School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Adaptive Forecasting in the presence of recent and ongoing structural change

Liudas Giraitis
Queen Mary, University of London

George Kapetanios
Bank of England and Queen Mary, University of London

Simon Price[1]
Bank of England and City University, London

Department of Economics
Discussion Paper Series
No. 12/02

[1] Department of Economics, City University London, Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK.
Email: simon.price@bankofengland.co.uk.

Adaptive forecasting in the presence of recent and ongoing structural change

Liudas Giraitis
Queen Mary, University of London

George Kapetanios*
Bank of England and Queen Mary, University of London

Simon Price
Bank of England and City University, London.

February 29, 2012

Abstract

We consider time series forecasting in the presence of ongoing structural change where both the time series dependence and the nature of the structural change are unknown. Methods that downweight older data, such as rolling regressions, forecast averaging over different windows and exponentially weighted moving averages, known to be robust to historical structural change, are found to be also useful in the presence of ongoing structural change in the forecast period. A crucial issue is how to select the degree of downweighting, usually defined by an arbitrary tuning parameter. We make this choice data dependent by minimizing forecast mean square error, and provide a detailed theoretical analysis of our proposal. Monte Carlo results illustrate the methods. We examine their performance on 191 UK and US macro series. Forecasts using data-based tuning of the data discount rate are shown to perform well.

Key words: recent and ongoing structural change, forecast combination, robust forecasts.
JEL Classifications: C100, C590.

*The views expressed are those of the authors, and not necessarily those of the Bank of England or Monetary Policy Committee.

1 Introduction

It is widely accepted that structural change is a crucial issue in econometrics and forecasting. Clements and Hendry suggest forcefully (in e.g. 1998a,b) that such change is the main source of forecast error; Hendry (2000) argues that the dominant cause of forecast failures is the presence of deterministic shifts. Convincing evidence of presence of structural change was offered by Stock and Watson (1996) who looked at many forecasting models of a large number of US time series, and found evidence for parameter instability in a substantial proportion. This remains relevant: the literature on forecasting in the presence of instabilities was surveyed in Rossi (2012) for the Handbook of Forecasting. In her conclusions, Rossi (2012) writes 'the widespread presence of forecast breakdowns suggests the need of improving ways to select good forecasting models in-sample.' Our work is a contribution to precisely this, taking a novel approach that is both robust and data driven. In general, model parameters may change continuously, drift smoothly over time or change at at discrete points, in an unknown manner, and both within the sample and over the forecast period. We therefore consider a very general setting accommodating an unknown model structure and structural change.

There is a large literature on the identification of breaks, and forecasting methods robust to them, and Rossi (2012) surveys the relevant literature. However, the deeply practical need to forecast after a recent structural change, or during a period of such change, has received very little attention. It is further compounded by the unknown and therefore unspecified nature of any structural change, since most forecast approaches are only effective in specific cases. Detection has a long history, mainly in the context of structural breaks (although see Kapetanios (2007) for the case of smooth structural change). The seminal paper on structural change where the break point is known was Chow (1960). Andrews (1993) introduced a testing methodology that allowed for unknown break-points, while another influential contribution on multiple structural changes in linear regression is Bai and Perron (1998). The question of amendment of forecasting strategies then arises. While this problem has been tackled by many authors, a major contribution has been made by Pesaran and Timmermann (2007), who combine a number of alternative forecasting strategies in the presence of breaks. They conclude that forecast pooling using a variety of estimation windows provides a reasonably good and robust forecasting performance.

Most of the existing work on forecasting assumes that change has occurred when sufficient time has elapsed for post-break estimation. However, the issue of change occurring in real time is a clear problem, which is addressed in Eklund, Kapetanios, and Price (2010). They consider a variety of forecasting strategies which can be grouped in two distinct groups. One group monitors for change and adjusts forecasts once change has been detected. The other group does not attempt to identify breaks, since that requires substantial time lags. Instead it uses break-robust forecasting strategies that essentially downweight data from older periods that are deemed to be irrelevant for the current conjuncture.

While moving in an interesting direction, Eklund, Kapetanios, and Price (2010) do not elaborate two closely related issues: how much to downweight, and whether older data should be downweighted monotonically. Clearly, any arbitrary discount factor is unlikely to be optimal. And neither may monotonicity: for example, if regimes (e.g., monetary policy) come and go then older data, from a period where the current regime held, would be more relevant than more recent data from other regimes.[1]

This paper suggest approaches that address these issues. Our main contribution is to introduce and analyse a cross-validation based method which selects a tuning parameter defining the downweighting rate of the older data. We show that the implied discount rate minimizes the MSE of the forecast in the weighting schemes we consider. Further, we consider a non-parametric approach for determining a flexible weighing scheme for past data, to be used in constructing forecasts. The latter does not assume any particular shape for the weight function, such as monotonic decline. We explore the properties of the new methods for variety of models in terms of theory, with a Monte Carlo exercise and empirically.

An interesting byproduct of our results is a novel and simple way to accommodate trends of a generic nature in forecasting. Unlike many forecasting approaches that require the removal of stochastic or other trends before forecasting, our methods are designed for, and work best, in relative terms compared to existing methods, when applied to the level of the forecast series.

The rest of the paper is organized as follows. Section 2 presents a new approach for forecasting in the presence of recent structural breaks. We provide its theoretical justification and asymptotic MSE, and describe some robust forecasting strategies. Section 3 includes an extensive Monte Carlo study in which all these strategies are evaluated. In Section 4 the forecast methods are used to examine a large number of US and UK macroeconomic time series, where we find results broadly consistent with the Monte Carlo study. Section 5 concludes. Proofs are reported in Appendix A.

[1] This might suggest a need to estimate regimes, but as our focus is on time series forecasting methods and not inference, one is free to be agnostic about the presence of particular regimes.

2 Adaptive forecasting: econometric framework

2.1 Forecasting strategies

Our framework may be summarised by the general model

$$y_{T,t} = \beta_{T,t} + u_t, \qquad t = 1, \cdots, T, \tag{2.1}$$

where $\beta_{T,t}$ is an unobserved process, and u_t is a stationary dependent noise process that is independent of $\beta_{T,t}$. Unlike most previous work we wish to place as little structure as possible on the process $\beta_{T,t}$. We do not specify whether $\beta_{T,t}$ is a stochastic or deterministic process, or whether it is discontinuous or smooth. We will consider two distinct but related approaches to the question of forecasting under such a model.

Eklund, Kapetanios, and Price (2010) find that simple robust forecasting, based on weighting schemes that discount past data, works very well in practice. Examples include exponentially weighted moving average forecasting, or forecast combinations based on different estimation windows. By varying a tuning parameter of parametrically defined weights, such methods simply impose different shapes on the weight functions that downweight past data. One crucial problem with most such methods is that the tuning parameters driving weight functions, are usually *a priori* pre-selected. Hence, they can be suboptimal and a data dependent method for choosing these parameters is of great interest.

One way to calibrate parameters of forecasting strategies is by optimizing on in-sample forecasting performance. Although this idea is not common in the literature, it is not new. For example, Kapetanios, Labhard, and Price (2006) suggest forecasts where models are averaged with weights that depend on the forecasting performance of each model in the recent past. That approach was found to work very well and to be preferable to alternative ways of determining the weights. In what follows we formalize the above ideas, presenting a data driven weighting strategy and its theoretical analysis. It is instructive to focus first on a simple location model such as (2.1). Extension of the results to models with regressors is briefly discussed.

The error process $\{u_t\}$ is stationary linear process with mean zero and finite variance σ_u^2 and independent of $\{\beta_{T,t}\}$. The persistent component $\beta_{T,t}$ is allowed to be a triangular array, and can be a deterministic trend, a unit root type process or a combination of both. This set-up provides sufficient flexibility to our theoretical analysis since it allows for $\beta_{T,t}$'s such as those used in locally stationary models (e.g. Dahlhaus (1996)), or persistent stochastic trend models. For simplicity of notation, we write $y_{T,t}$ as y_t and $\beta_{T,t}$ as β_t.

We consider the properties of a linear forecast of y_t, based on past values $y_{t-1}, \cdots, y_t$:

$$\hat{y}_{t|t-1} \equiv \hat{y}_{t|t-1}(H) = \sum_{j=1}^{t-1} w_{tj;H} y_{t-j} = w_{t1;H} y_{t-1} + \cdots + w_{t,t-1;H} y_1. \tag{2.2}$$

Such an averaging-type forecast involves standardized non-negative weights $w_{tj;H}$, $j = 1, \cdots, t-1$, $w_{t1;H} + \cdots + w_{t,t-1;H} = 1$, parameterized by a single tuning parameter H, controlling the rate at which past observations are downweighted (e.g., the width of the rolling window). We assume that $H \in I_T = [\alpha, T^{1-\delta}]$ where $\alpha > 0$ and $\delta > 0$ is small.

The class of weights $w_{tj;H}$ which we consider are described in the following assumption. $\dot{f}$ and $\ddot{f}$ denotes the first and second derivatives of a function f, $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$ and $I(A)$ is the indicator function. $a_T \sim b_T$ indicates that $a_T/b_T \to 1$, as T increases.

Assumption 1 *For* $t = 1, \cdots, T$, $T \geq 1$,

$$w_{tj,H} = \frac{K(\frac{t-j}{H})}{\sum_{k=1}^{t} K(\frac{k}{H})}, \qquad j = 1, \cdots, t, \qquad H \in I_T. \tag{2.3}$$

The function $K(x) \geq 0$, $x \geq 0$ *is a continuous function twice differentiable on its support and such that* $\int_0^\infty K(u)du = 1$,

$$\inf_{0 \leq x \leq 1/\alpha} K(x) > 0, \quad \max\{K(x), |\dot{K}(x)|, |\ddot{K}(x)|\} \leq \frac{C}{1+x^8}. \tag{2.4}$$

The main classes of weights that are commonly employed satisfy this Assumption. For example,

(i) *Rolling window*

$$K(u) = I(0 \leq u \leq 1). \tag{2.5}$$

(ii) *Exponential weighted moving average* (EWMA) weights

$$K(u) = e^{-u}, \qquad u \in [0, \infty). \tag{2.6}$$

Let $\rho = \exp(-1/H)$. Then $K(j/H) = \rho^j$ and $w_{tj;H} = \rho^{t-j}/\sum_{k=1}^{t-1} \rho^k$, $1 \leq j \leq t-1$.

(iii) *Triangular window*

$$K(u) = 2(1-u)I(0 \leq u \leq 1). \tag{2.7}$$

The forecasts $\hat{y}_{t|t-1}$ in (2.2) are based on (local) averaging. While the rolling window simply averages the H previous observations, the EWMA forecast uses all observations $y_1, \cdots, y_{t-1}$, increasinly downweighting more distant observations.

In practice, forecasting of y_t with a persistent unobserved deterministic or stochastic component β_t, e.g. a unit root or a linear trend, is often conducted by averaging over the last few observations. When persistence in β_t falls, wider windows may be expected to yield smaller forecast MSE. It is plausible that for a stationary process $\{y_t\}$, the sample mean average forecast $(y_t + \cdots + y_1)/t$ will be outperformed by a forecast discounting past data when dependence in $\{y_t\}$ becomes sufficiently strong. This hypothesis is supported by the theory presented below. The implication is that selection of H depends on the unknown and unspecified level of persistence in β_t and u_t, in contradiction to the usual practice of specifying an arbitrary value.

2.2 Selection of weights

Given a sample $y_1, \cdots, y_T$, to compute the forecast $y_{T+1|T}$ we need to select the parameter H. We use cross-validation, obtaining H by numerically minimizing the mean squared forecast error of the forecasts produced at desired horizons. The objective function associated with the above minimization problem is given by

$$Q_T(H) := \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{t|t-1} - y_t)^2, \qquad \hat{H} := \mathrm{argmin}_{H \in I_T} Q_T(H). \tag{2.8}$$

Subsequently we will show $\hat{H}$ defines 'optimal' weights for the forecast $\hat{y}_{T+1|T}(H)$ of y_{T+1}, minimizing the mean squared error (MSE), $e_T(H) := E(\hat{y}_{T+1|T}(H) - y_{T+1})^2$ in H, hence making the forecast procedure (2.2) operational and optimal. In addition, quantity $Q_T(\hat{H})$ will evaluate the forecast error as follows:

$$\begin{aligned} &\inf_{H \in I_T} e_T(H) \sim e_T(\hat{H}) \sim Q_T(\hat{H}), \qquad \text{or} \\ &\inf_{H \in I_T} e_T(H) - \sigma_u^2 \sim e_T(\hat{H}) - \sigma_u^2 \sim Q_T(\hat{H}) - \sigma_u^2, \end{aligned} \tag{2.9}$$

bearing in mind that in a number of settings discussed below, $e_T(\hat{H}) \to \sigma_u^2$.

Justification of minimization procedure (2.8) will require some restrictions on β_t and u_t, and some technical effort. To give a hint of how the data based selection of the tuning parameter H works, denote by $\hat{\sigma}_{T,u}^2 := T^{-1} \sum_{j=1}^{T} u_j^2$ the sample variance of error process $\{u_t\}$.

In the main set-ups of β_t, considered below, $Q_T(H)$ has the following property: as $T \to \infty$,

$$Q_T(H) = \hat{\sigma}^2_{T,u} + E[Q_T(H) - \hat{\sigma}^2_{T,u}](1 + o_P(1)),$$
$$= \hat{\sigma}^2_{T,u} + (A\frac{H^\gamma}{T^\delta} + \frac{B}{H})(1 + o_P(1)), \quad H \to \infty, \tag{2.10}$$

with some constants $A \geq 0$, $|B| < \infty$, and $\gamma, \delta \in \{0, 1, 2\}$. The term AH^γ/T^δ comes from β_t while B/H is contributed by u_t.

For a linear or stochastic unit root trend β_t, $\delta = 0$. Then, $\lim_T E[Q_T(H) - \hat{\sigma}^2_{T,u}] = Q(H)$ achieves its minimum on a bounded interval, and thus, $\hat{H}$ remains bounded. In particular, $\hat{H} \to \text{argmin}_{H \in I_T} Q(H)$. For a bounded smooth deterministic or bounded stochastic (unit root) trend β_t, minimization may reduce to $\hat{H} \sim \text{armin}_H\{A(H/T))^\delta + (B/H)\}$, $\delta > 0$, $B > 0$ which leads to $\hat{H} \sim cT^{\delta/(1+\delta)}$ increasing with T. For a sufficiently persistent stationary process $y_t = u_t$, $\hat{H}$ may stay bounded and minimize the limit $Q(H) := \lim_T E[Q_T(H) - \hat{\sigma}^2_{T,u}]$, whereas for i.i.d. or weakly dependent process, H will tend to take large values. Here (2.10) holds with $A = 0$. The results for a break in the mean of a stationary process occurring at the time $L < T$ show that the rolling window forecast will be built using the data from the post-brake period $[L, T]$ after time $\tau := T - L > \sqrt{T}$.

2.3 Properties of $\hat{H}$

Now we turn to the theoretical justification of the selection procedure of H and investigate the properties of $\hat{H}$.

The next assumption specifies restrictions on a stationary process $\{u_t\}$. Denote by $\gamma_u(k) = \text{Cov}(u_k, u_0)$, $k \geq 0$ the autocovariance function, and by $s^2_u := \sum_{k \in \mathbb{Z}} \gamma_u(k) > 0$ the long-run variance of $\{u_t\}$. Under Assumption 2 on $\{u_t\}$ below, $\sum_{k \in \mathbb{Z}} |\gamma_u(k)| < \infty$.

