b) \left(T b \lambda^2 + A\right)^2}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2}\right]$$

$$+ 2 w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) + 2 w_2^2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1),$$

$$= A \left[1 + \frac{T^2 b^2 \lambda^2 A + T b A^2 + (T - T b) \left(T^2 b^2 \lambda^4 + A^2 + 2 T b \lambda^2 A\right)}{\left[T\left(T b (1-b) \lambda^2 + A\right)\right]^2}\right]$$

$$+ 2 w_1^2 \sum_{s=2}^{T_b} \sum_{t=s}^{T_b} \gamma(t-s+1) + 2 w_2^2 \sum_{s=T_b+1}^{T} \sum_{t=s}^{T} \gamma(t-s+1),$$

$$=A\left[1+\frac{\begin{array}{r}T^2b^2\lambda^2A+TbA^2+T^3b^2\lambda^4+TA^2+2T^2b\lambda^2A\\-T^3b^3\lambda^4-TbA^2-2T^2b^2\lambda^2A\end{array}}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$=A\left[1+\frac{T^3b^2\lambda^4+TA^2+2T^2b\lambda^2A-T^3b^3\lambda^4-T^2b^2\lambda^2A}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$=A\left[1+\frac{\begin{array}{r}T^3b^2\lambda^4+TA^2+2T^2b\lambda^2A-T^3b^3\lambda^4-T^2b^2\lambda^2A-T^3b^3\lambda^4\\-T^2b^2\lambda^2A+T^3b^4\lambda^4+T^3b^3\lambda^4+T^2b^2\lambda^2A-T^3b^4\lambda^4\end{array}}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$=A\left[1+\frac{\begin{array}{r}T^3b^2\lambda^4-2T^3b^3\lambda^4+T^3b^4\lambda^4+TA^2+2T^2b\lambda^2A\\-2T^2b^2\lambda^2A+T^3b^3\lambda^4+T^2b^2\lambda^2A-T^3b^4\lambda^4\end{array}}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$=A\left[1+\frac{T\left(Tb\left(1-b\right)\lambda^2+A\right)^2+T\left(Tb\left(1-b\right)\lambda^2+A\right)\left(Tb^2\lambda^2\right)}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$=A\left[1+\frac{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]\left(Tb\left(1-b\right)\lambda^2+A+Tb^2\lambda^2\right)}{\left[T\left(Tb\left(1-b\right)\lambda^2+A\right)\right]^2}\right]$$

$$+2w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),$$

$$
\begin{aligned}
&= A\left[1+\frac{\left[T\left(T\,b\,(1-b)\,\lambda^2+A\right)\right]\left(T\,b\,\lambda^2+A\right)}{\left[T\left(T\,b\,(1-b)\,\lambda^2+A\right)\right]^2}\right]+2\,w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2\,w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),\\
&= A\left[1+\frac{1}{T}\,\frac{T\,b\,\lambda^2+A}{T\,b\,(1-b)\,\lambda^2+A}\right]+2\,w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+2\,w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),\\
&= A\,(1+w_2)+2\left[w_1^2\sum_{s=2}^{T_b}\sum_{t=s}^{T_b}\gamma(t-s+1)+w_2^2\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right], \qquad (4.24)
\end{aligned}
$$

□

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 (4.20) and (4.24) 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}{T}\left[1 - \frac{T\,b\,\lambda^2 + A}{T\,b\,(1-b)\,\lambda^2 + A}\right] + \frac{2}{T^2}\left[1 - \frac{\left[T\,b\,\lambda^2 + A\right]^2}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\right]\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1) \\
&+ \frac{2}{T^2}\sum_{s=2}^{T_b}\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), \\
&= \lambda^2 b^2 + \frac{A}{T}\left[\frac{T\,b\,\lambda^2 - T\,b^2\,\lambda^2 + A - T\,b\,\lambda^2 - A}{T\,b\,(1-b)\,\lambda^2 + A}\right], \\
&+ \frac{2}{T^2}\left[\frac{\begin{array}{r} T^2 b^2\lambda^4 + 2T^2 b^3\lambda^4 + T^2 b^4\lambda^4 + 2A\,T\,b\,\lambda^2 \\ - 2A\,T\,b^2\lambda^2 - T^2 b^2\lambda^4 - 2A\,T\,b\,\lambda^2 \end{array}}{\left[T\,b\,(1-b)\,\lambda^2 + A\right]^2}\right]\sum_{s=T_b+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
&+ \frac{2}{T^2}\sum_{s=2}^{T_b}\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), \\
&= \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

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

For this reason, we have

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

Proof of Theorem 4

Proof. Using the given linear regression model, we derive the one-step-ahead forecast as:

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

Let us set $v = \dfrac{T-T_v+1}{T}$, such that $T_v = T(1-v)+1$ and $T_b = T b$, then we obtain a simplified form of the above expression:

$$
\begin{aligned}
\hat{y}_{T+1} &= \frac{1}{T\,v}\left[\beta_1\left(T\,b-(T(1-v)+1)+1\right)+\beta_2\left(T-T\,b\right)+\sigma_\varepsilon \sum_{t=T_v}^{T}\varepsilon_t\right], \\
&= \frac{1}{T\,v}\left[\beta_1\left(T\,b-T+T\,v\right)+\beta_2\left(T-T\,b\right)+\sigma_\varepsilon \sum_{t=T_v}^{T}\varepsilon_t\right], \\
&= \frac{\beta_1\left(b-1+v\right)+\beta_2\left(1-b\right)}{v}+\frac{\sigma_\varepsilon}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t, \\
&= \frac{\beta_2\left(1-b\right)+\beta_1\left(v-(1-b\right)}{v}+\frac{\sigma_\varepsilon}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t, \\
&= \beta_2\{1-\mathbb{I}(v-(1-b))\}+\mathbb{I}(v-(1-b))\left[\frac{\beta_2\left(1-b\right)+\beta_1\left(v-(1-b\right)}{v}\right]+\frac{\sigma_\varepsilon}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t,
\end{aligned}
\tag{4.25}
$$

where $\mathbb{I}$ is an indicator function introduced to allow flexibility in cases whether the window contain a break or not, and equals to 1 if $c > 0$ and 0 otherwise.