Assumption 2 $\{u_t\}$ *is a stationary short memory linear process*

$$u_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \qquad \{\varepsilon_j\} \sim IID(0, \sigma^2_\varepsilon) \tag{2.11}$$

such that $|a_j| \leq Cj^{-1-v}$ *for some* $v > 0$ *and* $s^2_u > 0$, *with* ε_j *having all moments finite.*

We will write $\{u_t\} \in I(0)$ to denote that $\{u_t\}$ satisfies Assumption 2. We write $\{\beta_t\} \in I(1)$, if $\{\nabla \beta_t := \beta_t - \beta_{t-1}\}$ satisfies Assumption 2 with innovations having four moments finite. The class $I(1)$ contains unit root processes and will be used to model $\{\beta_t\}$. We denote by $\mathcal{G}$ the class of continuous functions $g(x), x \in (0, 1)$ with a bounded second derivative.

We will consider the following settings for $\beta_1, \cdots, \beta_T$:

b1.	Stationary process	$\beta_t = \mu, \qquad y_t = \mu + u_t.$
b2.	Unit root	$\{\beta_t\} \in I(1).$
b3.	Deterministic trend	$\beta_t = tg(t/T)$, where $g \in \mathcal{G}$.
b4.	Bounded unit root	$\beta_t = T^{-1/2}\tilde{\beta}_t$, $\{\tilde{\beta}_t\} \in I(1).$
b5.	Bounded deterministic trend	$\beta_t = g(t/T)$, where $g \in \mathcal{G}$.
b6.	Break in the mean	$\beta_t = \begin{cases} \mu_1, & t = 1, \cdots, L, \\ \mu_2, & t = L+1, \cdots, T, \quad (T/2 \le L < T). \end{cases}$ We assume that $\Delta := \mu_1 - \mu_2 \neq 0$.

We will use the following notation:

$\nu_{1,K} = (\int_0^\infty K(x)xdx)^2, \quad \nu_{2,K} = \int_0^\infty K(x)x^2dx, \quad \nu_{3,K} = \int\int_0^\infty K(x)K(y)(x \wedge y)dxdy.$

$\kappa(g) = \int_0^1 \dot{g}^2(x)dx,\ g \in \mathcal{G}; \quad Q_{u,T}(H) := T^{-1}\sum_{t=1}^T |\sum_{j=1}^{t-1}(u_j - u_t)|^2.$

We now are ready to analyze the properties of $Q_T(H)$, $\hat{H}$ and the MSE of the forecast error $e_T(H) = E(y_{T+1|T}(\hat{H}) - y_{T+1})^2$.

2.4 Stationary case

First we consider $\{y_t\}$ as a stationary process. Denote $K_2 = \int_0^\infty K^2(x)dx$, $K_0 = K(0)$,

$$Q_u(H) := \sum_{j,k=1}^\infty v_{j;H}v_{k;H}\gamma_u(j-k) - 2\sum_{j=1}^\infty v_{j;H}\gamma_u(j), \quad b_{u,K} := s_u^2\{K_2 - K_0\} + \sigma_u^2 K_0,$$

where $v_{j;H} = k_{j;H}/\sum_{k=1}^\infty k_{j;H}$, $j \ge 1$ with $k_{j;H} := K(j/H)$.

Theorem 1 *Let $y_t = \mu + u_t, t = 1, \cdots, T$, where $\{u_t\} \in I(0)$. Then, as $T \to \infty$, uniformly in $H \in I_T$,*

$$Q_T(H) = \hat{\sigma}_{u,T}^2 + E[Q_{u,T}(H) - \hat{\sigma}_{u,T}^2](1 + o_p(1)), \tag{2.12}$$

$$E[Q_{u,T}(H) - \hat{\sigma}_{u,T}^2] = Q_u(H)(1 + o(1)), \tag{2.13}$$

$$Q_u = H^{-1}b_{u,K} + o(H^{-1}), \qquad H \to \infty. \tag{2.14}$$

Theorem 1 shows that minimization of $Q_T(H)$ reduces to that of $Q_u(H) := \lim_T E[Q_T(H) - \hat{\sigma}_{u,T}^2]$, while by the corollary below, the forecast, based on $\hat{H}$ minimizes the MSE of the forecast with respect to H and allows its evaluation.

Corollary 1 *Under assumptions of Theorem 1,*

$$\inf_{H \in I_T} e_T(H) = \sigma_u^2 + \inf_{H \in I_T}\{Q_T(H) - \hat{\sigma}_{u,T}^2\}(1 + o(1)). \tag{2.15}$$

Moreover, if $Q_u(H)$ has a minimum at some $H_0 < T$, then

$$\inf_{H \in I_T} e_T(H) = e_T(\hat{H}) + o(1) = Q_T(\hat{H}) + o(1). \tag{2.16}$$

Otherwise, if function $Q_u(H) > 0$ is strictly positive, then $\hat{H} \sim T^{1-\delta}$ tends to take the largest possible value in I_T, and

$$\inf_{H \in I_T} e_T(H) = e_T(\hat{H})(1 + o_P(1)) = Q_T(\hat{H})(1 + o(1)). \tag{2.17}$$

Remark 1 Equation (2.15) indicates that minimization of $e_{H,T}$ in H is equivalent to that of $Q_T(H)$. Moreover, EWMA weights may lead to smaller MSE compared to a rolling window: for rolling windows, $b_{u,K} = \sigma_u^2 > 0$, while for EWMA weights $b_{u,K} = (2\sigma_u^2 - s_u^2)/2 \le \sigma_u^2$. In the latter case $b_{u,K} < 0$ if a stationary sequence $\{u_t\}$ has long-run variance such that $s_u^2 > 2\sigma_u^2$, e.g. for an AR(1) model with parameter $\phi \in (1/3, 1)$. Hence, the shape of weights and the strength of dependence in $\{u_t\}$ have a crucial impact on the the optimal forecast error and the rate of down-weighting the data. Moreover, for $b_{u,k} < 0$, $Q_u(H)$ achieves its minimum at some finite H_0. Thus, $\hat{H} \to H_0$ remains finite, and the forecast MSE, $e_T(\hat{H}) \to \sigma_u^2 + Q_u(H_0) < \sigma_u^2$, is smaller than that of the sample mean. However, if $b_{u,K} > 0$ and $Q_u(H) > 0$, $H \in I_T$, then by (2.13), $\hat{H}$ takes large values of order $T^{1-\delta}$ and the asymptotic forecast error $e_T(\hat{H}) \to \sigma_u^2$ is the same as for the sample mean forecast.

Unlike with EWMA weights, under rolling window weights $b_{u,T} > 0$ is always positive, and therefore it is hard to conclude if $Q_u(H)$ may be negative for any H. Monte Carlo simulation in Table 2 for u_t following an AR(1) model with parameter 0.7 indicates that the MSE for rolling window forecast is the same as that for the sample mean, whereas the EWMA weights reduce it by 33%.

2.5 Strong persistence

This corresponds to stochastic (b2) and deterministic (b3) trend $\{\beta_t\}$.
Set $\gamma_\beta(j,k) := E[(\beta_j - \beta_0)(\beta_k - \beta_0)]$, $j, k \ge 0$,
$V^2_{UR\beta,H} := \sum_{j,k=1}^{\infty} v_{j;H} v_{k;H}\, \gamma_\beta(j,k)$, $V_{tr,H} := \sum_{j=1}^{\infty} v_{j;H} j$.

Theorem 2 *Let $y_t = \beta_t + u_t, t = 1, \cdots, T$ with $\{\beta_t\}$ as in (b2) or (b3), $\{u_t\} \in I(0)$. Then, as $T \to \infty$, uniformly in $H \in I_T$,*

$$\begin{aligned} Q_T(H) &= \hat{\sigma}^2_{u,T} + \{E[Q_T(H) - \hat{\sigma}^2_{u,T}]\}(1 + o_p(1)) \qquad (2.18) \\ &= \hat{\sigma}^2_{u,T} + Q(H)(1 + o_p(1)), \end{aligned}$$

where

$$
\begin{aligned}
Q(H) &:= s^2_{\nabla\beta}V^2_{UR\beta,H} + Q_u(H) \quad \textit{in case (b2)}, \qquad (2.19)\\
&= \{Hs^2_{\nabla\beta}\nu_{3,K} + H^{-1}b_{u,K}\}(1+o(1)).\\
Q(H) &:= \kappa(\tilde{g})V^2_{tr,H} + Q_u(H) \quad \textit{in case (b3)}, \qquad (2.20)\\
&= \{H^2\kappa(\tilde{g})\nu_{1,K} + H^{-1}b_{u,K}\}(1+o(1)), \quad \tilde{g}(x) = xg(x), \quad x\in[0,1].
\end{aligned}
$$

Again, as in Theorem 1, minimization of $Q_T(H)$ reduces to that of $Q(H)$, and the forecast, based on $\hat{H}$ minimizes MSE, and the forecast MSE can be computed.

Corollary 2 *Under assumptions of Theorem 2, in cases* (b2) *and (b3),* $\hat{H} = O_p(1)$ *and*

$$\inf_{H\in I_T} e_T(H) = \sigma^2_u + \inf_{H\in I_T}\{Q_T(H) - \hat{\sigma}^2_{u,T}\}(1+o(1)). \qquad (2.21)$$

Moreover, if $Q(H)$ *has a unique minimum at some* H_0*, then* $\hat{H} \to_P H_0$ *and*

$$\inf_{H\in I_T} e_T(H) = e_T(\hat{H}) + o(1) = Q_T(\hat{H}) + o_p(1). \qquad (2.22)$$

Remark 2 Theorem 2 shows, that for strongly persistent β_t, $\hat{H} \to H_0 = argmin_{H\in I_T}Q(H)$, assuming H_0 is unique. By (2.21), $\hat{H}$ minimizes the forecast MSE, $e_T(\hat{H}) \to \sigma^2_u + Q(H_0)$.

To illustrate the selection of H for rolling window forecast, consider the case of a random walk $\{\beta_t\}$, when $\{\nabla\beta_t\} \sim IID(0,\sigma^2_b)$ and $\{u_t\} \sim IID(0,\sigma^2_u)$ are such that $\sigma^2_b/\sigma^2_u < 2/3$. Then,

$$
\begin{aligned}
&e_T(1) \sim \sigma^2_u + Q(1) = \sigma^2_u + (\sigma^2_b + \sigma^2_u), \qquad H = 1;\\
&e_T(2) \sim \sigma^2_u + Q(2) = \sigma^2_u + (7/4)\sigma^2_b + (1/2)\sigma^2_u < e_T(1), \quad H = 2.
\end{aligned}
$$

Hence, $\hat{H} \geq 2$. MSE could be minimized further by selecting K, giving smallest $Q_T(\hat{H})$.

2.6 Weakly persistent case and structural break

Next we consider the bounded stochastic trend (b4), deterministic trend (b5), and structural break (b6). Weaker persistence of these models results in $\hat{H}$ increasing with T.

Theorem 3 *Let* $y_t = \beta_t + u_t, t = 1,\cdots,T$*, where* $\{\beta_t\}$ *is as in (b4), (b5) or (b6),* $\{u_t\} \in I(0)$*. Then, as* $T\to\infty$*, uniformly in* $H \in I_T$*,*

$$
\begin{aligned}
Q_T(H) &= \hat{\sigma}^2_{u,T} + \{E[Q_T(H) - \hat{\sigma}^2_{u,T}]\}(1+o_p(1)) \qquad (2.23)\\
&= \hat{\sigma}^2_{u,T} + \bar{Q}_T(H)(1+o_p(1)),
\end{aligned}
$$

where

$$\bar{Q}_T(H) := T^{-1} s^2_{\nabla\beta} V^2_{UR\tilde{\beta},H} + Q_u(H) \quad \textit{in case (b4)}, \tag{2.24}$$
$$= \{(H/T) s^2_{\nabla\tilde{\beta}} \nu_{3,K} + H^{-1} b_{u,K}\}(1+o(1)), \qquad H \to \infty.$$
$$\bar{Q}_T(H) := (H/T)^2 \kappa(g) V^2_{tr,H} + Q_u(H) \quad \textit{in case (b5)}, \tag{2.25}$$
$$= \{(H/T)^2 \kappa(g) \nu_{1,K} + H^{-1} b_{u,K}\}(1+o(1)), \quad H \to \infty.$$

In case of the break (b6), for the rolling window weights, uniformly in H, with $\tau := T - L$,

$$\bar{Q}_T(H) \;:=\; Q_{br}(H) + Q_u(H), \tag{2.26}$$
$$Q_{br}(H) \;:=\; \Delta^2 T^{-1} \sum_{t=L}^{T} \big| \sum_{j=1}^{L} w_{tj;H} \big|^2 \tag{2.27}$$
$$= \Delta^2 \frac{H}{3T}(1+o(1)), \qquad \alpha \le H \le \tau;$$
$$= \frac{\Delta^2 \tau}{T}\{1 - \frac{\tau}{H} + \frac{\tau^2}{3H^2}\}(1+o(1)), \qquad \tau < H \le T.$$

Again, minimization of $Q_T(H)$ reduces to that of $\bar{Q}_T(H)$, and the forecast, based on $\hat{H}$ minimizes MSE.

Corollary 3 *Under assumptions of Theorem 3, in cases* (b4) *to* (b6),

$$\inf_{H \in I_T} e_T(H) = \sigma_u^2 + \inf_{H \in I_T} (Q_T(H) - \hat{\sigma}^2_{u,T})(1+o(1))\}; \tag{2.28}$$
$$= e_T(\hat{H}) + o_p(T^{-1/2}) = Q_T(\hat{H}) + o(T^{-1/2}) \quad \textit{in cases (b4) and (b6)},$$
$$= e_T(\hat{H}) + o(T^{-2/3}) = Q_T(\hat{H}) + o_p(T^{-2/3}) \quad \textit{in case (b5)}. \tag{2.29}$$

Remark 3 To illustrate selection of $\hat{H}$ and the order of $e_T(\hat{H})$ for the rolling window forecast in cases (b4) to (b6), assume that $\{u_t\} \sim IID(0, \sigma_u^2)$. Then $Q_u(H) = \sigma_u^2 > 0$, and Theorem 3 and Corollary 3 imply the following results about $\hat{H}$ and $e_T(\hat{H})$.

(b4) For the bounded unit root β_t,

$$\hat{H} \sim argmin_H \{ \frac{H}{T} s^2_{\nabla\tilde{\beta}} \nu_{3,K} + \frac{\sigma_u^2}{H} \} \sim \frac{\sigma_u}{s_{\nabla\tilde{\beta}} \sqrt{\nu_{3,K}}} T^{-1/2},$$
$$Q_T(\hat{H}) = \sigma_u^2 + 2 s_{\nabla\tilde{\beta}} \sqrt{\nu_{3,K}} \sigma_u T^{-1/2} + o_p(T^{-1/2}).$$

(b5) For the bounded smooth trend β_t,

$$\hat{H} \sim argmin_H \{\frac{H^2}{T^2}\kappa(g)\nu_{1,K} + \frac{\sigma_u^2}{H}\} \sim \Big(\frac{\sigma_u^2}{2\kappa(g)\nu_{1,K}}\Big)^{1/3} T^{-2/3},$$
$$Q_T(\hat{H}) = \sigma_u^2 + 2\sigma_u\{2\sigma_u\kappa(g)\nu_{1,K}\}^{1/3}T^{-2/3} + o_p(T^{-2/3}).$$

(b6) For the forecast under break, when $\tau = T - L >> T^{1/2}$,

$$\hat{H} \sim (\sigma_u/\Delta)\sqrt{3T}, \tag{2.30}$$
$$Q_T(\hat{H}) = \sigma_u^2 + 2\sigma_u\Delta/\sqrt{3T} + o_p(T^{-1/2}).$$

To verify (2.30), notice that

$$\inf_{H\le\tau} \bar{Q}_T(H) \sim \inf_{H\le\tau} \big(\frac{\Delta^2 H}{3T} + \frac{\sigma_u^2}{H}\big) \sim \frac{2\sigma_u\Delta}{\sqrt{3T}}, \quad \hat{H}_1 := argmin_{H\le\tau}\bar{Q}_T(H) \sim \frac{\sigma_u\sqrt{3T}}{\Delta},$$

whereas $\tilde{Q}_T(H) \sim \Delta^2\tau/T > T^{-1/2}$, for $H > \tau$. Hence, for breaks such that $\tau > T^{1/2}$, $\hat{H} \sim \hat{H}_1 \sim (\sigma_u/\Delta)\sqrt{3T}$ and the forecast error $e_T(\hat{H})$ is as in (2.30). In finite samples, if $\tau > (\sigma_u/\Delta)\sqrt{3T}$, then $\hat{H}_1 \le \tau$, and the forecast will be based on the data from the post-break period. However, for more recent breaks, such that $\tau < (\sigma_u/\Delta)\sqrt{T}$, it holds $\hat{H} \sim argmin_H H^{-1}(\sigma_u^2 + o(1))$, indicating that the forecast will not be affected by the break and not switching to the post-break period.

Similar patterns are observed in the case of EWMA weights. The above examples show that the tuning parameter $\hat{H}$ is robustly adjusted to the unknown structure of the data optimizing the MSE of the forecast. The range of $\hat{H}$ may stretch over all of the interval I_T.

2.7 Examples

In order to get a better feel for the behaviour of the data-selected tuning parameters, we consider one single realization of sequentially computed $\hat{H}_t$, $t = T_0, T_0 + 1, \cdots, T$ for two structural change experiments used in our Monte Carlo study below. We look at rolling window forecasts. Figures 1 and 2 report the beginning (dashed line) of the data selected rolling window for a structural break in the mean (Experiment 4 of our Monte Carlo study) and a unit root model (Experiment 11), respectively. The sample size T is 200 and the forecasting starts at $T_0 = 100$.[2] For comparison, we also report the first observation in the

[2] Details on how the parameter $\hat{H}_t$ is estimated are given in Section 3.

data-estimated rolling window when the model has no structural change (Experiment 1 in the Monte Carlo study), based on the same realizations of the noise u_t, as in the previous two cases (solid line). The vertical distance between the diagonal and the dashed (solid) line for a given $t = 100, \cdots, 200$ shows the time span of observations (graphical realization of the tuning parameter) used for forecasting, that is $t - \hat{H}_t, t$. It is clearly seen, that under structural change the estimated tuning parameter selects a much smaller sample used for forecast than in its absence. Figure 1 shows, that for the structural break (at observation 110) the data dependent method is attempting to get more information about the change immediately after the break by initially using a larger sample for forecasting. This then becomes smaller than that in the no-change case, as more data after the breakpoint accrue. Interestingly, after observation 125, the starting point of the rolling window is the first post break observation 111 (the dotted line), as suggested by theory. 125 is close to the roughly estimated theoretical switching time $110 + \sqrt{3(110)} = 128$ (see Remark 3). Moreover, it remains at that point for much of the rest of the sample. In Figure 2, we can see that with a unit root, the window remains short throughout the sample. A final diagnostic for the method may be obtained by considering the value of the estimated mean squared error obtained in real time. This is given in Figure 3, where the solid line relates to the stationary case, the long-dashed line to the structural break case and the short-dashed line to the unit root case. The smallest MSE is obtained in the stationary case followed by the structural break case and finally the unit root case, which is the ranking one would expect.

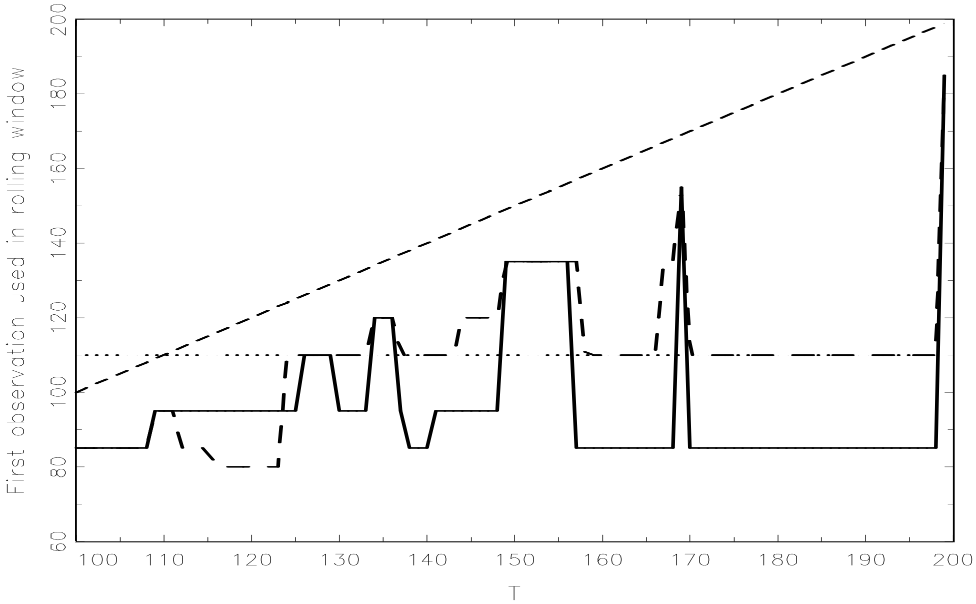
2.8 Extensions

Our proposed method extends in several practically relevant ways. In this section we briefly discuss some of these extensions.

Sub-samples

The first relates to the possibility that the forecast MSE may be evaluated and minimized over different sample periods, in order to select the optimal subsample and a specific tuning parameter. Theory indicates that an optimal tuning parameter and subsample may be selected evaluating MSE over different sample periods $[k, \cdots, T]$. Selecting H, one may wish to consider only the recent (and most relevant data in the evaluation of the MSE to reflect the evolution of structural change). This is achieved by an extended two-parameter minimization

Figure 1: Realization of the data selected rolling window for a structural break. The solid line represents beginning of the window when there is no structural change, and the dashed line (long dashes) the starting point of the window for a structural break model with a break at observation 110 (Experiment 4 of the Monte Carlo study). The dotted line indicates the first post break observation and the dashed line (short dashes) represents the last observation in the window.

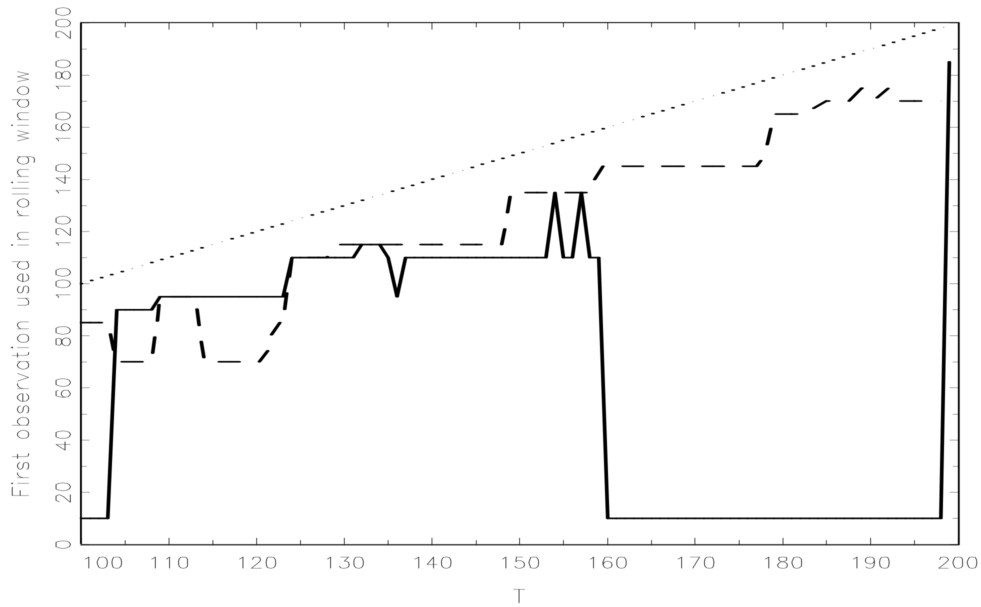


procedure given by

$$Q_T(H,k) := \frac{1}{T-k}\sum_{t=k}^{T}(\hat{y}_{t|t-1} - y_t)^2, \qquad \{\hat{H},\hat{k}\} := \text{argmin}_{H\in I_T, k\in\{k_{\min},\cdots,k_{\max}\}} Q_T(H,k). \tag{2.31}$$

The selected values of $(\hat{H},\hat{k})$ can then be used to construct forecasts based on the subsample $[\hat{k},\cdots,T]$. This value of H may be different from that obtained by the optimization in (2.8). Such a procedure, when building forecasts, seeks for an optimal subsample $y_{\hat{k}},\cdots,y_T$ ('stability period') and an optimal tuning parameter $\hat{H} = \hat{H}(\hat{k})$ for it. Observe that for the rolling window forecast, obviously $\hat{k} \le \hat{H} \le T$, however using exponential downweighting, only data $y_{\hat{k}},\cdots,y_T$ should be used.

Figure 2: Realization of the data selected rolling window for a unit root. The solid line shows the beginning of the window when there is no structural change, and the dashed line the starting point of the window for a unit root model (Experiment 11 of the Monte Carlo study). The dotted line shows the last observation in the window.

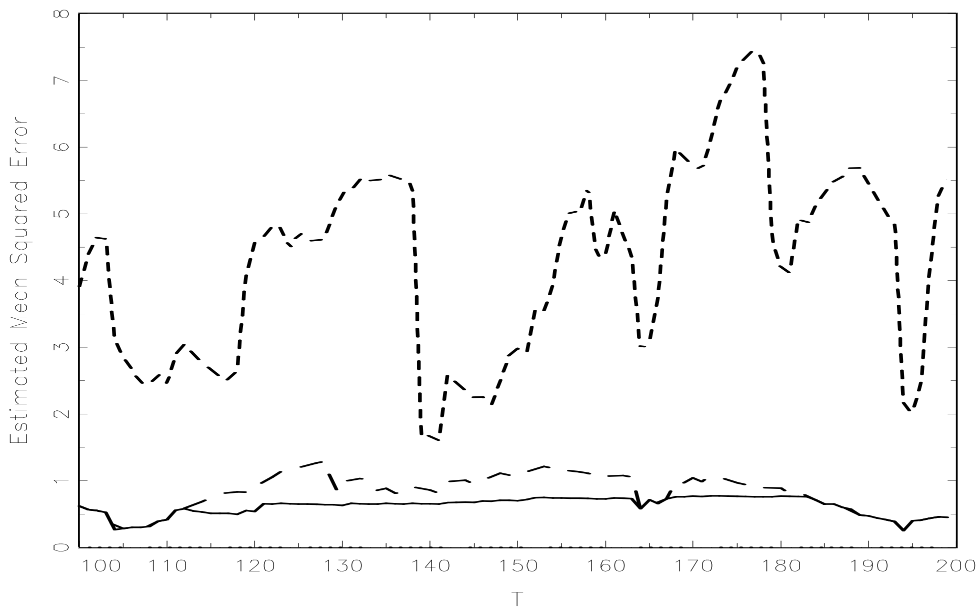


The advantage of the two parameter procedure becomes obvious in rolling window forecasts under the break in the mean, discussed in Remark 3. If the rolling window is selected using all the data in a large sample $y_1, \cdots, y_T$, then it takes $\sqrt{T}$ time lags for the forecast to switch to the postbreak data. However, the switch may be faster when less observations are used (i.e., when $\hat{k} >> 1$ is selected, reducing the weight of irrelevant past information). Our theoretical findings show that the two parameter minimization will minimize the forecast MSE leading to smallest possible MSE with optimal downweighting and the most relevant data subsample.

Non-parametric method

The above analysis presupposes a particular parametric form for the weight function. While that might be desirable from the usual motivation of parsimony, in some circumstances it

Figure 3: Realization of the estimated MSE. The solid line shows the MSE for the stationary case, the long-dashed line for the structural break case and the short-dashed line for the unit root case.



will be restrictive. For example, monotonic downweighting might be counterproductive when data may come from a processes that follows a finite number of regimes. Data from the same regime as that holding during the latest forecast period may be more relevant compared to more recent data. To account for such possibilities, we construct a nonparametric weighting scheme.

Again we focus on the simple location model (2.1) assuming that $\beta_{T,t}$ is some smooth deterministic function of t and u_t is a standardized $iid(0,1)$ noise. We consider forecasts of y_{t-1} of the form

$$\hat{y}_{t|t-1} = \sum_{j=1}^{t-1} w_{tj} y_{t-j}. \tag{2.32}$$

We wish to determine a nonparametric set of weights w_{Tj}, $j = 1, \cdots, T-1$, such that the forecast MSE of $\hat{y}_{T|T-1}$ is minimised subject to $\sum_{i=1}^{T-1} w_{Tj} = 1$. Letting $\tilde{\beta}_t = \beta_t - \beta_T$,

$$E(\hat{y}_{T|T-1} - y_T)^2 = \big(\sum_{j=1}^{T-1} w_{Tj}\tilde{\beta}_{T-j}\big)^2 + \sigma_u^2 \sum_{j=1}^{T-1} w_{Tj}^2.$$

We construct the Lagrangean

$$L(\lambda, w_{T1}, \cdots, w_{T,T-1}) = \big(\sum_{j=1}^{T-1} w_{Tj}\tilde{\beta}_{T-j}\big)^2 + \sigma_u^2 \sum_{j=1}^{T-1} w_{Tj}^2 - \lambda\big(\sum_{i=1}^{T-1} w_{Tj} - 1\big).$$

Taking derivatives of L w.r.t. w_{Tj}'s and equalling them to zero, gives $T-1$ equations

$$(\tilde{\beta}_{T-j}^2 + \sigma_u^2)w_{Tj} + \tilde{\beta}_{T-j} \sum_{i=1, i\neq j}^{T-1} \tilde{\beta}_{T-i} w_{Ti} = \lambda/2, \quad j = 1, \cdots, T-1.$$

We need to solve this set of $T-1$ equations. As a system they are written as

$$(\tilde{B} + \sigma_u I)w_T = (\lambda/2)1, \qquad \text{or} \quad Bw_T = \Lambda, \tag{2.33}$$

where $\tilde{B} = (\tilde{\beta}_{T-j}\tilde{\beta}_{T-k})_{j,k=1,\dots,T-1}$ is $(T-1)\times(T-1)$ matrix, I is $(T-1)\times(T-1)$ unit matrix, $w_T = (w_{Tj})_{j=1,\cdots,T-1}$ is $(T-1)\times 1$ vector and 1 is $(T-1)\times 1$ unit vector.

Whence $w_T = B^{-1}\Lambda$, and λ is determined such that the sum of the elements of $B^{-1}\Lambda$ is unity. This is not an operational procedure as β_T is unknown at time $T-1$. We suggest setting $\beta_t = \hat{\beta}_t$, $t = 1, \cdots, T-1$ and $\beta_T = \hat{\beta}_T = \hat{\beta}_{T-1}$ where $\hat{\beta}_t$ denotes some estimator of β_t. This approach does not allow for a dependent u_t, but we discuss possible extensions of (2.1) below that make the assumption of a serially uncorrerelated u_t more plausible.

The method can be extended to allow for time varying variances $E(u_t^2) = \sigma_{u,t}^2$. Then, the forecast MSE takes the form

$$E(\hat{y}_{T|T-1} - y_T)^2 = \big(\sum_{j=1}^{T-1} w_{Tj}\tilde{\beta}_{T-j}\big)^2 + \sum_{j=1}^{T-1} w_{Tj}^2 \sigma_{u,T-j}^2.$$

Following the steps of the previous argument gives the following system of equations

$$(\tilde{B} + \tilde{I})w_T = (\lambda/2)1, \qquad \text{or} \quad Bw_T = \Lambda,$$

where $\tilde{I} = diagonal(\sigma_{u,T-1}^2, \cdots, \sigma_{u,1}^2)$ is $(T-1)\times(T-1)$ diagonal matrix. Once again this procedure becomes operational by replacing $\sigma_{\epsilon,t}^2$ with an estimate. We note that estimation of β_t and $\sigma_{\epsilon,t}^2$ is discussed widely in the literature when β_t and $\sigma_{\epsilon,t}^2$ are deterministic functions of time (see, e.g., Orbe, Ferreira, and Rodriguez-Poo (2005) and Kapetanios (2007)), and is discussed in Giraitis, Kapetanios, and Yates (2011) for stochastic β_t.