Using the equation (4.25), the one-step-ahead forecast error is:

$$
\begin{aligned}
\hat{e}_{T+1} &= y_{T+1} - \hat{y}_{T+1}, \\
&= \beta_{T+1} + \sigma_\varepsilon\, \varepsilon_{T+1} - \beta_2\{1 - \mathbb{I}(v-(1-b))\} - \mathbb{I}(v-(1-b)) \left[\frac{\beta_2\,(1-b) + \beta_1\,(v-(1-b)}{v}\right] - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \beta_2 + \sigma_\varepsilon\, \varepsilon_{T+1} - \beta_2\{1 - \mathbb{I}(v-(1-b))\} - \mathbb{I}(v-(1-b)) \left[\frac{\beta_2\,(1-b) + \beta_1\,(v-(1-b)}{v}\right] - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b))\,\beta_2 - \mathbb{I}(v-(1-b)) \left[\frac{\beta_2\,(1-b) + \beta_1\,(v-(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b)) \left[\beta_2 - \frac{\beta_2\,(1-b) + \beta_1\,(v-(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b)) \left[\frac{\beta_2\, v - \beta_2\,(1-b) - \beta_1\,(v-(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b)) \left[\frac{\beta_2\,(v-(1-b)) - \beta_1\,(v-(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b))(\beta_2 - \beta_1) \left[\frac{v-(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \\
&= \mathbb{I}(v-(1-b))(\beta_2 - \beta_1) \left[1 - \frac{(1-b)}{v}\right] + \sigma_\varepsilon\, \varepsilon_{T+1} - \frac{\sigma_\varepsilon}{T\,v} \sum_{t=T_v}^{T} \varepsilon_t, \qquad (4.26)
\end{aligned}
$$

Next, the expected squared forecast error normalized by σ_ε^2 is:

$$
\begin{aligned}
E\left[\sigma_\varepsilon^{-2} e_{T+1}^2\right] =& E\left[\varepsilon_{T+1}^2)\right] + \mathbb{I}(v-(1-b))\,E\left[\left(\frac{\beta_2-\beta_1}{\sigma_\varepsilon}\right)^2\left(1-\frac{(1-b)}{v}\right)^2\right] + E\left[\left(\frac{1}{T\,v}\sum_{t=T_v}^{T}\varepsilon_t\right)^2\right], \\
=& A+\lambda^2\left[1-\frac{(1-b)}{v}\right]^2 \mathbb{I}(v-(1-b)) + \frac{1}{T^2 v^2}\sum_{t=T_v}^{T} E(\varepsilon_t^2) + \frac{2}{T^2 v^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T} E(\varepsilon_{s-1}\,\varepsilon_t), \\
=& A+\lambda^2\left[1-\frac{(1-b)}{v}\right]^2 \mathbb{I}(v-(1-b)) + \frac{1}{T^2 v^2} A\,(T-T_v+1) + \frac{2}{T^2 v^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\lambda^2\left[1-\frac{(1-b)}{v}\right]^2 \mathbb{I}(v-(1-b)) + \frac{1}{T^2 v^2} A\,T\,v + \frac{2}{T^2 v^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\lambda^2\left[1-\frac{(1-b)}{v}\right]^2 \mathbb{I}(v-(1-b)) + \frac{A}{T\,v} + \frac{2}{T^2 v^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1),
\end{aligned} \tag{4.27}
$$

where $\lambda = \dfrac{\beta_2-\beta_1}{\sigma_\varepsilon}$ and $b = T_b/T$. □

Proof of Theorem 5

Proof. If we consider the window that contains the break so that $\mathbb{I}(v-(1-b)) = 1$, the optimal window size, v^0, is:

$$
v^0 = \begin{cases} \dfrac{(1-b) + \frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{1-\frac{A}{2\lambda^2(1-b)T}}, & \text{if} \quad \lambda^2 \geq \frac{A\,T}{2\,(T-T_b)\,T_b} \\ 1, \quad \text{if} \quad \lambda^2 < \frac{A\,T}{2\,(T-T_b)\,T_b}. \end{cases} \tag{4.28}
$$

Now, we substitute equation (4.28) into equation (4.27) to obtain the derivation of the scaled MSFE for the optimal window observations:

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

$$+\frac{A}{T(1-b)\dfrac{1+\dfrac{4}{2\lambda^{2}(1-b)T^{2}}\sum\limits_{s=T_{v}+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}{1-\dfrac{A}{2\lambda^{2}(1-b)T}}},$$

$$=A+\lambda^{2}\left[1-\left(1-\frac{\dfrac{A}{2\lambda^{2}(1-b)T}}{1+\dfrac{4}{2\lambda^{2}(1-b)T^{2}}\sum\limits_{s=T_{v}+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}\right)\right]^{2}$$

$$+\frac{A}{T(1-b)}\left(1-\frac{\dfrac{A}{2\lambda^{2}(1-b)T}}{1+\dfrac{4}{2\lambda^{2}(1-b)T^{2}}\sum\limits_{s=T_{v}+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}\right),$$

$$=A+\lambda^{2}\left[\frac{\dfrac{A}{2\lambda^{2}(1-b)T}}{1+\dfrac{4}{2\lambda^{2}(1-b)T^{2}}\sum\limits_{s=T_{v}+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)}\right]^{2}$$

$$+\frac{A}{T(1-b)}-\frac{\dfrac{A^{2}}{2\lambda^{2}(1-b)^{2}T^{2}}}{1+\dfrac{4}{2\lambda^{2}(1-b)T^{2}}\sum\limits_{s=T_{v}+1}^{T}\sum\limits_{t=s}^{T}\gamma(t-s+1)},$$

$$=A+\frac{\frac{A^2}{4\lambda^2(1-b)^2T^2}}{\left(1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right)^2}$$

$$+\frac{A}{T(1-b)}-\frac{\frac{A^2}{2\lambda^2(1-b)^2T^2}}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)},$$

$$=A+\frac{A}{T(1-b)}$$

$$-\frac{\frac{A^2}{2\lambda^2(1-b)^2T^2}}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\left[1-\frac{1}{2+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\right],$$

$$=A+\frac{A}{T(1-b)}$$

$$-\frac{\frac{A^2}{2\lambda^2(1-b)^2T^2}}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\left[\frac{2+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)-1}{2+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\right],$$

$$=A+\frac{A}{T(1-b)}$$

$$-\frac{\frac{A^2}{2\lambda^2(1-b)^2T^2}}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\left[\frac{1+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{2+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}\right],$$

$$
\begin{aligned}
&= A + \frac{A}{T(1-b)} \\
&- \frac{\frac{A^2}{2\lambda^2(1-b)^2T^2}}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)} \left[\frac{\frac{1}{2}+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)} \right], \\
&= A + \frac{A}{T(1-b)} - \frac{A^2}{4\lambda^2(1-b)^2T^2} \left[\frac{1+\frac{4}{\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)}{\left(1+\frac{4}{2\lambda^2(1-b)T^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1)\right)^2} \right], \quad (4.29)
\end{aligned}
$$

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

Proof of Theorem 6

Proof. If we consider the windows that contain no break so that $\mathbb{I}(v-(1-b)) = 0$, so we substitute the size of the windows with no break, $v^0_{v\leq(1-b)} = (1-b)$, into equation (4.27) and then we obtain the derivation of the scaled MSFE for the post-break window observations:

$$
\begin{aligned}
E\left[\sigma_\varepsilon^{-2} e_{T+1}^2 | v = (1-b)\right] &= A + \lambda^2 \left[1 - \frac{(1-b)}{v}\right]^2 0 + \frac{A}{T(1-b)} + \frac{2}{T^2(1-b)^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
&= A\left[1+\frac{1}{T(1-b)}\right] + \frac{2}{T^2(1-b)^2}\sum_{s=T_v+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \quad (4.30)
\end{aligned}
$$