Dynamic weighting and regression models

Another alternative and simple way to allow for extra flexibility in the weight function is to allow the first p weights $w_1, \cdots, w_p$ $(p \geq 0)$ to vary freely by specifying

$$\tilde{w}_{tj,H} = \begin{cases} w_{t-j}, & j = t-1, \cdots, t-1, \\ K(\frac{t-j}{H}), & j = 1, \cdots, t-p-1, \qquad H \in I_T, \end{cases}$$

and standardizing the weights: $w_{tj,H} = \frac{\tilde{w}_{tj,H}}{\sum_{j=1}^{t} \tilde{w}_{tj,H}}$. This allows the first few lags of y_t to enter freely into the forecast rather than through a given parametric function. Then, Q_T can be minimized jointly over $H, \tilde{w}_1, \cdots, \tilde{w}_p$, and, potentially, even p.

Hitherto we have been dealing with a simple location model, which, although allowing for dependent u_t and accommodating a wide range of behaviours for y_t, may be considered somewhat restrictive. It can be extended to a regression model of the form

$$y_{T,t} = \beta'_{T,t} x_t + u_t, \qquad t = 1, \cdots, T, \quad T \geq 1, \tag{2.34}$$

where x_t is a $K \times 1$ vector of predetermined (stochastic) variables, β_t's are $K \times 1$ vectors of parameters, and u_t is a stationary dependent noise process that is independent of x_t. Setting $\beta_t = \left(E\left(x_t x_t'\right)\right)^{-1} E\left(x_t' y_t\right) = \left(\Sigma_t^{xx}\right)^{-1} \Sigma_t^{xy}$, where $\Sigma_t^{xx} = [\sigma_{ij,t}^{xx}]$, and $\Sigma_t^{xy} = [\sigma_{i,t}^{xy}]$ are corresponding covariance matrices, we allow the relevant expectations to be time-varying. Here, the main task of interest becomes to estimate the expectations Σ_t^{xx} and Σ_t^{xy} over time by the robust methods outlined above. To achieve that, we write down $z_{ij,t} = x_{i,t} x_{j,t}$ and $z_{i,t} = x_{i,t} y_t$ as simple location models: $z_{ij,t} = \sigma_{ij,t}^{xx} + u_{ij,t}$, and $z_{i,t} = \sigma_{i,t}^{xy} + u_{i,t}$. This way, the regression (2.34) can be reduced to estimation of a sequence of simple location models. Subsequently, some practical questions regarding estimation of the tuning parameters arise, i.e. whether each of those simple location models should be treated independently or pooled, which is more straightforward to handle.

An alternative and perhaps more attractive way to accommodate regressors is to modify (2.2) so that

$$\hat{y}_{t|t-1} \equiv \hat{y}_{t|t-1}(H_1, \cdots, H_{K+1}) = \sum_{j=1}^{t-1} w_{tj;H_1} y_{t-j} + \sum_{i=1}^{K} \sum_{j=1}^{t-1} w_{tj;H_{i+1}} x_{i,t-j}, \tag{2.35}$$

and then minimise $Q_T(H_1, \cdots, H_{K+1})$ with respect to $\boldsymbol{H} = (H_1, \cdots, H_{K+1})'$, where $Q_T(H_1 , \cdots, H_{K+1})$ is defined similarly to $Q_T(H)$ in (2.8). It is equally easy to consider multi-step ahead forecasts by simply setting $\hat{y}_{t|t-s} \equiv \hat{y}_{t|t-s}(H) = \sum_{j=s}^{t-1} w_{tj;H}^{(s)} y_{t-j}$, and then minimizing the relevant MSE.

2.9 Theoretical conclusions

We conclude this section by noting some important implications from our analysis. Firstly, the dominant tendency in the forecasting literature of using models developed for other purposes such as impulse response or policy analysis, to obtain forecasts may be counterproductive. Our arguments suggest that if good forecasting is the aim, then forecasting by averaging and appropriately weighting down past data, without engaging in further modelling, is a viable strategy.

Secondly, appropriately downweighting past can provide a general approach for handling trends of any nature. Our theoretical results show that this method applies for stochastic, linear or nonlinear deterministic trends and structural breaks without knowledge of the nature of the trend. It is therefore a tractable method for forecasting the levels of apparently nonstationary processes. As a result it bypasses difficult problems of combining appropriate detrending of level series with the subsequent forecasting of stationary processes. Importantly, the proposed forecasting approach continues to be valid if a series is actually stationary.

3 Monte Carlo Study

In this section we explore the finite sample performance of the forecasting strategies discussed in Section 2.1. The simulation study consides Monte Carlo experiments for the forecast of y_{T+1} based on the sample $y_1, \cdots, y_T$ for a number of specific designs for the simple location model (2.1) with β_t as in (b1) to (b6). For the one-step ahead forecasts, the benchmark is the sample mean forecast $\hat{y}_{benchmark,T+1} = T^{-1}\sum_{t=1}^{T} y_t$, while for two-step ahead forecasts, it is $(T-1)^{-1}\sum_{t=1}^{T-1} y_t$. The benchmark disregards the possibility of structural change. We compare the performance of forecasts in terms of relative root MSE.

3.1 Design: data generating processes

We consider the following location shift model for generating the data:

$$y_t = \beta_t + u_t, \qquad t = 1, \cdots, T,$$

namely a version of (2.1), where u_t is a standard normal $iid(0,1)$ noise, or an $AR(1)$ process with parameter $\rho = 0.7$ and standard normal iid innovations. The process β_t is either a deterministic or stochastic trend as in (b1) to (b5), or a process with a break in the mean as in (b6). We consider the following data generating processess, denoted in tables as $Ex1$–$Ex11$:

1. $y_t = u_t$.
2. $y_t = 0.05t + 5u_t$.
3. $y_t = 0.05t^{0.5+0.75(t/T)} + 5u_t$.
4. $y_t = \begin{cases} \epsilon_t, & t \leq (11/20)T, \\ 1 + \epsilon_t, & t > (11/20)T. \end{cases}$
5. $y_t = 2\sin\left(\frac{2\pi t}{T}\right) + 3u_t$.
6. $y_t = 5\sin\left(\frac{2\pi t}{T}\right) + 3\epsilon_t$.
7. $y_t = (0.025t - 2.5)^2 + 5u_t$.
8. $y_t = (0.025t - 2.5)^2 + 3u_t$.
9. $y_t = \frac{2}{\sqrt{T}}\sum_{i=1}^{t} v_i + u_t$.
10. $y_t = \frac{2}{\sqrt{T}}\sum_{i=1}^{t} v_i + 0.05t + u_t$.
11. $y_t = 2\sum_{i=1}^{t} v_i + u_t$,

where v_t is a standard normal $iid(0,1)$ sequence.

The selection of deterministic trends provides a variety of shapes of the functions driving the structural change in the unconditional mean of y_t.

$Ex1$ is the case of no structural change. Here, as long as the noise u_t is an iid or very weakly dependent process, the benchmark (sample mean) forecast should do best, and the robust methods at most should not lag far behind the benchmark. However, when u_t is sufficiently persistent, such as, e.g., an AR(1) process with $\rho = 0.7$, then the robust forecast with EWMA weights should outperform the benchmark, see Remark 1. Theory indicates, that the exponential weights should outperform the rolling window, but it leaves open the possibility that the rolling window can outperform the benchmark when a stationary process y_t becomes persistent. Table 2 indicates that the latter is not true for rolling window, but it is obvious for exponential weights.

The functional forms in $Ex2$ and $Ex3$ are respectively a linear and nonlinear monotonic trend of type (b2). While such trends may be unrealistic, at least for processes which have been detrended applying filters or differencing, they provide a useful benchmark. Further, these trends are sufficiently subtle and minor to be swamped visually by the noise process. Functions in $Ex7$ and $Ex8$ provide hump shaped quadratic trends which again are likely to be relevant in practice. According to the theory, for such (b2)-type trends, robust methods should obviously outperform the benchmark. Moreover, performance of the robust methods should improve when the level of the noise (or $\mathrm{Var}(u_t)$) decreases, which is confirmed by simulations in Tables 1 and 2 comparing $Ex7$ and $Ex8$. The tables also show strong benefits from the use of robust forecasting when the noise u_t becomes more dependent, see $Ex7$ and $Ex8$ in Tables 1 and 2. Obviously, presence of a stronger nonlinear trend improves the effect of using a robust forecast, as seen from $Ex2$ and $Ex3$ in Tables 1 and 2.

The purpose of $Ex4$ is to introduce a break in the mean, to see if our robust methods can help under traditional structural change specifications. The break occurs at time $L \sim T/2$, and the after-break time $T/2$ is greater than $\sqrt{T}$, as required in (b6). Hence the break is not 'too recent' and it will be taken into account by the robust forecasting method, leading to significant improvement of forecast quality comparing to the benchmark, see Tables 1 and 2. Moreover, the effect is amplified by the increase of dependence in the error process u_t.

Functions in $Ex5$ and $Ex6$ represent smooth cyclical bounded trends. These are more likely to remain after standard detrending and provide a realistic scenario. Tables 1 and 2 show that presence of cyclical trends is taken into account by the adaptive forecast. Moreover, wider oscillation of the trend in $Ex6$ seems to lead to a stronger deterioration of relative performance of the benchmark.

Finally, $Ex9-10$ deal with a bounded stochastic trend β_t which is relevant to popular time-varying coefficient specifications in the macroeconometric and forecasting literature, while $Ex11$ deals with a random walk (unit root) process, observed under noise. Tables 1 and 2 show that robust forecast outperforms the benchmark in $Ex9-10$, and by more than 80% in $Ex11$. Moreover, exponential weights outperform rolling windows.

3.2 Forecast methods

We examine the robust forecasting using three classes of parametric weight functions.

Rolling window. This is a flat weight function

$$\begin{aligned} w_{tj,H} &= 1/H, \qquad t-H \le j \le t-1, \quad \text{if } H \le t-1, \\ &= 1/t, \qquad 1 \le j \le t-1, \quad \text{if } H \ge t, \\ &= 0, \qquad \text{otherwise,} \end{aligned}$$

giving equal weight to recent data and zero weight to older data, see (2.5).

We denote it in tables by '*Rolling H*' where H is the window size.

Exponential weights (EWMA). For $0 < \rho < 1$,

$$w_{tj;\rho} = \frac{\rho^{t-j}}{\sum_{k=1}^{t-1} \rho^k} \qquad 1 \le j \le t-1.$$

Here the main weight is placed on the last few data, downweighting others to zero exponentially fast when ρ is small, and more equally when ρ is close to 1, see (2.6).

We denote this as '*Exponential* ρ'.

Polynomial weights. For $\alpha > 0$,

$$w_{tj;\rho} = \frac{(t-j)^{-\alpha}}{\sum_{k=1}^{t-1} k^{-\alpha}} \qquad 1 \le j \le t-1. \tag{3.1}$$

They are downweighting the past slower than exponential weights. We denote them in tables by "*Polynomial* α".

We consider forecasts with fixed values of H and ρ, and data selected values $\hat{H}$, $\hat{\rho}$ and $\hat{\alpha}$ for the tuning parameters. In case of polynomial weights we do not use fixed values of parameters. We set $H = 20, 30$ for rolling window; $\rho = 0.99, 0.95, 0.9, 0.8, 0.7$ and 0.5 for exponential weights. Using fixed values will allow us to compare the performance of the forecast with data tuned parameter with the best fixed one that gives the smallest Monte Carlo forecast MSE among fixed tuning parameters, roughly speaking with the best possible. Our objective is to verify in simulations that these two MSE's are comparable, as indicated by Corollaries 1 to 3.

Non-parametric method. We also consider the non-parametric forecast method as in (2.32) and (2.33) based on the non-parametric weighting scheme. In the Tables 1-3 the corresponding results are referred to as *'Non-parametric'*.

'Rolling ($\hat{k}\hat{H}$) method' This is the rolling window forecast where $\hat{k}$ and $\hat{H}$ are selected minimizing $Q_{T,k}(H)$ in H and k as in (2.31).

Averaging method. The final robust method we examine is the *averaging method* of rolling window forecasts over different periods advocated by Pesaran and Timmermann (2007):

$$\bar{y}_{T+1|T} = \frac{1}{T}\sum_{H=1}^{T} \hat{y}^{(H)}_{T+1|T}, \qquad \hat{y}^{(H)}_{T+1|T} = \frac{1}{H}\sum_{t=T-H+1}^{T} y_t. \tag{3.2}$$

It combines rolling window forecasts of y_{T+1} using all possible windows that include the last available observation. One major advantage of this method is that it does not require selecting tuning parameter apart from choosing the mimimum sample size used for forecasting, which choice is usually of minor significance for the performance of this method. We refer to this method as *Averaging*.

3.3 Monte Carlo results

Tables 1-3 are produced as follows. We choose a particular starting point in time T_0 of the forecasting by any given method. Then, one-step ahead forecasts $y_{T_0|T_0-1}, \cdots, y_{T|T-1}$ are computed for the subsequent period ending at T. To compare different forecast methods, as the performance criterion we use the forecast root MSE relative to the benchmark of the sample mean of all data (MSE_{RR}) . For method "i", we compute $MSE_i = \frac{1}{T-T_0}\sum_{t=T_0}^{T}(\hat{y}^{(i)}_{t|t-1} - y_t)^2$ and define the relative root $MSE_{RR} = \sqrt{MSE_i}/\sqrt{MSE_0}$ where MSE_0 correspond to the benchmark forecast by the sample mean. For all experiments, forecasting starts at $T_0 = 100$, an the samples size is $T = 200$. MSE_{RR} below unity shows that the forecast method outperforms the sample mean.

The relative root MSE results for models $Ex1$ to $Ex11$ obtained applying various forecasting methods with data selected and fixed tuning parameters are reported Tables 1 and 2. In Table 1, the noise u_t is an iid standard normal process, whereas in Table 2, u_t are dependent variables, generated by a stationary AR(1) process with parameter $\rho = 0.7$ and iid standard normal innovations. Finally, in Table 3 we report similar results as in Table 1 but for two-step ahead forecasts. The results in Table 3 are, in general, similar to those in Table 1 and so we focus mainly on Tables 1 and 2.

The first column, labelled $Ex1$, corresponds to the stationary case $y_t = u_t$, or no-change baseline. In the iid case, as expected, the sample mean outperforms the forecasts for each method, especially those penalised by the loss of information from strong discounting. However, for sufficiently dependent u_t, discounting improves the forecast as indicated by Remark 1.

For the other experiments, in almost all cases, downweighting beats the sample mean in the sense that the MSE_{RR} is considerably below unity. Generally, all these methods are useful, including the simple rolling window and averaging method. In the case of a fixed tuning parameter, for the model with a strong trend, the largest reduction of MSE_{RR} comes from the exponential weights with the highest discount rates. In the set of experiments with iid noise u_t, for this particular design, the exponential weights with a $\rho = 0.9$ discount perform well in the sense that it has the largest number of minimum MSE_{RR}'s. Although the tuned exponential weights are not the best, they are where they should be according to theory:

comparable to the best fixed value methods and never among the poor performers. Note, that the exponential weights with a $\rho = 0.9$ discount can perform considerably worse than the tuned exponential weights in cases such as $Ex3$ and $Ex11$ illustrating the importance of data-dependent tuning.

Given that optimal fixed ρ for exponential weights cannot be observed in practice, our simulation study suggests the efficiency and usefulness of data based downweighting. The nonparametric method similarly offers a powerful alternative, for iid noise u_t slightly beating the tuned parameter methods in all cases except Ex11, see Table 1. However, being designed for an iid noise u_t, in case of a dependent AR(1) noise this method is outperformed by the parametric tuning methods, see Table 2.

Comparing exponential, rolling window and polynomial methods, exponential method outperforms rolling windows while the latter beats polynomial windows when the noise u_t is dependent and is outperformed by it when the noise is iid. The averaging method outperforms the benchmark but is beaten by the rolling windows with data selected $\hat{H}$. The rolling window forecast using a data dependent window, $\hat{H}$, and an evaluation period $[\hat{k}, T]$, is equivalent to a rolling window with $\hat{H}$ and $k = 1$ under the iid noise, but outperforms it when the noise, u_t, is dependent.

It is worth noting that, in applications, one could select from a set of available forecasts with data dependent and fixed discounting rates, the one minimizing the criterion function $Q_T(H)$ of (2.8), and respectively, the forecast MSE, $\sigma_u^2 + E(Q_T(H) - \hat{\sigma}_u^2)$. This possibility illustrates the more wide relevance our cross-validation approach.

Simulation results suggest that robust forecasting methods with data selected parametric downweighting are effective in the face of a variety of types of structural change, and that data-dependent tuning is a viable approach, in some cases preventing significant errors. For models with iid noise, nonparametric methods can be very effective. It remains to be seen in the next section whether they are effective in practice.

4 Empirical Application

In this section we examine how our methods would have fared when applied to a large range of UK and US quarterly data series.[3] We are not trying to establish the best methods for particular data series, but instead to get an impression of whether the issues identified above

[3]We take no account of real-time data revisions.

are important in practice. In all cases we transform series to stationarity and employ either a simple location or an AR(1) forecasting model. We use data on 94 series for the UK and 97 for the US, taken for ease of comparison from Eklund, Kapetanios, and Price (2010). [4] The data span 1977Q1 to 2008Q2 for the UK (1960Q1 to 2008Q3 for the US). We examine two forecast evaluation sub-periods within this (1992Q1 to 1999Q4 and 2000Q1 to 2008Q2 for the UK and 1992Q2 to 200Q1 and 2000Q2 to 2008Q3 for the US) so that the periods evaluated are the same length for comparability. For each series, we compare RMSFEs to that from the benchmark simple location model estimated using equal weights on all data. The methods we report relate to those in the Monte Carlo study, and are rolling window forecasts, averaging across estimation periods, exponentially weighted moving average forecasts, polynomially weighted moving average forecasts and forecasts produced using nonparametric weights.

Table 4 reports results for the location model over the two samples we examine. We report the median RRMSFE (relative to the full sample (equal weight) benchmark. We also report some other summary statistics for the RRMSFE. Namely, the minimum, maximum, variance and skewness. DM1 and DM2 report the number of significant Diebold-Mariano tests where the null is equality of the robust method and the benchmark. The alternative for DM1 is that the benchmark is the better forecast, and for DM2 that the robust method is superior.

In almost all cases, the methods beat the benchmark for the median. Results are particularly good for the first half of the sample. Clearly, looking at the minima, in some cases there is an enormous gain, whereas the penalty in the worst case, although large in some cases, is several orders of magnitude lower. As for the Monte Carlo, the EWMA with a fixed large discount ($\rho = 0.50$) performs poorly. Tuning parameters for EWMA provides a good median performance, whereas (in contrast to the Monte Carlo) the nonparametric methods perform relatively badly. Nevertheless, all methods significantly outperform the equal weight benchmark in at least 19% of cases.

Qualitatively, for the US the results are even stronger. The median reduction in the optimised EWMA is large, over 30% in both samples. Generally, most methods preform relatively better than for the UK data.

One final issue we examine relates to the well known fact that AR models are a good benchmark for many macroeconomic series. We have not attempted to incorporate lags into the procedure (as described in Section 2.8). Despite this, when we compare the tuned expo-

[4] See Eklund, Kapetanios, and Price (2010) for details of the data.

nential method to a benchmark AR process, the median relative RMSE for the UK remains below unity (0.925 for the first subsample and 0.962 for the second). For all other methods, the relative RMSE slightly exceeds unity. Rather less favourable results hold for the US (0.992 and 0.996), but the median RMSE is still below unity. This is further evidence for the usefulness of the data based tuning.

5 Conclusions

Forecast methods that are known to be robust to historical structural change, have been recently found to be useful forecasting tools under ongoing structural change. They include rolling regressions, forecast averaging over different windows and exponentially weighted moving averages. However, the, *a priory* set, degree of downweighting older data, which is a common feature shared by these methods, is suboptimal by its nature. The alternative approach suggested here is that, although we do not know the structure of the model and the nature of structural change, we can make the choice of the tuning parameter data-dependent and select it by cross-validation using in sample forecast performance. As we have shown, such discounting has a number of attractive properties. It minimizes asymptotic forecast MSE over the class of parametrically weighted moving average forecasts. Rather remarkably, it allows also the evaluation of the forecast error, and provides a framework for a number of new developments for forecasting under ongoing structural change. Our theory and small sample evidence suggests that exponential weighting may be most helpful and efficient, and that data selected tuning can provide a useful framework for avoiding large forecast errors.

The simulation study and the empirical exercise using over 190 UK and US macroeconomic series show that fixed low-discount EWMA weighting is often good, but is outperformed by the data selected downweighting. This is strong support for our approach, motivated by the impossibility of knowing the optimal degree of discounting *ex ante*.

References

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–56.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

CHOW, G. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–603.

CLEMENTS, M. P., AND D. F. HENDRY (1998a): *Forecasting economic time series*. CUP, Cambridge.

CLEMENTS, M. P., AND D. F. HENDRY (1998b): "Intercept corrections and structural change," *Journal of Applied Econometrics*, 11, 475–94.

DAHLHAUS, R. (1996): "Fitting Time Series Model to Nonstationary Processes," *Annals of Statistics*, 25, 1–37.

EKLUND, J., G. KAPETANIOS, AND S. PRICE (2010): "Forecasting in the presence of recent structural change," Bank of England Working Paper 406.

GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2011): "Inference on stochastic time-varying coefficient models," Queen Mary, University of London Working Paper no. 540.

GIRAITIS, L., H. KOUL, AND D. SURGAILIS (2012): *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

HENDRY, D. F. (2000): "On detectable and non-detectable structural change," *Structural Change and Economic Dynamics*, 11, 45–65.

KAPETANIOS, G. (2007): "Estimating Deterministically Time Varying Variances in Regression Models," *Economics Letters*, 97(2), 97–104.

KAPETANIOS, G., V. LABHARD, AND S. PRICE (2006): "Forecasting using predictive likelihood model averaging," *Economics Letters*, 91, 373–379.

Orbe, S., E. Ferreira, and J. Rodriguez-Poo (2005): "Nonparametric Estimation of Time Varying Parameters under Shape Restrictions," *Journal of Econometrics*, 126, 53–77.

Pesaran, M. H., and A. Timmermann (2007): "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137, 134–61.

Rossi, B. (2012): "Advances in Forecasting Under Instability," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann. Elsevier-North Holland.

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

A Appendix: Proofs

A.1 Proof of Theorems 1-3 and Corollaries 1-3

We decompose the objective function $Q_T(H)$ and the forecast MSE, $e_T(H) = E(\hat{y}_{T+1|T} - y_T)^2$, into terms corresponding to β_j and u_j of (2.1) which will be analyzed separately.

Let

$$v_{tj;H} := w_{t,t-j}, \quad 1 \le j \le t-1; \qquad v_{tj;H} := 0, \quad t \le j < \infty. \tag{A.1}$$

Then

$$\begin{aligned} \hat{y}_{t|t-1} - y_t &= \sum_{j=1}^{t-1} w_{tj;H}(y_j - y_t) = \sum_{j=1}^{t-1} v_{tj;H}(y_{t-j} - y_t) \\ &= \sum_{j=1}^{t-1} v_{tj;H}(\beta_{t-j} - \beta_t) + \sum_{j=1}^{t-1} v_{tj;H}(u_{t-j} - u_t) =: e_{\beta,tH} + e_{u,tH}. \end{aligned} \tag{A.2}$$

Hence, $(\hat{y}_{t|t-1} - y_t)^2 = e^2_{\beta,tH} + e^2_{u,tH} + 2e_{\beta,t}e_{u,tH}$, and

$$\begin{aligned} Q_T(H) = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_{t|t-1} - y_t)^2 &= \frac{1}{T}\sum_{t=1}^{T} e^2_{\beta,tH} + \frac{1}{T}\sum_{t=1}^{T} e^2_{u,tH} + \frac{2}{T}\sum_{t=1}^{T} e_{\beta,tH}e_{u,tH} \\ &=: Q_{\beta,T}(H) + Q_{u,T}(H) + 2Q_{\beta u,T}(H). \end{aligned}$$

Because $\{\beta_t\}$ and $\{u_t\}$ are independent,

$$\begin{aligned} E[Q_T(H) - \hat{\sigma}_u^2] &= E[Q_{u,T}(H) - \sigma_u^2] + EQ_{\beta,T}(H), \\ \mathrm{Var}(Q_T(H)) &\le C\{\mathrm{Var}(Q_{u,T}(H)) + \mathrm{Var}(Q_{\beta,T}(H)) + |EQ_{\beta u,T}(H)|\}. \end{aligned} \tag{A.3}$$

Similarly,

$$e_T(H) = E[e^2_{\beta,(T+1)H}] + E[e^2_{u,(T+1)H}] = e_{\beta,T}(H) + e_{u,T}(H). \tag{A.4}$$

In the next lemmas we derive asymptotic of the terms on the r.h.s. of (A.3) required for the proof of the main results.

Claims of Theorems 1, 2 and 3 and Corollaries 1, 2 and 3 is straightforward implication of the Lemmas A.1, A.2, A.3 and A.4 below.

Recall that $I_T = [\alpha,\, T^{1-\delta}]$ where $\delta > 0$ is assumed to be small.

A.2 Properties of $Q_{\beta,T}(H)$

First we consider the case when β_t is a unit root process as in (b2). Then $\zeta_j := \nabla\beta_t = \beta_t - \beta_{t-1}$ is stationary $I(0)$ process, and the correlation

$$\gamma_\beta(j,k) := E[(\beta_j - \beta_0)(\beta_k - \beta_0)] = E[(\beta_t - \beta_{t-j})(\beta_t - \beta_{t-k})], \qquad 0 \le j,k \le t$$

does not depend on t. Since $\beta_j - \beta_0 = \sum_{l=1}^{j} \zeta_l$, then $E(\beta_j - \beta_0)^2 \sim j s^2_{\nabla\beta},\ j \to \infty$, where $s^2_{\nabla\beta}$ is the long-run variance of $\{\zeta_j\}$, and

$$\begin{aligned} \gamma_\beta(j,k) &= s^2_{\nabla\beta}(j \wedge k) + o(j \wedge k), \quad (j \wedge k) \to \infty, \\ |\gamma_\beta(j,k)| &\le C(jk)^{1/2}. \end{aligned} \tag{A.5}$$

First we approximate the mean $E[Q_{\beta,T}(H)]$ by $V^2_{\beta,H} := \sum_{j,k=1}^{\infty} v_{j;H} v_{k;H} \gamma_\beta(j,k)$. By standard argument, from definition of weights $v_{j,H}$, using (A.5) and (A.51) it follows

$$\begin{aligned} V^2_{\beta,H} &\le C\Big(\sum_{j=1}^{\infty} v_{j,H} j^{1/2}\Big)^2 \le CH, \quad H \in I_T, \\ V^2_{\beta,H} &= H s^2_{\nabla\beta} \sum_{j,k=1}^{\infty} v_{j,H} v_{k,H} (j \wedge k) + o(H^{-1}) = H s^2_{\nabla\beta} \nu_{3,K} + o(H^{-1}), \quad H \to \infty. \end{aligned} \tag{A.6}$$

The next lemma deals with the case of a unit root trend β_j (b3), and implies the corresponding results for the bounded unit root trend (b4).

Lemma A.1 *Let β_t be as in (b2) and Assumption 1 be satisfied. Then uniformly in $H \in I_T$, the following holds.*

(i) As $T \to \infty$,

$$\sup_{H \in I_T} V^{-2}_{\beta,H} |E[Q_{\beta,T}(H)] - V^2_{\beta,H}| \to 0, \tag{A.7}$$

$$\sup_{H \in I_T} V^{-2}_{\beta,H} |e_{\beta,T}(H) - V^2_{\beta,H}| \to 0. \tag{A.8}$$

(ii) In addition,

$$\sup_{H \in I_T} V^{-2}_{\beta,H} |Q_{\beta,T}(H) - E[Q_{\beta,T}(H)]| = o_p(1). \tag{A.9}$$

Proof. (i) We first show that for $H \in I_T$,

$$|E[e^2_{\beta,tH}] - V^2_{\beta,H}| \le CH^2(t \vee H)^{-1}. \tag{A.10}$$

Denote by i_t the l.h.s. of (A.10). Since by (A.5), $|\gamma_\beta(j,k)| \le C(jk)^{1/2}$, and for any real numbers $(a_j a_k - b_j b_k) = (a_j - b_j)(a_k - b_k) + b_k(a_j - b_j) + b_j(a_k - b_k)$, then

$$\begin{aligned} i_t &\le \sum_{j,k=1}^{\infty} |v_{tj;H}v_{tk;H} - v_{j;H}v_{k;H}|\, |\gamma_\beta(j,k)| \qquad \text{(A.11)} \\ &\le C\sum_{j,k=1}^{\infty} |v_{tj;H}v_{tk;H} - v_{j;H}v_{k;H}|(jk)^{-1/2} \le C\{p_{1,t}^2 + p_{1,t}p_{2,t}\} \end{aligned}$$

where $p_{1,t} := \sum_{j=1}^{\infty} |v_{tj;H} - v_{j;H}| j^{1/2}$ and $p_{2,t} := \sum_{k=1}^{\infty} v_{k;H} k^{1/2}$. By (A.45), $p_{1,t} \le CH^{3/2}(t \vee H)^{-1}$, and by (A.51), $p_{2,t} \le CH^{1/2}$. Hence, $i_t \le C\{H^3(t\vee H)^{-2} + H^2(t \vee H)^{-1}\} \le CH^2(t \vee H)^{-1}$. Therefore,

$$\begin{aligned} H^{-1}|E[Q_{\beta,T}(H)] - V_{\beta,H}^2| &\le C(HT)^{-1}\sum_{t=1}^{T} E|e_{\beta,tH}^2 - V_{\beta,H}^2| \\ &\le CHT^{-1}\sum_{t=1}^{T} t^{-1} \le CHT^{-1}\log T \le T^{-\delta}\log T \end{aligned}$$

which proves (A.7), bearing in mind (A.6).

Finally, by (A.10), $H^{-1}|e_T(H) - V_{\beta,H}^2| \le CHT^{-1} \le T^{-\delta} \to 0$, which together with (A.6) implies (A.8) and completes the proof of (i).

(ii) Let $h_{t,j} := \beta_{t-j} - \beta_t$ and $\bar v_{tj;H} := H^{-1/2}v_{tj;H}$. Recall that $Q_{\beta,T}(H) := T^{-1}\sum_{t=1}^{T} e_{\beta,tH}^2$. We will approximate

$$\bar e_{\beta,tH} := H^{-1/2}e_{\beta,tH} = \sum_{j=1}^{t-1} \bar v_{tj;H}h_{t,j} \quad \text{by} \quad e_{\Delta\beta,tH} = \sum_{j=1}^{[LH]} v_{tj;H}^{\Delta}\bar h_{t,j}, \qquad \text{(A.12)}$$

$$H^{-1}Q_{\beta,T}(H) := T^{-1}\sum_{t=1}^{T} \bar e_{\Delta\beta,tH}^2 \quad \text{by} \quad Q_{\Delta\beta,T}(H) := T^{-1}\sum_{t=1}^{T} e_{\Delta\beta,tH}^2,$$

where $e_{\Delta\beta,tH}$ is defined below. We shall show that, as $T \to \infty$,

$$E[\sup_{H\in I_T} |H^{-1}Q_{\beta,T}(H) - Q_{\Delta\beta,T}(H)|] \to 0, \qquad \text{(A.13)}$$

$$\sup_{H\in I_T} Q_{\Delta\beta,T}(H) = o_p(1), \qquad \text{(A.14)}$$

which proves (A.9).