□

Proof of Theorem 7

Proof. Using the given linear regression model, the one-step-ahead forecast for AveW is:

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

where

$$\begin{aligned}
\hat{y}_{T+1}(v_{(i)}) &= \frac{1}{T - T_{v_{(i)}} + 1}\sum_{t=T_{v_{(i)}}}^{T} y_t,\\
&= \frac{1}{T - T_{v_{(i)}} + 1}\sum_{t=T_{v_{(i)}}}^{T} (\beta_t + \sigma_\varepsilon\,\varepsilon_t),\\
&= \frac{1}{T - T_{v_{(i)}} + 1}\left[\sum_{t=T_{v_{(i)}}}^{T_b}\beta_1 + \sum_{t=T_b+1}^{T}\beta_2 + \sigma_\varepsilon\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],\\
&= \frac{1}{T - T_{v_{(i)}} + 1}\left[\beta_1\,(T_b - T_{v_{(i)}} + 1) + \beta_2\,(T - T_b) + \sigma_\varepsilon\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],
\end{aligned}$$

We set $v_{(i)} = (T - T_{v_{(i)}} + 1)/T$ such that $T_{v_{(i)}} = T(1 - v_{(i)}) + 1$ and $T_b = T\,b$, then we obtain a simplified form of the above expression:

$$\begin{aligned}
\hat{y}_{T+1}(v_{(i)}) &= \frac{1}{T\,v_{(i)}}\left[\beta_1\,(T\,b-(T\,(1-v_{(i)})+1)+1)+\beta_2\,(T-T\,b)+\sigma_\varepsilon\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],\\
&= \frac{1}{T\,v_{(i)}}\left[\beta_1\,(T\,b-T+T\,v_{(i)})+\beta_2\,(T-T\,b)+\sigma_\varepsilon\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],\\
&= \frac{\beta_1\,(b-1+v_{(i)})+\beta_2\,(1-b)}{v_{(i)}}+\frac{\sigma_\varepsilon}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\\
&= \frac{\beta_2\,(1-b)+\beta_1\,(v_{(i)}-(1-b)}{v_{(i)}}+\frac{\sigma_\varepsilon}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\\
&= \beta_2\{1-\mathbb{I}(v_{(i)}-(1-b))\}+\mathbb{I}(v_{(i)}-(1-b))\left[\frac{\beta_2\,(1-b)+\beta_1\,(v_{(i)}-(1-b)}{v_{(i)}}\right]+\frac{\sigma_\varepsilon}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,
\end{aligned} \tag{4.31}$$

By using equation (4.31), we proceed with the one-step-ahead forecast for AveW:

$$\begin{aligned}
\hat{y}_{T+1} &= \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(v_{(i)}),\\
&= \frac{1}{m}\sum_{i=1}^{m}\left[\beta_2\{1-\mathbb{I}(v_{(i)}-(1-b))\}+\mathbb{I}(v_{(i)}-(1-b))\left[\frac{\beta_2\,(1-b)+\beta_1\,(v_{(i)}-(1-b)}{v_{(i)}}\right]\right.\\
&\quad\left. + \frac{\sigma_\varepsilon}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],\\
&= \frac{1}{m}\left[\beta_2\sum_{i=1}^{m}\{1-\mathbb{I}(v_{(i)}-(1-b))\}+\beta_2\,(1-b)\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right.\\
&\quad\left. + \beta_1\sum_{i=1}^{m}\frac{(v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))+\sigma_\varepsilon\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right],\\
&= \frac{\beta_2}{m}\sum_{i=1}^{m}\{1-\mathbb{I}(v_{(i)}-(1-b))\}+\frac{\beta_2\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\\
&\quad + \frac{\beta_1}{m}\sum_{i=1}^{m}\frac{(v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))+\frac{\sigma_\varepsilon}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\\
&= \frac{\beta_2}{m}\sum_{i=1}^{m}\{1-\mathbb{I}(v_{(i)}-(1-b))\}+\frac{\beta_2\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\\
&\quad - \frac{\beta_1\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))+\frac{\beta_1}{m}\sum_{i=1}^{m}\mathbb{I}(v_{(i)}-(1-b))+\frac{\sigma_\varepsilon}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t,\\
&= \beta_2+\frac{\beta_2\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))-\frac{\beta_1\,(1-b)}{m}\sum_{i=1}^{m}\frac{1}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\\
&\quad + \frac{\beta_1}{m}\sum_{i=1}^{m}\mathbb{I}(v_{(i)}-(1-b))-\frac{\beta_2}{m}\sum_{i=1}^{m}\mathbb{I}(v_{(i)}-(1-b))+\frac{\sigma_\varepsilon}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t, \qquad (4.32)
\end{aligned}$$

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

$$\begin{aligned}
\hat{e}_{T+1} &= y_{T+1} - \hat{y}_{T+1}, \\
&= \beta_{T+1} + \sigma_\varepsilon \, \varepsilon_{T+1} - \beta_2 - \frac{(\beta_2-\beta_1)(1-b)}{m} \sum_{i=1}^{m} \frac{1}{v_{(i)}} \mathbb{I}(v_{(i)} - (1-b)) \\
&+ \frac{(\beta_2-\beta_1)}{m} \sum_{i=1}^{m} \mathbb{I}(v_{(i)} - (1-b)) - \frac{\sigma_\varepsilon}{m} \sum_{i=1}^{m} \frac{1}{T\, v_{(i)}} \sum_{t=T_{v_{(i)}}}^{T} \varepsilon_t, \\
&= \beta_2 + \sigma_\varepsilon \, \varepsilon_{T+1} - \beta_2 - \frac{(\beta_2-\beta_1)(1-b)}{m} \sum_{i=1}^{m} \frac{1}{v_{(i)}} \mathbb{I}(v_{(i)} - (1-b)) \\
&+ \frac{(\beta_2-\beta_1)}{m} \sum_{i=1}^{m} \mathbb{I}(v_{(i)} - (1-b)) - \frac{\sigma_\varepsilon}{m} \sum_{i=1}^{m} \frac{1}{T\, v_{(i)}} \sum_{t=T_{v_{(i)}}}^{T} \varepsilon_t, \\
&= \sigma_\varepsilon \, \varepsilon_{T+1} - \frac{(\beta_2-\beta_1)(1-b)}{m} \sum_{i=1}^{m} \frac{1}{v_{(i)}} \mathbb{I}(v_{(i)} - (1-b)) \\
&+ \frac{(\beta_2-\beta_1)}{m} \sum_{i=1}^{m} \mathbb{I}(v_{(i)} - (1-b)) - \frac{\sigma_\varepsilon}{m} \sum_{i=1}^{m} \frac{1}{T\, v_{(i)}} \sum_{t=T_{v_{(i)}}}^{T} \varepsilon_t, \\
&= \sigma_\varepsilon \, \varepsilon_{T+1} + \frac{(\beta_2-\beta_1)}{m} \sum_{i=1}^{m} \left(1 - \frac{(1-b)}{v_{(i)}}\right) \mathbb{I}(v_{(i)} - (1-b)) - \frac{\sigma_\varepsilon}{m} \sum_{i=1}^{m} \frac{1}{T\, v_{(i)}} \sum_{t=T_{v_{(i)}}}^{T} \varepsilon_t, \\
&= \sigma_\varepsilon \, \varepsilon_{T+1} + \frac{(\beta_2-\beta_1)}{m} \sum_{i=1}^{m} \frac{v_{(i)} - (1-b)}{v_{(i)}} \mathbb{I}(v_{(i)} - (1-b)) - \frac{\sigma_\varepsilon}{m} \sum_{i=1}^{m} \frac{1}{T\, v_{(i)}} \sum_{t=T_{v_{(i)}}}^{T} \varepsilon_t, \quad (4.33)
\end{aligned}$$