We set $L = \log T$, whereas $v_{tj;H}^{\Delta}$ is a step function in H: letting $\Delta := \log^{-4} T$ we split the interval $I_T = \cup_{i=0}^{N}[H_i,\, H_i + \Delta)$ into small subintervals, where $H_i = \alpha + \Delta i$, $i = 0, \cdots, N = [T^{1-\delta}] + 1$. We define

$$v_{tj;H}^{\Delta} := \bar v_{tj;H_i}, \qquad H \in [H_i,\, H_i + \Delta). \qquad \text{(A.15)}$$

Variables $\bar h_{t,j}$ are m-dependent having all moments finite. Recall that $\bar h_{t,j} = \beta_{t-j} - \beta_t = -\sum_{l=t-j+1}^{t} \zeta_j$, where by Assumption 2, $\zeta_j = \nabla\beta_j = \sum_{k=0}^{\infty} b_k \eta_{j-k}$ is a linear process, $\eta_j \sim IID(0, \sigma_\eta^2)$, and $|b_k| \le Ck^{-1-v}$ for some $v > 0$. Set $\bar\zeta_j = \sum_{k=1}^{m} b_k \bar\eta_{j-k}$ where $\bar\eta_j := \eta_j I\{|\eta_j| \le (\log T)^4\}$ is the truncated noise with $m = m_T = (\log T)^p$, $p \ge 8/v$. Define $\bar h_{t,s} = -\sum_{l=t-j+1}^{t} \bar\zeta_j$. Notice that $\bar\eta_j$'s are m-dependent r.v.'s, whereas by construction $e_{\Delta\beta,tH}, e^2_{\Delta\beta,tH}$ are $\tilde m$-dependent r.v.'s, with $\tilde m := (\log T)^p + [LH] \le CH(\log T)^p$

Proof of (A.13). First we show that

$$\begin{aligned} |\bar e_{\beta,tH} - e_{\Delta\beta,tH}| &\le C[\{\Delta + L^{-5}\}S_T + S_T^*], \qquad (A.16) \\ |\bar e_{\beta,tH}| + |e_{\Delta\beta,tH}| &\le CS_T \end{aligned}$$

where $S_T := \sum_{j=1}^{t-1} j^{-3/2} h_{t,j}$ and $S_T^* := \sum_{j=1}^{T} j^{-3/2} |h_{t,j} - \bar h_{t,j}|$ do not depend on H. The first claim follow applying to the r.h.s. of

$$|\bar e_{\beta,tH} - e_{\Delta\beta,tH}| \le \sum_{j=1}^{T} |\bar v_{tj;H} - v^{\Delta}_{tj;H}|\, |h_{t,j}| + \sum_{j=[LH]+1}^{t-1} v^{\Delta}_{tj;H} |h_{t,j}| + \sum_{j=1}^{[LH]} v^{\Delta}_{tj;H} |h_{t,j} - \bar h_{t,j}|$$

the bound (A.47) and (A.46),

$$\begin{aligned} &\bar v_{tj;H} \le Cj^{-3/2}, \quad 1 \le j \le t; \qquad \bar v_{tj;H} \le CL^{-5} j^{-3/2}, \quad [LH] \le j \le t,\ L \ge 1; \quad (A.17) \\ &|\bar v_{tj;H} - \bar v_{tj;H+\theta}| \le C|\Delta| j^{-3/2}, \quad 1 \le j \le t, \quad \theta \in [0, \Delta], \end{aligned}$$

which also imply the second claim of (A.16).

Hence, by (A.16) and equality $a^2 - b^2 = (a-b)(a+b)$,

$$\begin{aligned} |\bar e^2_{\beta,tH} - e^2_{\Delta\beta,tH}| &\le \big(\{\Delta + L^{-5}\}S_T^2 + S_T S_T^*\big), \qquad (A.18) \\ |H^{-1} Q_{\beta,T}(H) - Q_{\Delta\beta,T}(H)| &\le T^{-1} \sum_{t=1}^{T} |\bar e^2_{\beta,tH} - e^2_{\Delta\beta,tH}| \\ &= C(\{\Delta + L^{-5}\}S_T^2 + S_T S_T^*). \end{aligned}$$

By the bound $E|h_{t,j} h_{t,k}| \le C(jk)^{1/2}$,

$$ES_T^2 \le C \sum_{j,k=1}^{T} (jk)^{-3/2} E[|h_{t,j} h_{t,k}|] \le C \log^2 T.$$

Similarly, by definition of $\bar h_{t,j}$, using $E(\eta_j - \bar\eta_j)^2 = E[\eta_j^2 I(|\eta_j| \ge \log^4 T)] \le C(\log T)^{-4} E[\eta_j^4]$, then

$$E(h_{t,j}-\bar h_{t,j})^2 \;\le\; Cj\{(\sum_{k=m}^{\infty}|b_k|)^2+E(\eta_j-\bar\eta_j)^2\} \qquad (A.19)$$
$$\le\; Cj\{(\sum_{k=m}^{\infty}k^{-1-v})^2+(\log T)^{-8}\}\le Cj(\log T)^{-8}.$$

Therefore, with $L=\log T$ and $\Delta=(\log T)^{-4}$,

$$ES_T^{*\,2}\le C(\log T)^{-8}\big(\sum_{j=1}^{T}j^{-1}\big)^2\le C(\log T)^{-6},$$
$$E[(\Delta+L^{-5}\}S_T^2+S_TS_T^*]\le C[(\Delta+L^{-5}\}ES_T^2+(ES_T^2ES_T^{*\,2})^{-1}]\le C(\log T)^{-2}=o(1),$$

which together with (A.18) proves (A.13).

Proof of (A.14). It suffices to show that, as $T\to\infty$,

$$\max_{i=1,\cdots,N}|Q_{\Delta\beta,T}(H_i)-EQ_{\Delta\beta,T}(H_i)|=o_P(1). \qquad (A.20)$$

Notice that $Q_{\Delta\beta,T}(H_i)=T^{-1}\sum_{t=1}^{T}e^2_{\Delta\beta,tH_i}$ is the sum of m_i-dependent r.v.'s with $m_i\le H(\log T)^p\le CT^{1-\delta/2}$ ($H\le T^{1-\delta}$), and $N\le CT^{1-\delta}\Delta^{-1}\le CT$. Thus, by Lemma 2, (A.20) holds if

$$\max_{t,i}E(e^2_{\Delta\beta,tH_i})^{2k}\le C(\log T)^{8k}. \qquad (A.21)$$

To show the latter, firstly observe, that is $\bar h_{t,j}$ the sum of linear variables with i.i.d. innovations $\bar\eta_j$, and therefore its moments satisfy

$$E\bar h_{t,j}^{2k}\le C\frac{E[\bar\eta_j^{2k}]}{(E\bar\eta_j^2)^k}(E[\bar h_{t,j}^2])^k,\quad k\ge 1,$$

see, e.g., Proposition 4.4.3 in Giraitis, Koul, and Surgailis (2012). Since $E[\bar\eta_j^{2k}]\le(\log T)^{8k}$, $E\bar\eta_j^2\to E\eta_j^2>0$ and $E\bar h_{t,j}^2\le Eh_{t,j}^2\le Cj$, we conclude that

$$E\bar h_{t,j}^{2k}\le Cj^k(\log T)^{8k},\quad 1\le j\le t-1. \qquad (A.22)$$

By (A.22) and Cauchy inequality,

$$\sum_{j=1}^{t-1}|\bar v_{tj;H}|\le H^{-1/2}, \qquad (A.23)$$
$$E[e^{4k}_{\Delta\beta,tH}]\equiv E\Big(\sum_{j=1}^{[HL]}\bar v_{tj;H}\bar h_{t,j}\Big)^{4k}\le\Big(\sum_{j=1}^{t-1}\bar v_{tj;H}\Big)^{4k}\max_{1\le j\le LH}E\bar h_{t,j}^{2k}$$
$$\le H^{-2k}\{(LH)^{2k}(\log T)^{8k}\}\le C(\log T)^{10k},$$

which implies (A.21) and completes the proof of the lemma. □

In the next lemma we consider the case of a bounded deterministic trend β_j (b5) and the break in the mean (b6), while the corresponding results for the deterministic trend (b3) follow straightforwardly from (b5).

In case (b5), recall $V_{tr,H} := \sum_{j=1}^{\infty} v_{j,H}(j/H)$, $\kappa(g) = \int_0^1 \dot{g}(x)^2 dx$. By (A.51), $V^2_{tr,H} \to \nu_{1,K}$ as $H \to \infty$.

Lemma A.2 (i) *Let β_t be as in (b5). Then, $T \to \infty$,*

$$\sup_{H\in I_T}\Big|(\frac{T}{H})^2 Q_{\beta,T} - \kappa(g)V^2_{tr,H}\Big| \to 0, \qquad \sup_{H\in I_T}\Big|(\frac{T}{H})^2 e_{\beta,T}(H) - \dot{g}(1)^2 V^2_{tr,H}\Big| \to 0. \tag{A.24}$$

(ii) *Let β_t be as in (b6). Then (2.26) holds.*

Proof. (i) For simplicity, denote $g((t-j)/T) = g_{(t-j)/T}$. Then $e_{\beta,tH} = \sum_{j=1}^{t-1} v_{tj;H}(g_{(t-j)/T} - g_{t/T})$. Let $\bar{e}_{\beta,tH} = -(H/T)\dot{g}_{t/T} v_H$. By Taylor expansion,

$$\begin{aligned} |e_{\beta,tH} - \bar{e}_{\beta,tH}| &\le \sum_{j=1}^{t-1}\{|v_{tj;H} - v_{j;H}||g_{(t-j)/T} - g_{t/T}| + v_{j;H}\big|g_{(t-j)/T} - g_{t/T} + (H/T)\dot{g}_{t/T}\big|\} \\ &\le C(H/T)\sum_{j=1}^{t-1}\{|v_{tj;H} - v_{j;H}|(j/H) + v_{j;H}(j/H)^2(H/T)\}. \end{aligned}$$

For $H \in I_T$, $(H/T) \le T^{-\delta}$, by (A.45) $\sum_{j=1}^{t-1}|v_{tj;H} - v_{j;H}|(j/H) \le CH(t \vee H)^{-1}$ and by (A.51), $\sum_{j=1}^{t-1} v_{j;H}(j/H)^2 \le C$. So,

$$\begin{aligned} (T/H)|e_{\beta,tH} - \bar{e}_{\beta,tH}| &\le C\{H(t\vee H)^{-1} + T^{-\delta}\}, \\ (T/H)(|e_{\beta,tH}| + |\bar{e}_{\beta,tH}|) &\le C\sum_{j=1}^{t-1}(v_{tj;H}(j/H) + 1) \le CH(t\wedge H)^{-1}, \\ (T/H)^2|e^2_{\beta,tH} - \bar{e}^2_{\beta,tH}| &\le C\{H(t\vee H)^{-1} + T^{-\delta}\}H(t\wedge H)^{-1} \le C(Ht^{-1} + T^{-\delta}). \end{aligned} \tag{A.25}$$

Notice that

$$\begin{aligned} (T/H)^2 T^{-1}\sum_{t=1}^{T}\bar{e}^2_{\beta,tH} &= T^{-1}\sum_{t=1}^{T}\dot{g}(t/T)^2 V^2_{tr,H} = \kappa(g)V^2_{tr,H} + o(1), \\ (T/H)^2 T^{-1}\sum_{t=1}^{T}|e^2_{\beta,tH} - \bar{e}^2_{\beta,tH}| &\le CT^{-1}\sum_{t=1}^{T}(Ht^{-1} + T^{-\delta}) \le CT^{-\delta} \to 0, \end{aligned}$$

which proves the first claim of (A.24). The second claim follows from (A.25):

$$\begin{aligned} \Big|(\frac{T}{H})^2 e_{\beta,T}(H) - \dot{g}(1)^2 V^2_{tr,H}\Big| &= (T/H)^2|e^2_{\beta,(T+1)H} - \bar{e}^2_{\beta,(T+1)H}| \\ &\le C(HT^{-1} + T^{-\delta}) \le CT^{-\delta} \to 0, \end{aligned}$$

which completed the proof of the lemma.

(ii) In case of the break in the mean β_t of (b6), (2.26) follows by standard calculus. $\square$

A.3 Properties of $Q_{u,T}(H)$

Lemma A.3 *Let $y_t = u_t$ be a stationary process. Under Assumptions of Theorem 1, as $T \to \infty$,*

$$(i) \qquad \sup_{H \in I_T} H\Big|E[Q_{u,T}(H)] - \sigma_u^2 - Q_u(H)\Big| \;\to\; 0, \tag{A.26}$$

$$\sup_{H \in I_T} H|e_T(H) - \sigma_u^2 - Q_u(H)| \;\to\; 0, \tag{A.27}$$

where $Q_u(H) = H^{-1}b_{u,K} + o(H^{-1})$ as $H \to \infty$.

$$(ii) \qquad \sup_{H \in I_T} H\Big|Q_{u,T}(H) - \hat{\sigma}_u^2 - E[Q_{u,T}(H) - \hat{\sigma}_u^2]\Big| \;\to_p\; 0, \tag{A.28}$$

Proof. (i) Because of stationarity of $\{u_j\}$,

$$\begin{aligned} E[Q_{u,T}(H) - \hat{\sigma}_u^2] &= T^{-1}\sum_{t=1}^{T}\{E(\hat{y}_{t|t-1} - y_t)^2 - \sigma_u^2\}, \qquad (A.29)\\ E(\hat{y}_{t|t-1} - y_t)^2 - \sigma_u^2 &= \sum_{j,k=1}^{\infty} v_{tj;H}v_{tk;H}\gamma_u(j-k) - 2\sum_{j=1}^{\infty} v_{tj;H}\gamma_u(j),\\ Q_u(H) &= \sum_{j,k=1}^{\infty} v_{j;H}v_{k;H}\gamma_u(j-k) - 2\sum_{j=1}^{\infty} v_{j;H}\gamma_u(j). \end{aligned}$$

Below we show that

$$|\{E(\hat{y}_{t|t-1} - y_t)^2 - \sigma_u^2\} - Q_u(H)| \le Ct^{-1}, \qquad t \ge 2, \tag{A.30}$$

where C does not depend on t and $H \in I_T$. Then,

$$\begin{aligned} H\Big|E[Q_{u,T}(H) - \hat{\sigma}_u^2] - Q_u(H)\Big| &\le CHT^{-1}\sum_{t=1}^{T} t^{-1}\\ &\le CHT^{-1}\log T \le T^{-\delta/2}, \end{aligned}$$

which proves (A.26). In addition, by (A.30)

$$H|e_T(H) - \sigma_u^2 - Q_u(H)| \le C(H/T) \le CT^{-\delta},$$

which implies (A.27). The last claim of the lemma about $Q_u(H)$ is shown in (A.52).

(ii) Proof of (A.28). We shall use the following notation:

$$\begin{aligned} h_{t,j} &:= \sum_{l=1}^{j} u_{t-j}, \quad j=1,\cdots,t-2; \qquad h_{t,t-1} := \sum_{l=1}^{t-1} u_{t-j}, \\ \bar{v}_{tj;H} &:= H^{1/2}(v_{tj;H} - v_{t,j+1;H}), \quad j=1,\cdots,t-2, \qquad \bar{v}_{tj;H} := H^{1/2} v_{t,t-1;H}. \end{aligned}$$

Using summation by parts, write

$$\bar{e}_{u,tH} := H^{1/2} \sum_{j=1}^{t-1} v_{tj;H} u_{t-j} = \sum_{j=1}^{t-1} \bar{v}_{tj;H} h_{t,j}. \tag{A.31}$$

Then,

$$\begin{aligned} H(Q_{u,T}(H) - \hat{\sigma}_u^2) &= T^{-1} \sum_{t=1}^{T} (\bar{e}_{u,tH} - H^{1/2} u_t)^2 - H\hat{\sigma}_u^2 \qquad (A.32) \\ &= T^{-1} \sum_{t=1}^{T} \bar{e}_{u,tH}^2 - 2u_t H^{1/2} T^{-1} \sum_{t=1}^{T} \bar{e}_{u,tH} =: Q_{u,T}^{(1)}(H) - 2Q_{u,T}^{(2)}(H). \end{aligned}$$

To prove (A.28) it suffices to show that

$$\sup_{H \in I_T} \left| Q_{u,T}^{(i)}(H) - E[Q_{u,T}^{(i)}(H)] \right| = o_p(1), \qquad i = 1, 2. \tag{A.33}$$

Case $i = 1$. The proof follows by the same argument as the proof of Lemma A.7. By (A.49) and (A.48), the weights $\bar{v}_{tj;H}$ have properties (A.17) and (A.23). Moreover, h_{tj} is a sum of linear variables satisfying $Eh_{tj}^2 \le Cj$. Therefore, similarly as in (A.15), defining $v_{utj;H}^{\Delta} := \bar{v}_{tj;H_i}$, $H \in [H_i,\, H_i + \Delta)$ and $e_{u\Delta,tH} := \sum_{j=1}^{[LH]} v_{utj;H}^{\Delta} \bar{h}_{t,j}$, we obtain same bounds as

$$\begin{aligned} |\bar{e}_{u,tH} - e_{u\Delta,tH}| &\le C[\{\Delta + L^{-5}\} S_T + S_T^*], \qquad (A.34) \\ |\bar{e}_{u,tH}| + |e_{u\Delta,tH}| &\le C S_T, \end{aligned}$$

and (A.33) follows by the argument as in the proof of (A.9).

Consider the case $i = 2$. To handle the addition factor $H^{1/2}$, we modify definition (A.15) as follows:

$$\begin{aligned} \tilde{v}_{tj;H}^{\Delta} &:= H_i^{1/2} \bar{v}_{tj;H_i}, \qquad H \in [H_i,\, H_i + \Delta),\ i = 0, \cdots, N; \qquad (A.35) \\ \tilde{e}_{\Delta u,tH} &:= \sum_{j=1}^{t-1} \tilde{v}_{tj;H}^{\Delta} u_{t-j}, \qquad Q_{\Delta u,T}^{(2)}(H) := T^{-1} \sum_{t=1}^{T} u_t \tilde{e}_{\Delta u,tH}. \end{aligned}$$

The bound (A.48) with $\gamma = 1$ yields

$$|H^{1/2}\bar{v}_{tj;H} - v^{\Delta}_{tj;H}| \le C\Delta j^{-2}, \quad |H^{1/2}\bar{e}_{u,tH} - \bar{e}_{\Delta u,tH}| \le C\Delta j^{-2}. \tag{A.36}$$

Recall that $\Delta = (\log T)^{-4}$. Hence,

$$\big|Q^{(2)}_{u,T}(H) - \tilde{Q}^{(2)}_{\Delta u,T}(H)\big| \le T^{-1}\sum_{t=1}^{T}|u_t\{H^{1/2}\bar{e}_{u,tH} - \tilde{e}_{\Delta u,tH}\}| \le C\Delta T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} j^{-2}|h_{t,s}u_t|,$$

where the r.h.s. does not depend on H. Since $E|h_{t,s}u_t| \le (Eh^2_{t,s}Eu^2_t)^{1/2} \le Cj^{1/2}$, then

$$E\big[\sup_{H\in I_T}\big|\bar{Q}^{(2)}_{u,T}(H) - \tilde{Q}^{(2)}_{\Delta u,T}(H)\big|\big] \le C\Delta T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{\infty} j^{-3/2} \le C\Delta \to 0.$$

Since $\tilde{Q}^{(2)}_{\Delta u,T}(H)$ is a step function in H, it remains to show that

$$\max_{i=1,\cdots,N}|\tilde{Q}^{(2)}_{\Delta u,T}(H_i) - E[\tilde{Q}^{(2)}_{\Delta u,T}(H_i)]| = o_P(1).$$

As in (A.58), it suffices to verify, that there exist $k \ge 1$ and $\gamma > 1$, such that

$$\max_{i=1,\cdots,N} E|q_{iT}|^{2k} \le CT^{-\gamma}, \quad q_{iT} := \tilde{Q}^{(2)}_{\Delta u,T}(H_i) - E[\tilde{Q}^{(2)}_{\Delta u,T}(H_i)]. \tag{A.37}$$

To prove (A.37), let $H = H_i$. Notice that

$$\tilde{Q}^{(2)}_{\Delta u,T}(H_i) \;=\; H_iT^{-1}\sum_{t=1}^{T}u_t\sum_{j=1}^{t-1}v_{tj;H}u_{t-j} = HT^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}w_{t,j}u_tu_{t-j}$$

is a centered quadratic form with weights $w_{t,j}$. Set $w_{t,j} = 0$, $j = t, \cdots, T$.

The $2k$-th moment, $k \ge 1$, of a general centered quadratic form $p_T := \sum_{t,j=1}^{T} b_{T,tj}u_tu_{t-j}$ of a linear process u_j (2.11) with i.i.d. innovations ε_j satisfies the bounds

$$E(p_T - E[p_T])^{2k} \le CA_T^k, \qquad A_T := \sum_{t,t',j,j'=1}^{T} b_{T,tj}b_{T,t'j'}\gamma_u(t-t')\gamma_u(j-j'), \tag{A.38}$$

as long as $E\varepsilon_1^{4k} < \infty$. For a linear process u_j, for $k = 1$ (A.38) follows from Lemma 4.5.1. in Giraitis, Koul, and Surgailis (2012), while its generalization for $k > 2$ is also straightforward.

Notice that for $b_{T,tj} = w_{tj;H}$,

$$A_T \le C(\log H + TH^{-1}) \tag{A.39}$$

because $w_{tj}^2 \le CK((t-j)/H)C_{tH}^{-2}$, $C_{tH} := \sum_{k=1}^{t-1} K(k/H)$, $\sum_{j\in\mathbb{Z}} |\gamma_u(j)| < \infty$, and $C_{tH} \le C(t\wedge H)^{-1}$ of (A.43) imply

$$\begin{aligned} A_T &\le \sum_{t,t',j,j'=1}^{T} w_{tj}^2 |\gamma_u(t-t')\gamma_u(j-j')| \le C\sum_{t=1}^{T} C_{tH}^{-2}\{\sum_{j=1}^{t-1} K(j/H)\}\{\sum_{j'\in\mathbb{Z}} |\gamma_u(j')|\}^2 \\ &\le C\sum_{t=1}^{T} C_{tH}^{-1} \le C(\log T + TH^{-1}). \end{aligned}$$

Consequently, from (A.38) and (A.39), for $H \le I_T$, $H/T \le T^{-\delta}$, we obtain

$$\begin{aligned} Eq_{iT}^{2k} &= (H/T)^{2k}(p_T - E[p_T])Ep_T^{2k} \le C(H/T)^{2k}A_T^k \\ &\le C(H/T)^{2k}(\log H + TH^{-1})^k \le CT^{-\delta k}, \end{aligned}$$

which implies (A.37) choosing $k > 1/d$. This completes the proof of the lemma. □

A.4 Properties of $Q_{\beta u,T}$

In the following lemma we set $d_\beta = 1$ for β_t and in (b2), (b4) and (b6), and $d_\beta = 2$ for β_t and in (b3) and (b5).

Lemma A.4 (i) *Under assumption of Theorems 2 and 3,*

$$\sup_{H\in I_T}\{H^{-d_\beta}\big|Q_{\beta u,T}(H)\big|\} \to_p 0, \quad \textit{as in (b2) and (b3)}; \tag{A.40}$$

$$\sup_{H\in I_T}\{[(\frac{T}{H})^{d_\beta}\wedge H]\big|Q_{\beta u,T}(H)\big|\} \to_p 0, \quad \textit{as in (b4), (b5) and (b6)}. \tag{A.41}$$

Proof. Consider the case when β_t is a unit root process (b2). Then $E[Q_{\beta u,T}(H)] = 0$, and using notation (A.2), $\bar{e}_{\beta,tH}$ of (A.12), $\bar{e}_{u,tH}$ of (A.31), we can write

$$\begin{aligned} e_{\beta,tH}e_{u,tH} &= e_{\beta,tH}\sum_{j=1}^{t-1} v_{tj;H}(u_{t-j}-u_t) = \bar{e}_{\beta,tH}\bar{e}_{u,tH} - e_{\beta,tH}u_t, \\ Q_{\beta u,T}(H) &= T^{-1}\sum_{t=1}^{T} e_{\beta,tH}e_{u,tH} = T^{-1}\sum_{t=1}^{T}\bar{e}_{\beta,tH}\bar{e}_{u,tH} - (HT)^{-1}\sum_{t=1}^{T} H^{1/2}\bar{e}_{\beta,tH}u_t \\ &=: q_{T,H}^{(i)} - H^{-1}q_{T,H}^{(2)}. \end{aligned}$$

We will verify that

$$\sup_{H\in I_T} |q_{T,H}^{(i)}| \to_p 0, \quad i = 1, 2. \tag{A.42}$$

The terms $\bar{e}_{\beta,tH}$ and $\bar{e}_{u,tH}$ are of the similar type, satisfy conditions (A.16) and (A.34) and $E[\bar{e}_{\beta,tH}\bar{e}_{u,tH}] = 0$. Therefore, the same argument as in the proof of (A.9) yields (A.42) for $i = 1$. For $i = 2$ the latter follows using (A.36) and the argument used in the proof of (A.9).

Clearly this implies (A.40) for the rescaled unit root process β_t, (b4).

The bound (A.41) for β_t (b5) and (b6) follows combining the approached used in the proofs of Lemmas A.1, A.4 and A.3. The case (b5) also implies (A.40) for (b3). $\square$

A.5 Auxiliary results

Denote $C_{tH} := \sum_{j=1}^{t-1} K(j/H), \quad t \geq 1$. Recall (A.1).

Lemma 1 *Under Assumption 1, uniformly in $H \in I_T$, $T \geq 1$, the following holds.*
(i) There exists $C > 0$ such that for all $t \geq 1$ and $j \geq 1$.

$$C_{tH}^{-1} \leq C(t \wedge H), \qquad v_{tj;H} \leq C(t \wedge H)^{-1}, \quad v_{j;H} \leq C(H \vee j)^{-1}, \tag{A.43}$$

$$|v_{tj;H} - v_{j;H}| \leq Ct^{-1}, \qquad j \geq 1, \tag{A.44}$$

$$\sum_{j=1}^{\infty} |v_{tj;H} - v_{j;H}|(j/H)^{\gamma} \leq CH(t \vee H)^{-1}, \qquad (0 \leq \gamma \leq 2). \tag{A.45}$$

(ii) The following holds uniformly in t, $H, H' \in I_T$ and $L \geq 1$.
(a) Let $0 \leq \gamma \leq 2$ and $\bar{v}_{tj;H} := H^{-\gamma} v_{tj;H}$, $1 \leq j \leq t-1$. Then

$$\bar{v}_{tj;H} \leq Cj^{-\gamma-1}, \quad \bar{v}_{tj;H} \leq CL^{-5} j^{-\gamma-1}, \tag{A.46}$$

$$|\bar{v}_{tj;H'} - \bar{v}_{tj;H}| \leq C|H' - H| j^{-\gamma-2}, \quad |H' - H| \leq 1. \tag{A.47}$$

(b) Let $0 \leq \gamma \leq 1$ and $\bar{v}_{tj;H} := H^{\gamma}(v_{tj;H} - v_{t,j+1;H})$, $1 \leq j \leq t-2$ and $\bar{v}_{t,t-1;H} := H^{\gamma} v_{t,t-1;H}$. Then

$$\bar{v}_{tj;H} \leq Cj^{-2+\gamma}, \quad \bar{v}_{tj;H} \leq CL^{-5} j^{-2+\gamma}, \tag{A.48}$$

$$|\bar{v}_{tj;H'} - \bar{v}_{tj;H}| \leq C|H' - H| j^{-3+\gamma}, \quad |H' - H| \leq 1. \tag{A.49}$$

(iii) As $H \to \infty$,

$$\sum_{j=1}^{\infty} v_{j;H}^2 = H^{-1} \int_R K^2(x)dx + o(H^{-1}), \qquad v_{0;H} = H^{-1}K(0) + o(H^{-1}); \tag{A.50}$$

$$\sum_{j=1}^{\infty} v_{j;H}(\frac{j}{H})^{\gamma} = \int_R K(x)x^{\gamma}dx + o(H^{-1}), \qquad 0 \leq \gamma \leq 2, \tag{A.51}$$

$$Q_u(H) = H^{-1}b_{u,K} + o(H^{-1}), \tag{A.52}$$

where $b_{u,K}$ is as in Theorem 1.

Proof (i) *Proof of (A.43).* Let $\epsilon > 0$ be a small number. Assume that $H > 1/\epsilon$. Then for $1 \le j \le \epsilon H$, $K(j/H) \ge \inf_{0\le u\le \epsilon} K(u) =: c_* > 0$ since $K(0) > 0$ and K is continuous at 0 by Assumption 1. Hence, for $t \ge 2$, $C_{tH} \ge c_* \min(t, \epsilon H) \ge c_*\epsilon(t \wedge H)$, for some $c_* > 0$ implying (A.43). Assume that $H \le 1/\epsilon$, then by assumption (2.4), $C_{tH} \ge K(1/H) \ge \inf_{0\le u\le 1/\alpha} K(u) \ge c(t \wedge H) > 0$, for some $c > 0$, which completes the proof of (A.43).

The second claim of (A.43) follows from the first because K is bounded, while the third claim follows for $j \ge H$ from $v_{tj;H} \le CH^{-1}K(j/H) \le Cj^{-1}$, because $K(x) \le Cx^{-1}$, and for $j \le H$ it holds $v_{tj;H} \le C(t \vee H)^{-1}$.

Proof of (A.44). For $t \le H$, $|v_{tj;H} - v_{j;H}| \le Ct^{-1}$ by (A.43).

Let $t \ge H$. First notice that by (A.43) and $K(x) \le Cx^2$,

$$\sum_{j=t}^{\infty} v_{j;H} \le H^{-1}\sum_{t=t}^{\infty}(H/j)^2 \le CHt^{-1}. \tag{A.53}$$

This and (A.43) imply

$$|v_{tj;H} - v_{j;H}| = v_{tj;H}\sum_{j=t}^{\infty} v_{j;H} \le CH^{-1}\{Ht^{-1}\} \le Ct^{-1} \tag{A.54}$$

completing the proof.

Proof of (A.45). Write

$$\sum_{j=1}^{\infty} |v_{tj;H} - v_{j;H}|(\frac{j}{H})^{\gamma} = \sum_{j=1}^{H}[\cdots] + \sum_{j=H+1}^{\infty}[\cdots] =: s_{1H} + s_{2H}.$$

Then $s_{1H} \le \sum_{j=1}^{H}(v_{tj;H} + v_{j;H}) \le 1$, whereas by (A.53) and (A.54),

$$\begin{aligned} s_{2,H} &\le \sum_{j=1}^{\infty} v_{tj;H}(\frac{j}{H})^{\gamma}\{\sum_{j=t}^{\infty} v_{j;H}\} \le \frac{C_{\infty H}}{C_{tH}}\{\sum_{j=1}^{\infty} v_{j;H}(j/H)^{\gamma}\}(H\wedge t)t^{-1} \\ &\le C\frac{H}{t\wedge H}\{C\}\frac{H\wedge t}{t} = CHt^{-1}, \end{aligned}$$

in view of (A.43) and (A.51), completing the proof.