Next, the MSFE for AveW forecast is:

$$
\begin{aligned}
E\left[\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | v_{\min}\right] =& E\left[\varepsilon_{T+1}^{2})\right] + E\left[\left(\frac{\beta_2-\beta_1}{\sigma_{\varepsilon}}\right)^{2}\left(\frac{1}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right)^{2}\right] \\
&+E\left[\left(\frac{1}{m}\sum_{i=1}^{m}\frac{1}{T\,v_{(i)}}\sum_{t=T_{v_{(i)}}}^{T}\varepsilon_t\right)^{2}\right], \\
=& A+\left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right]^{2}+\frac{1}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{t=T_{v_{(i)}}}^{T}E(\varepsilon_t^2) \\
&+\frac{2}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}E(\varepsilon_{s-1}\,\varepsilon_t), \\
=& A+\left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right]^{2}+\frac{1}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{t=T_{v_{(i)}}}^{T}E(\varepsilon_t^2) \\
&+\frac{2}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right]^{2}+\frac{1}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}T\,v_{(i)}A\left(1+2(i-1)\right) \\
&+\frac{2}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \\
=& A+\left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{v_{(i)}-(1-b)}{v_{(i)}}\mathbb{I}(v_{(i)}-(1-b))\right]^{2}+\frac{A}{m^2}\sum_{i=1}^{m}\frac{1+2(i-1)}{T\,v_{(i)}} \\
&+\frac{2}{m^2}\sum_{i=1}^{m}\frac{1}{T^2 v_{(i)}^2}\sum_{s=T_{v_{(i)}}+1}^{T}\sum_{t=s}^{T}\gamma(t-s+1), \qquad (4.34)
\end{aligned}
$$

where $\lambda = \dfrac{\beta_2 - \beta_1}{\sigma_\varepsilon}$, $b = T_b/T$ and m is the number of windows. □

Appendix B - 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 M. H. Pesaran, Pick, 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 $\delta = 0.75$. We also estimate the break dates $\hat{b}$ as suggested by Lavielle and Moulines 2000, and conditional on these estimates, we obtain the break size estimates $\hat{\lambda}$. We report the 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.

However, in Table 4.2, Table 4.3 and Table 4.4, 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 4.5, we observe that in most cases, the forecasting performance of the optimal proposed methods perform better than those in Table 4.1, due to the increase in the actual break date, $b = 0.2$ for the time period $T = 1000$.

λ	0.5				—		1				—		2			
T = 1000																
b = 0.1																
Estimated Equal 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	0.35	0.12	0.9769	0.9769	—	0.21	1.07	0.10	0.9892	0.9892	—	0.32	2.16	0.1	1.2218	1.2218
0.23	0.61	0.10	1.2573	1.2573	—	0.28	1.20	0.08	1.1759	1.1759	—	0.30	1.65	0.1	1.2407	1.2407
0.28	1.16	0.05	1.5774	1.5774	—	0.29	0.86	0.07	1.5287	1.5287	—	0.44	2.60	0.1	1.5092	1.5092
0.46	1.75	0.02	2.5214	2.5214	—	0.39	1.48	0.12	2.3473	2.3473	—	0.52	2.24	0.1	2.0668	2.0668
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	0.35	0.12	0.9515	0.9513	—	0.21	1.07	0.10	0.9275	0.9266	—	0.32	2.16	0.1	1.0114	1.0109
0.23	0.61	0.10	1.1242	1.1233	—	0.28	1.20	0.08	1.1125	1.1118	—	0.30	1.65	0.1	1.07127	1.07118
0.28	1.16	0.05	1.1638	1.1621	—	0.29	0.86	0.07	1.2954	1.2948	—	0.44	2.60	0.1	1.3813	1.3808
0.46	1.75	0.02	1.5410	1.5341	—	0.39	1.48	0.12	1.5769	1.5756	—	0.52	2.24	0.1	1.5442	1.5438
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	0.35	0.12	0.9575	0.9575	—	0.21	1.07	0.10	0.9275	0.9275	—	0.32	2.16	0.1	1.0115	1.0115
0.23	0.61	0.10	1.1326	1.1326	—	0.28	1.20	0.08	1.1128	1.1128	—	0.30	1.65	0.1	1.0712	1.0712
0.28	1.16	0.05	1.1699	1.1699	—	0.29	0.86	0.07	1.2945	1.2945	—	0.44	2.60	0.1	1.3812	1.3812
0.46	1.75	0.02	1.5442	1.5442	—	0.39	1.48	0.12	1.5796	1.5796	—	0.52	2.24	0.1	1.5446	1.5446
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	0.35	0.12	0.9399	0.9391	—	0.21	1.07	0.10	0.9274	0.9265	—	0.32	2.16	0.1	1.0114	1.0106
0.23	0.61	0.10	1.1318	1.1289	—	0.28	1.20	0.08	1.1104	1.1097	—	0.30	1.65	0.1	1.0712	1.0705
0.28	1.16	0.05	1.1920	1.1812	—	0.29	0.86	0.07	1.3287	1.3233	—	0.44	2.60	0.1	1.3813	1.3805
0.46	1.75	0.02	1.5704	1.5396	—	0.39	1.48	0.12	1.6054	1.5710	—	0.52	2.24	0.1	1.5487	1.5459
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	0.35	0.12	0.9417	0.9417	—	0.21	1.07	0.10	0.9266	0.9266	—	0.32	2.16	0.1	1.0059	1.0059
0.23	0.61	0.10	1.1869	1.1869	—	0.28	1.20	0.08	1.1053	1.1053	—	0.30	1.65	0.1	1.0642	1.0642
0.28	1.16	0.05	1.3345	1.3345	—	0.29	0.86	0.07	1.3681	1.3681	—	0.44	2.60	0.1	1.3357	1.3357
0.46	1.75	0.02	1.9106	1.9106	—	0.39	1.48	0.12	1.8258	1.8258	—	0.52	2.24	0.1	1.5669	1.5669

Table 4.1: Simulation results of the MSFEs of each method applied on fractionally integrated time series for the time period $T = 1000$ and break date $b = 0.1$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$ and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{0.75}$ in a single, discrete break in a simple regression model.

λ	0.5				—		1				—		2			
T = 500																
b = 0.1																
Estimated Equal 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
0.16	0.64	0.09	1.0929	1.0929	—	0.23	1.04	0.10	0.9735	0.9735	—	0.31	1.80	0.1	1.1066	1.1066
0.24	0.35	0.55	1.2916	1.2916	—	0.29	0.98	0.05	1.2887	1.2887	—	0.32	1.83	0.1	1.2364	1.2364
0.35	0.83	0.86	1.7357	1.7357	—	0.35	-0.44	0.55	1.5216	1.5216	—	0.41	2.00	0.1	1.4437	1.4437
0.35	0.49	0.54	2.4612	2.4612	—	0.37	0.82	0.47	2.4005	2.4005	—	0.44	2.51	0.1	2.1195	2.1195
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
0.16	0.64	0.09	1.0613	1.0607	—	0.23	1.04	0.10	0.9263	0.9243	—	0.31	1.80	0.1	0.9838	0.9826
0.24	0.35	0.55	1.0167	1.0153	—	0.29	0.98	0.05	1.2208	1.2196	—	0.32	1.83	0.1	1.0700	1.0691
0.35	0.83	0.86	1.2555	1.2523	—	0.35	-0.44	0.55	1.2752	1.2727	—	0.41	2.00	0.1	1.2365	1.2435
0.35	0.49	0.54	1.4593	1.4576	—	0.37	0.82	0.47	1.5224	1.5211	—	0.44	2.51	0.1	1.5720	1.5713
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
0.16	0.64	0.09	1.0773	1.0773	—	0.23	1.04	0.10	0.9265	0.9265	—	0.31	1.80	0.1	0.9837	0.9837
0.24	0.35	0.55	1.0335	1.0335	—	0.29	0.98	0.05	1.2239	1.2239	—	0.32	1.83	0.1	1.0702	1.0702
0.35	0.83	0.86	1.2655	1.2655	—	0.35	-0.44	0.55	1.2794	1.2794	—	0.41	2.00	0.1	1.2366	1.2366
0.35	0.49	0.54	1.4687	1.4687	—	0.37	0.82	0.47	1.5273	1.5273	—	0.44	2.51	0.1	1.5726	1.5726
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
0.16	0.64	0.09	1.0701	1.0670	—	0.23	1.04	0.10	0.9263	0.9255	—	0.31	1.80	0.1	0.9839	0.9826
0.24	0.35	0.55	1.0994	1.0880	—	0.29	0.98	0.05	1.2137	1.2104	—	0.32	1.83	0.1	1.0702	1.0691
0.35	0.83	0.86	1.2655	1.2464	—	0.35	-0.44	0.55	1.2808	1.2635	—	0.41	2.00	0.1	1.2369	1.2220
0.35	0.49	0.54	1.4731	1.4675	—	0.37	0.82	0.47	1.5667	1.5536	—	0.44	2.51	0.1	1.5724	1.4844
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
0.16	0.64	0.09	1.0743	1.0743	—	0.23	1.04	0.10	0.9168	0.9168	—	0.31	1.80	0.1	0.9831	0.9831
0.24	0.35	0.55	1.1867	1.1867	—	0.29	0.98	0.05	1.2110	1.2110	—	0.32	1.83	0.1	1.0538	1.0538
0.35	0.83	0.86	1.4582	1.4582	—	0.35	-0.44	0.55	1.3099	1.3099	—	0.41	2.00	0.1	1.2068	1.2068
0.35	0.49	0.54	1.7956	1.7956	—	0.37	0.82	0.47	1.7945	1.7945	—	0.44	2.51	0.1	1.5721	1.5721

Table 4.2: Simulation results of the MSFEs of each method applied on fractionally integrated time series for the time period $T = 500$ and break date $b = 0.1$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$ and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{0.75}$ in a single, discrete break in a simple regression model.

λ	0.5				—			1			—			2		
								T = 300								
								b = 0.1								
								Estimated Equal 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
-0.005	0.82	0.03	1.2139	1.2139	—	0.20	0.96	0.11	1.1315	1.1315	—	0.34	2.27	0.1	1.1146	1.1146
0.20	-1.11	0.03	1.3890	1.3890	—	0.26	0.48	0.15	1.2940	1.2940	—	0.39	1.93	0.1	1.2672	1.2672
0.28	1.17	0.89	1.8776	1.8776	—	0.28	0.84	0.87	1.7887	1.7887	—	0.40	1.81	0.23	1.5750	1.5750
0.54	-1.51	0.21	2.4709	2.4709	—	0.36	-0.95	0.89	2.2586	2.2586	—	0.49	1.59	0.12	2.0250	2.0250
								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
-0.005	0.82	0.03	1.0587	1.0575	—	0.20	0.96	0.11	1.0626	1.0619	—	0.34	2.27	0.1	1.0053	1.0047
0.20	-1.11	0.03	1.0713	1.0679	—	0.26	0.48	0.15	1.0899	1.0886	—	0.39	1.93	0.1	1.1302	1.1285
0.28	1.17	0.89	1.2764	1.2655	—	0.28	0.84	0.87	1.2943	1.2939	—	0.40	1.81	0.23	1.2519	1.2504
0.54	-1.51	0.21	1.3664	2.3654	—	0.36	-0.95	0.89	1.4770	1.4686	—	0.49	1.59	0.12	1.5662	1.5649
								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
-0.005	0.82	0.03	1.1004	1.1004	—	0.20	0.96	0.11	1.0642	1.0642	—	0.34	2.27	0.1	1.0051	1.0051
0.20	-1.11	0.03	1.1024	1.1024	—	0.26	0.48	0.15	1.0957	1.0957	—	0.39	1.93	0.1	1.1305	1.1305
0.28	1.17	0.89	1.2968	1.2968	—	0.28	0.84	0.87	1.3010	1.3010	—	0.40	1.81	0.23	1.2521	1.2521
0.54	-1.51	0.21	1.3844	1.3844	—	0.36	-0.95	0.89	1.4850	1.4850	—	0.49	1.59	0.12	1.5685	1.5685
								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
-0.005	0.82	0.03	1.0716	1.0678	—	0.20	0.96	0.11	1.0663	1.0645	—	0.34	2.27	0.1	1.0056	1.0047
0.20	-1.11	0.03	1.1129	1.0961	—	0.26	0.48	0.15	1.1051	1.0999	—	0.39	1.93	0.1	1.1308	1.1296
0.28	1.17	0.89	1.3022	1.3015	—	0.28	0.84	0.87	1.3210	1.3185	—	0.40	1.81	0.23	1.2528	1.2488
0.54	-1.51	0.21	1.4223	1.6459	—	0.36	-0.95	0.89	1.5207	1.6989	—	0.49	1.59	0.12	1.5686	2.8876
								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
-0.005	0.82	0.03	1.1541	1.1541	—	0.20	0.96	0.11	1.0846	1.0846	—	0.34	2.27	0.1	1.0010	1.0010
0.20	-1.11	0.03	1.2512	1.2512	—	0.26	0.48	0.15	1.1572	1.1572	—	0.39	1.93	0.1	1.1240	1.1240
0.28	1.17	0.89	1.5395	1.5395	—	0.28	0.84	0.87	1.4705	1.4705	—	0.44	2.60	0.1	1.2397	1.2397
0.54	-1.51	0.21	1.8127	1.8127	—	0.36	-0.95	0.89	1.6772	1.6772	—	0.49	1.59	0.12	1.5202	1.5202

Table 4.3: Simulation results of the MSFEs of each method applied on fractionally integrated time series for the time period $T = 300$ and break date $b = 0.1$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$ and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{0.75}$ in a single, discrete break in a simple regression model.

λ		0.5	—				1			—				2		
T = 250																
b = 0.1																
Estimated Equal 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
0.19	2.20	0.02	1.2761	1.2761	—	0.18	0.72	0.18	1.1282	1.1282	—	0.38	1.67	0.1	1.2459	1.2459
0.26	0.46	0.16	1.5771	1.5771	—	0.27	0.75	0.42	1.3069	1.3069	—	0.41	2.10	0.1	1.2617	1.2617
0.31	-0.58	0.74	1.7819	1.7819	—	0.33	-0.93	0.81	1.6537	1.6537	—	0.47	1.28	0.65	1.4395	1.4395
0.51	2.12	0.06	2.3493	2.3493	—	0.32	-0.56	0.44	2.4064	2.4064	—	0.52	1.78	0.20	2.3062	2.3062
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
0.19	2.20	0.02	1.0697	1.0686	—	0.18	0.72	0.18	1.0370	1.0366	—	0.38	1.67	0.1	1.1039	1.1016
0.26	0.46	0.16	1.1242	1.1198	—	0.27	0.75	0.42	1.1310	1.1278	—	0.41	2.10	0.1	1.1391	1.1378
0.31	-0.58	0.74	1.2114	1.2085	—	0.33	-0.93	0.81	1.2969	1.2942	—	0.47	1.28	0.65	1.1181	1.1174
0.51	2.12	0.06	1.4205	1.4188	—	0.32	-0.56	0.44	1.5360	1.4932	—	0.52	1.78	0.20	1.5655	1.5647
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
0.19	2.20	0.02	1.1117	1.1117	—	0.18	0.72	0.18	1.0432	1.0432	—	0.38	1.67	0.1	1.1037	1.1037
0.26	0.46	0.16	1.1762	1.1762	—	0.27	0.75	0.42	1.1512	1.1512	—	0.41	2.10	0.1	1.1386	1.1386
0.31	-0.58	0.74	1.2344	1.2344	—	0.33	-0.93	0.81	1.3141	1.3141	—	0.47	1.28	0.65	1.1193	1.1193
0.51	2.12	0.06	1.4323	1.4323	—	0.32	-0.56	0.44	1.5495	1.5495	—	0.52	1.78	0.20	1.5684	1.5684
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
0.19	2.20	0.02	1.1012	1.0917	—	0.18	0.72	0.18	1.0318	1.0287	—	0.38	1.67	0.1	1.1042	1.1028
0.26	0.46	0.16	1.1791	1.1630	—	0.27	0.75	0.42	1.1580	1.1386	—	0.41	2.10	0.1	1.1394	1.1386
0.31	-0.58	0.74	1.2147	1.2123	—	0.33	-0.93	0.81	1.3057	1.2890	—	0.47	1.28	0.65	1.1290	1.1276
0.51	2.12	0.06	1.4593	1.4582	—	0.32	-0.56	0.44	1.5361	1.5356	—	0.52	1.78	0.20	1.5696	1.5685
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
0.19	2.20	0.02	1.2014	1.2014	—	0.18	0.72	0.18	1.0641	1.0641	—	0.38	1.67	0.1	1.1008	1.1008
0.26	0.46	0.16	1.3876	1.3876	—	0.27	0.75	0.42	1.1833	1.1833	—	0.41	2.10	0.1	1.1276	1.1276
0.31	-0.58	0.74	1.4570	1.4570	—	0.33	-0.93	0.81	1.3763	1.3763	—	0.47	1.28	0.65	1.1025	1.1025
0.51	2.12	0.06	1.7313	1.7313	—	0.32	-0.56	0.44	1.6892	1.6892	—	0.52	1.78	0.20	1.6088	1.6088

Table 4.4: Simulation results of the MSFEs of each method applied on fractionally integrated time series for the time period $T = 250$ and break date $b = 0.1$ with different break date estimates $\hat{b}$, break size estimates $\hat{\lambda}$ and modified LW estimates $\hat{d}$ based on bandwidth $m = T^{0.75}$ in a single, discrete break in a simple regression model.

λ		0.5	—				1			—				2		
T = 1000																
b = 0.2																
Estimated Equal 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
0.12	0.44	0.19	1.0135	1.0135	—	0.23	0.99	0.20	1.1300	1.1300	—	0.36	2.12	0.20	1.7511	1.7511
0.22	0.56	0.10	1.2620	1.2620	—	0.31	1.15	0.21	1.3152	1.3152	—	0.36	1.75	0.20	1.7334	1.7334
0.28	1.11	0.05	1.5825	1.5825	—	0.30	0.66	0.23	1.5740	1.5740	—	0.46	2.65	0.20	1.9435	1.9435
0.46	1.70	0.02	2.5184	2.5184	—	0.41	1.44	0.19	2.3341	2.3341	—	0.50	1.86	0.18	2.4768	2.4768
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
0.12	0.44	0.19	0.9323	0.9316	—	0.23	0.99	0.20	0.9271	0.9263	—	0.36	2.12	0.20	1.0108	1.0093
0.22	0.56	0.10	1.1027	1.1019	—	0.31	1.15	0.21	1.1105	1.1093	—	0.36	1.75	0.20	1.0690	1.0682
0.28	1.11	0.05	1.1562	1.1554	—	0.30	0.66	0.23	1.3129	1.3117	—	0.46	2.65	0.20	1.3752	1.3747
0.46	1.70	0.02	1.5483	1.5477	—	0.41	1.44	0.19	1.5886	1.5871	—	0.50	1.86	0.18	1.5590	1.5583
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
0.12	0.44	0.19	0.9316	0.9316	—	0.23	0.99	0.20	0.9268	0.9268	—	0.36	2.12	0.20	1.0109	1.0109
0.22	0.56	0.10	1.1046	1.1046	—	0.31	1.15	0.21	1.1104	1.1104	—	0.36	1.75	0.20	1.0692	1.0692
0.28	1.11	0.05	1.1594	1.1594	—	0.30	0.66	0.23	1.3133	1.3133	—	0.46	2.65	0.20	1.3751	1.3751
0.46	1.70	0.02	1.5523	1.5523	—	0.41	1.44	0.19	1.5908	1.5908	—	0.50	1.86	0.18	1.5589	1.5589
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
0.12	0.44	0.19	0.9354	0.9348	—	0.23	0.99	0.20	0.9270	0.9264	—	0.36	2.12	0.20	1.0108	1.0094
0.22	0.56	0.10	1.1188	1.1186	—	0.31	1.15	0.21	1.1105	1.1094	—	0.36	1.75	0.20	1.0691	1.0683
0.28	1.11	0.05	1.1770	1.1716	—	0.30	0.66	0.23	1.3194	1.3169	—	0.46	2.65	0.20	1.3753	1.3747
0.46	1.70	0.02	1.5857	1.5566	—	0.41	1.44	0.19	1.6030	1.5854	—	0.50	1.86	0.18	1.5592	1.5587
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
0.12	0.44	0.19	0.9418	0.9418	—	0.23	0.99	0.20	0.9316	0.9316	—	0.36	2.12	0.20	1.0200	1.0200
0.22	0.56	0.10	1.1623	1.1623	—	0.31	1.15	0.21	1.1090	1.1090	—	0.36	1.75	0.20	1.0730	1.0730
0.28	1.11	0.05	1.3036	1.3036	—	0.30	0.66	0.23	1.3171	1.3171	—	0.46	2.65	0.20	1.3312	1.3312
0.46	1.70	0.02	1.8949	1.8949	—	0.41	1.44	0.19	1.7572	1.7572	—	0.50	1.86	0.18	1.5316	1.5316

Table 4.5: Simulation results of the MSFEs of each method applied on fractionally integrated time series for the time period $T = 1000$ 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 $m = T^{0.75}$ in a single, discrete break in a simple regression model.

This table presents the empirical application to inflation rates forecasting in Germany and Australia. We observe that the estimated optimal window and estimated AveW methods provide the best forecasts of the inflation rates of Germany and Australia, respectively. In contrast, the estimated post-break window performs poorly, displaying the highest MSFEs among all methods in both 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	0.2504	−0.2222	0.3033	0.1047	0.1028	0.1168	0.1168	0.1077	0.0694	0.1111	0.1111
AUT	0.1092	−0.2711	0.3377	0.0374	0.0373	0.0441	0.0441	0.0391	0.0361	0.0331	0.0331

Table 4.6: MSFE results 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 $m = T^{3/4}$ in a simple regression model.

Appendix C - Figures

This figure presents the series of inflation rates for Germany and Australia, respectively, where the red vertical lines represent their corresponding estimated break points, $\hat{T_b^G} = 185$ and $\hat{T_b^A} = 206$.

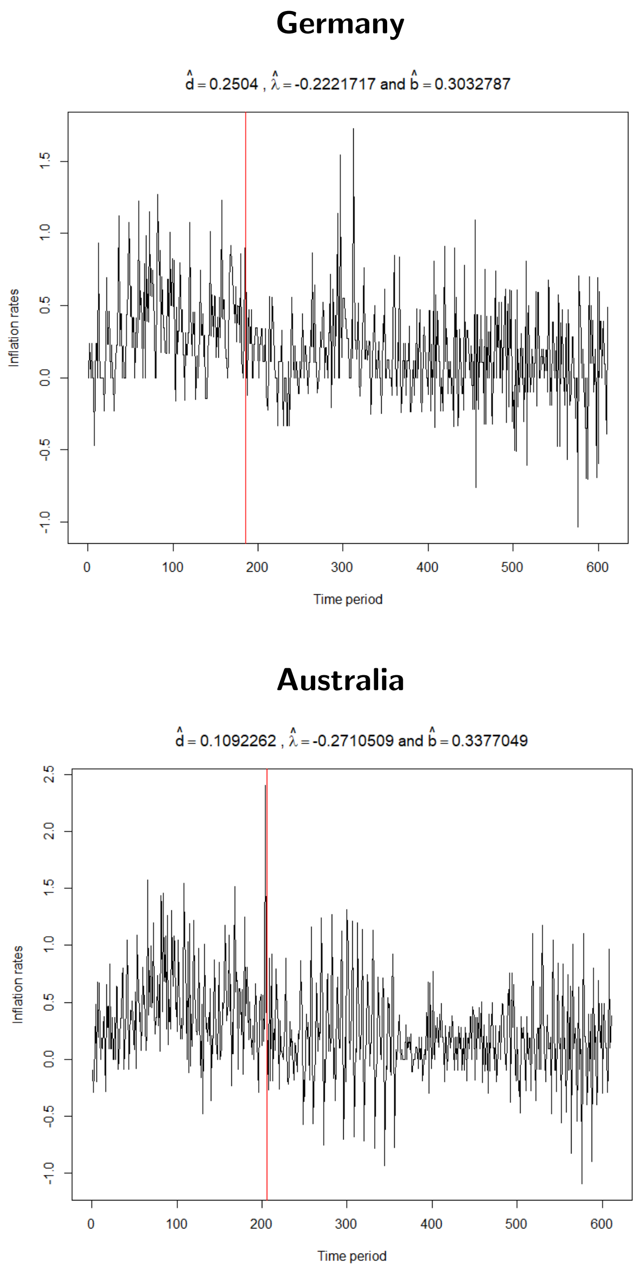

Figure 4.1: Inflation Rates for Germany and Australia with their respective memory estimate $\hat{d}$, break size estimate $\hat{\lambda}$ and break date estimate $\hat{b}$. The red vertical line indicate the break point estimate.

Bibliography

Aue, A., Hörmann, S., Horváth, L., & Reimherr, M. (2009). Break Detection in the Covariance Structure of Multivariate Time Series Models. *The Annals of Statistics*, *37*(6B), 4046–4087.

Aue, A., & Horváth, L. (2013). Structural Breaks in Time Series. *Journal of Time Series Analysis*, *34*(1), 1–16.

Aue, A., Horváth, L., Hušková, M., & Ling, S. (2009). On Distinguishing Between Random Walk and Change in the Mean Alternatives. *Econometric Theory*, *25*(2), 411–441. https://doi.org/10.1017/S0266466608090130

Bai, J., & Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, *66*(1), 47–78. http://www.jstor.org/stable/2998540

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

Bos, C. S., Franses, P. H., & 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. https://doi.org/10.48550/ARXIV.1803.10883

Casini, A. (2021). Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models. https://doi.org/10.48550/ARXIV.2103.02981

Casini, A., Deng, T., & Perron, P. (2021). Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference. https://doi.org/10.48550/ARXIV.2103.01604

Chen, W. W., & Hurvich, C. M. (2006). Semiparametric Estimation of Fractional Cointegrating Subspaces. *The Annals of Statistics*, *34*(6), 2939–2979.

Clements, A., Hurn, A., & Volkov, V. (2016). Common Trends in Global Volatility. *Journal of International Money and Finance*, *67*, 194–214.

Clements, M. P., & Hendry, D. F. (2000). *Forecasting Non-Stationary Economic Time Series*. mit Press.

Diebold, F. X., & Inoue, A. (2001). Long Memory and Regime Switching. *Journal of Econometrics*, *105*(1), 131–159. https://doi.org/10.3386/t0264

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 251–276. http://www.jstor.org/stable/1913236

Fleming, J., & Kirby, C. (2011). Long Memory in Volatility and Trading Volume. *Journal of Banking & Finance*, *35*(7), 1714–1726. https://doi.org/10.1016/j.jbankfin.2010.11.007

Gadea, M. D., Sabaté, M., & 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. https://doi.org/10.1016/j.intfin.2003.05.001

Giacomini, R., & Rossi, B. (2009). Detecting and Predicting Forecast Breakdowns. *The Review of Economic Studies*, *76*(2), 669–705.

Granger, C. W. (1980). Long Memory Relationships and the Aggregation of Dynamic Models. *Journal of Econometrics*, *14*(2), 227–238.

Granger, C. W., & Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, *73*(1), 61–77. https://doi.org/10.1016/0304-4076(95)01733-x

Granger, C. W., & 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. https://doi.org/10.1016/j.jempfin.2003.03.001

Granger, C. W., & Joyeux, R. (1980). An Introduction to Long-Memory Time Series Mdels and Fractional Differencing. *Journal of Time Series Analysis*, *1*(1), 15–29. https://doi.org/10.1111/j.1467-9892.1980.tb00297.x

Hassler, U., & Wolters, J. (1995). Long Memory in Inflation Rates: International Evidence. *Journal of Business & Economic Statistics*, *13*(1), 37–45.

Hou, J., & 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. https://doi.org/10.1016/j.jeconom.2014.05.004

Hsu, C.-C. (2005). Long Memory or Structural Changes: An Empirical Examination on Inflation Rates. *Economics Letters*, *88*(2), 289–294. https://doi.org/10.1016/j.econlet.2005.02.017

Hyung, N., Franses, P. H., & 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. https://EconPapers.repec.org/RePEc:eee:riibaf:v:20:y:2006:i:1:p:95-110

Inoue, A., & Rossi, B. (2011). Identifying the Sources of Instabilities in Macroeconomic Fluctuations. *Review of Economics and Statistics*, *93*(4), 1186–1204.

Koop, G., & Potter, S. M. (2007). Estimation and Forecasting in Models with Multiple Breaks. *The Review of Economic Studies*, *74*(3), 763–789.

Lavielle, M., & Moulines, E. (2000). Least-Squares Estimation of An Unknown Number of Shifts in a Time Series. *Journal of Time Series Analysis*, *21*(1), 33–59. https://doi.org/10.1111/1467-9892.00172

Lu, Y. K., & Perron, P. (2010). Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model. *Journal of Empirical Finance*, *17*(1), 138–156. https://doi.org/10.1016/j.jempfin.2009.10.001

Maheu, J. M., & Gordon, S. (2008). Learning, Forecasting and Structural Breaks. *Journal of Applied Econometrics*, *23*(5), 553–583.

Maheu, J. M., & 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.

Martins, L. F., & Rodrigues, P. M. (2014). Testing for Persistence Change in Fractionally Integrated Models: An Application to World Inflation Rates. *Computational Statistics & Data Analysis*, *76*, 502–522.

Mikosch, T., & 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. https://doi.org/10.1162/003465304323023886

Nguyen, D. B. B., Prokopczuk, M., & Sibbertsen, P. (2020). The Memory of Stock Return Volatility: Asset Pricing Implications. *Journal of Financial Markets*, *47*, 100487.

Paye, B. S., & Timmermann, A. (2006). Instability of Return Prediction Models. *Journal of Empirical Finance*, *13*(3), 274–315.

Perron, P., & Yamamoto, Y. (2021). Testing for Changes in Forecast Performance. *Journal of Business and Economic Statistics*, *39*(1), 148–165.

Pesaran, H., & Pick, A. (2011). Forecast Combination Across Estimation Windows. *Journal of Business & Economic Statistics*, *29*(2), 307–318. https://doi.org/10.2307/25800802

Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. *The Review of Economic Studies*, *73*(4), 1057–1084.

Pesaran, M. H., Pick, A., & Pranovich, M. (2013). Optimal Forecasts in the Presence of Structural Breaks. *Journal of Econometrics*, *177*(2), 134–152.

Pesaran, M. H., & 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. https://doi.org/10.2139/ssrn.1238142

Robinson, P. M., & Yajima, Y. (2002). Determination of Cointegrating Rank in Fractional Systems. *Journal of Econometrics*, *106*(2), 217–241.

Rossi, B. (2013). Advances in Forecasting Under Instability. *Handbook of Economic Forecasting* (pp. 1203–1324). Elsevier.

Shao, X., & Zhang, X. (2010). Testing for Change Points in Time Series. *Journal of the American Statistical Association*, *105*(491), 1228–1240.

Shimotsu, K. (2010). Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend. *Econometric Theory*, *26*(2), 501–540.

Sibbertsen, P., & Kruse, R. (2009). Forecasting Long Memory Time Series Under a Break in Persistence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1508404

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

Souza, I. V. M., Reisen, V. A., Franco, G. d. C., & Bondon, P. (2018). The Estimation and Testing of the Cointegration Order Based on the Frequency Domain. *Journal of Business & Economic Statistics*, *36*(4), 695–704.

Stock, J. H., & Watson, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics*, *14*(1), 11–30.

Tsay, W.-J. (2000). Long Memory Story of the Real Interest Rate. *Economics Letters*, *67*(3), 325–330. https://doi.org/10.1016/S0165-1765(99)00272-4

Wang, B., Wang, M., & Chan, N. H. (2015). Residual-Based Test for Fractional Cointegration. *Economics Letters*, *126*, 43–46.

Wenger, K., Leschinski, C., & Sibbertsen, P. (2018). The Memory of Volatility. *Quantitative Finance and Economics*, *2*(1), 137–159. https://doi.org/10.3934/QFE.2018.1.137