(ii) (a) *Proof of (A.46).* Let $t < H$, $j < t$. Then similarly as above $v_{tj;H}H^{-\gamma} \le CK(j/H)t^{-1}H^{-\gamma} \le Cj^{-1-\gamma}$. Next let $j \le H$. Then $t \ge H$ and using $K(x) \le x^{-\gamma-1}$, $v_{tj;H}H^{-\gamma} \le CK(j/H)H^{-1-\gamma} \le j^{-1-\gamma}$. The proves the first claim of (A.46), while the second claim follows using $K(x) \le Cx^{-\gamma-5}$, which for $t \ge H$ and $j/H \le L$ gives $v_{tj;H}H^{-\gamma} \le CK(j/H)H^{-1-\gamma} \le j^{-1-\gamma}L^{-5}$.

Proof of (A.47) By the mean value theorem,

$$H'^{-\gamma} v_{tj;H'} - H^{-\gamma} v_{tj;H} = \frac{K(j/H')}{H'^{\gamma} C_{tH'}} - \frac{K(j/H)}{H'^{\gamma} C_{tH}} = (H' - H) g(\tilde{H}), \quad \tilde{H} \in [H,\ H'],$$

$$g(H) = \frac{d}{dH}\{\frac{K(j/H)}{H^{\gamma} C_{tH}}\} = \frac{\dot{K}(j/H)(j/H)H^{-1}}{H^{\gamma} C_{tH}} - K(j/H)\{\frac{1/2}{H^{\gamma+1} C_{tH}} + \frac{(d/dH)C_{tH}}{H^{\gamma} C_{tH}^2}\}.$$

Let $t \geq H$. Then by (A.43) $C_{tH} \geq cH$, and by (2.4), $|\dot{K}(x)x| + K(x) \leq cx^{-3/2}$. So,

$$g(H) \leq CH^{-\gamma-2}(|\dot{K}(j/H)(j/H)| + K(j/H)) \leq Cj^{-\gamma-2}.$$

Let $t < H$. Then by (A.43) $C_{tH} \geq ct \geq cj$, $H \geq j$, and by (2.4), we can bound $|\dot{K}(x)x| + K(x) \leq C$. So, $g(H) \leq Cj^{-\gamma-2}$, which completes the proof of (A.47).

(b) To show (A.47) and (A.46) for the weights (b) $\bar{v}_{tj;H} := H^{\gamma}(v_{tj;H} - v_{t,j+1;H})$, $1 \leq j \leq t-2$, apply the mean value theorem to obtain $\bar{v}_{tj;H} := H^{-(1-\gamma)}(\dot{K}(\tilde{j}/H)/C_{tH})$, $\tilde{j} \in [j, j+1]$ and use the same argument as for the weights (a). Proof of (A.47) and (A.46) for $\bar{v}_{t,t-1;H}$ follows similarly as above.

(iii) *Proof of (A.50) and (A.51).* Under (2.4), these claims follow using standard argument of approximation of a sum by an integral.

Proof of (A.52). Let $\theta_{s;H} := \sum_{k=1}^{\infty} v_{k,H} v_{k+|s|,H}$ and $f_{s;H} := \theta_{|s|;H} - \theta_{0;H} - v_{|s|,H} + v_{0,H}$, $s \in \mathbb{Z}$. Then

$$\begin{aligned} Q_u(H) &= \sum_{j,k=1}^{\infty} v_{j,H} v_{k,H} \gamma_u(j-k) - 2\sum_{j=1}^{\infty} v_{j,H}\gamma_u(j) \\ &= \sum_{s=-\infty}^{\infty} (\theta_{|s|;H} - v_{s,H})\gamma_u(s) + v_{0,H}\gamma_u(0) \\ &= \Big\{(\theta_{0;H} - v_{0,H}) \sum_{js=-\infty}^{\infty} \gamma_u(s) + v_{0,H}\gamma_u(0)\Big\} + \sum_{js=-\infty}^{\infty} f_{s;H}\gamma_u(s) =: i_H + r_H. \end{aligned}$$

By (A.50), $i_H = H^{-1}s_u^2(K_2 - K_0) + H^{-1}K_0 + o(H^{-1}) = H^{-1}b_{u,K} + o(H^{-1})$. To complete the proof, it suffices to show that

$$r_H = o(H^{-1}). \tag{A.55}$$

Notice that by (A.43),

$$H|v_{k+|s|,H} - v_{k,H}| \leq C|K(\frac{k+|s|}{H}) - K(\frac{s}{H})| \leq C\sup_x K(x), \qquad \text{for all } s,$$

$$\leq C\sup_x |\dot{K}(x)|\frac{|s|}{H} \to 0, \quad \forall s \text{ fixed}.$$

Hence,

$$\max_s H|f_{s;H}| \leq \sum_{k=1}^{\infty} v_{k,H} H|v_{k+|s|,H} - v_{|s|,H}| + H|v_{k,H} - v_{0,H}| \leq C \sum_{k=1}^{\infty} v_{k,H} < \infty,$$
$$H|f_{s;H}| \leq C|s|H^{-1} \to 0, \quad \forall s \text{ fixed}.$$

Since $\sum_{k=-\infty}^{\infty} |\gamma_u(s)| < \infty$, this by dominated convergence theorem implies (A.55). □

Lemma 2 *Let* $S_{Ti} = T^{-1} \sum_{t=1}^{T} z_{Tt,i}$, $i = 1, \cdots, N$, $N \leq T$, *be the sums of arrays of zero mean* m_i*-dependent r.v.'s* $z_{Tt,i}$ *such and* $\max_i m_i \leq T^{1-\delta}$, $0 < \delta < 1$. *If for some integer* $k \geq (1/2) + 1/(2\delta)$,

$$\max_{t,i} E[z_{Tt,j}^{2k}] \leq C(\log T)^p, \qquad (\exists p \geq 1), \tag{A.56}$$

then as $T \to \infty$, $\max_{i=1,\cdots,N_T} |S_{Ti}| = o_p(1)$.

Proof. It suffices to show that, for some $k > 1$ and $\gamma > 1$,

$$\max_{i=1,\cdots N} ES_{Ti}^{2k} \leq CT^{-\gamma}, \tag{A.57}$$

since then for any $a > 0$,

$$P(\max_{i=1,\cdots,N} |S_{Ti}| \geq a) \leq \sum_{i=1}^{N} P(|S_{Ti}| \geq a) \leq a^{-2k} \sum_{i=1}^{N} E[S_{Ti}^{2k}] \leq CNT^{-\gamma} \to 0. \tag{A.58}$$

By assumption, $\gamma^* = (2k-1)\delta > 0$. Because r.v.'s $z_{Tt,i}$ are m_i-dependent, then

$$\begin{aligned} E[S_{Ti}^{2k}] &\leq CT^{-2k} \sum_{1 \leq t_{2k} \leq \cdots \leq t_{2k} \leq T:\, t_1 - t_{2k} \leq 2km_i} E\Big[\prod_{s=1}^{2k} (z_{Tt,i} - z_{Tt,i})\Big] \\ &\leq C\{\max_{t=1,\cdots,T} E[z_{Tt,i}^{2k}]\}(m_i/T)^{2k-1} \leq C(\log T)^p T^{-\delta(2k-1)} \leq C(\log T)^p T^{-\gamma*}, \end{aligned}$$

which proves (A.57). □

Table 1: Monte Carlo Results. $T = 200$. One-Step Ahead Forecasts. Table reports relative root mean square error using a full sample mean benchmark

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	1.045	0.700	0.168	0.773	0.805	0.337	0.985	0.826	0.674	0.696	0.170
Rolling	$H = \hat{H}$	1.134	0.745	0.203	0.826	0.866	0.373	1.041	0.877	0.756	0.726	0.334
Rolling	$H = 20$	1.047	0.667	0.211	0.755	0.767	0.342	0.940	0.797	0.670	0.667	0.291
	$H = 30$	1.028	0.666	0.272	0.764	0.775	0.384	0.940	0.820	0.693	0.664	0.361
Exponential	$\rho = 0.99$	1.002	0.836	0.754	0.896	0.909	0.765	0.990	0.973	0.865	0.835	0.738
	$\rho = 0.95$	1.020	0.671	0.301	0.757	0.779	0.407	0.940	0.829	0.681	0.668	0.339
	$\rho = 0.90$	1.048	0.672	0.194	0.742	0.769	0.333	0.941	0.793	0.649	0.667	0.231
	$\rho = 0.80$	1.103	0.705	0.164	0.763	0.803	0.325	0.984	0.815	0.658	0.698	0.181
	$\rho = 0.70$	1.169	0.749	0.163	0.802	0.851	0.338	1.042	0.861	0.688	0.739	0.167
	$\rho = 0.50$	1.317	0.846	0.178	0.897	0.961	0.378	1.174	0.970	0.769	0.833	0.166
Averaging		1.005	0.754	0.644	0.844	0.858	0.630	0.989	0.966	0.799	0.753	0.610
Nonparametric		1.108	0.686	0.166	0.759	0.786	0.321	0.966	0.798	0.661	0.683	0.210
Polynomial	$\alpha = \hat{\alpha}$	1.010	0.773	0.444	0.941	0.917	0.555	1.005	0.952	0.789	0.767	0.354
Rolling	$H = \hat{H}$, $k = \hat{k}$	1.145	0.780	0.210	0.853	0.900	0.436	1.051	0.891	0.780	0.747	0.281

Table 2: Monte Carlo Results. $T = 200$. One-Step Ahead Forecasts. $u_t \sim AR(0.7)$. Table reports relative root mean square error using a full sample mean benchmark

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.660	0.394	0.087	0.631	0.466	0.188	0.582	0.483	0.410	0.407	0.121
Rolling	$H = \hat{H}$	1.016	0.660	0.132	0.863	0.620	0.282	0.788	0.666	0.561	0.568	0.141
Rolling	$H = 20$	1.028	0.645	0.204	1.013	0.772	0.329	0.890	0.768	0.692	0.648	0.332
	$H = 30$	1.020	0.655	0.264	1.029	0.788	0.371	0.916	0.803	0.721	0.656	0.407
Exponential	$\rho = 0.99$	0.989	0.816	0.745	0.984	0.900	0.758	0.971	0.959	0.867	0.825	0.759
	$\rho = 0.95$	0.938	0.598	0.282	0.924	0.723	0.376	0.845	0.753	0.656	0.606	0.372
	$\rho = 0.90$	0.870	0.536	0.166	0.851	0.644	0.278	0.760	0.646	0.573	0.543	0.248
	$\rho = 0.80$	0.781	0.473	0.119	0.758	0.567	0.232	0.678	0.566	0.497	0.481	0.178
	$\rho = 0.70$	0.715	0.429	0.100	0.691	0.514	0.208	0.624	0.519	0.449	0.439	0.146
	$\rho = 0.50$	0.639	0.384	0.085	0.619	0.456	0.185	0.566	0.469	0.400	0.395	0.120
Averaging		0.996	0.730	0.632	0.989	0.852	0.623	0.967	0.950	0.808	0.743	0.649
Nonparametric		1.019	0.585	0.138	0.984	0.717	0.274	0.854	0.682	0.624	0.590	0.233
Polynomial	$\alpha = \hat{\alpha}$	0.713	0.472	0.176	0.743	0.559	0.352	0.635	0.529	0.545	0.448	0.205
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.941	0.597	0.117	0.718	0.509	0.228	0.658	0.545	0.448	0.467	0.107

Table 3: Monte Carlo Results. $T = 200$. Two-Step Ahead Forecasts. Table reports relative root mean square error using a full sample mean benchmark

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	1.046	0.691	0.169	0.760	0.796	0.335	0.975	0.824	0.663	0.701	0.191
Rolling	$H = \hat{H}$	1.069	0.702	0.202	0.822	0.807	0.344	0.970	0.913	0.685	0.716	0.175
Rolling	$H = 20$	1.039	0.662	0.219	0.741	0.766	0.341	0.936	0.797	0.664	0.674	0.293
	$H = 30$	1.026	0.660	0.283	0.753	0.774	0.384	0.936	0.823	0.687	0.672	0.360
Exponential	$\rho = 0.99$	1.001	0.836	0.759	0.893	0.911	0.767	0.990	0.975	0.866	0.843	0.740
	$\rho = 0.95$	1.017	0.666	0.312	0.746	0.779	0.409	0.938	0.833	0.678	0.678	0.347
	$\rho = 0.90$	1.046	0.664	0.202	0.730	0.765	0.333	0.938	0.792	0.643	0.675	0.242
	$\rho = 0.80$	1.102	0.696	0.168	0.753	0.795	0.324	0.980	0.811	0.651	0.706	0.199
	$\rho = 0.70$	1.169	0.737	0.165	0.792	0.842	0.337	1.038	0.853	0.680	0.747	0.188
	$\rho = 0.50$	1.318	0.830	0.177	0.889	0.950	0.377	1.171	0.954	0.757	0.841	0.193
Averaging		1.004	0.753	0.650	0.839	0.860	0.633	0.987	0.968	0.799	0.763	0.611
Nonparametric		1.104	0.678	0.171	0.745	0.780	0.319	0.960	0.797	0.653	0.691	0.219
Polynomial	$\alpha = \hat{\alpha}$	1.005	0.776	0.401	0.960	0.932	0.567	0.998	0.958	0.780	0.789	0.353
Rolling	$H = \hat{H}$, $k = \hat{k}$	1.062	0.706	0.195	0.810	0.819	0.349	0.952	0.898	0.674	0.721	0.154

Table 4: Empirical relative root mean square error results for the UK.

		Second subsample							First subsample						
Method		Median	Min	Max	Var	Skew	DM1	DM2	Median	Min	Max	Var	Skew	DM1	DM2
Exponential	$\rho = \hat{\rho}$	0.858	0.006	1.280	0.309	-1.233	2	21	0.803	0.012	1.445	0.373	-0.552	4	29
Rolling	$H = \hat{H}$	0.886	0.006	1.503	0.300	-1.207	2	19	0.899	0.010	1.592	0.300	-0.939	7	26
Rolling	$H = 20$	0.887	0.005	1.518	0.309	-1.063	4	18	0.927	0.009	1.519	0.314	-0.955	12	22
	$H = 30$	0.903	0.006	1.845	0.312	-0.821	6	19	0.897	0.010	1.329	0.268	-1.193	10	24
Exponential	$\rho = 0.99$	0.927	0.462	1.060	0.127	-1.703	3	26	0.946	0.699	1.019	0.077	-1.315	0	40
	$\rho = 0.95$	0.858	0.007	1.252	0.270	-1.437	5	22	0.839	0.100	1.111	0.239	-1.145	2	37
	$\rho = 0.90$	0.858	0.005	1.254	0.299	-1.233	6	20	0.812	0.012	1.222	0.304	-0.909	4	33
	$\rho = 0.80$	0.884	0.005	1.273	0.327	-1.078	9	21	0.817	0.010	1.358	0.355	-0.673	7	31
	$\rho = 0.70$	0.929	0.006	1.409	0.360	-0.907	12	20	0.841	0.011	1.468	0.398	-0.516	10	29
	$\rho = 0.50$	1.047	0.007	1.755	0.438	-0.623	22	19	0.927	0.013	1.716	0.486	-0.324	13	27
Averaging		0.883	0.069	1.203	0.235	-1.625	3	22	0.884	0.258	1.180	0.193	-1.321	2	33
Nonparametric		0.926	0.034	1.699	0.351	-0.870	8	20	0.899	0.038	1.591	0.380	-0.765	7	21
Polynomial	$\alpha = \hat{\alpha}$	0.863	0.011	1.263	0.266	-1.203	0	22	0.817	0.017	1.365	0.330	-0.821	0	26
Rolling	$H = \hat{H}, k = \hat{k}$	0.860	0.005	1.292	0.292	-1.266	1	22	0.821	0.010	1.158	0.266	-1.202	2	30

Table 5: Empirical relative root mean square error results for the US.

		Second subsample							First subsample						
Method		Median	Min	Max	Var	Skew	DM1	DM2	Median	Min	Max	Var	Skew	DM1	DM2
Exponential	$\rho = \hat{\rho}$	0.639	0.020	1.246	0.382	-0.178	0	37	0.647	0.007	1.288	0.387	-0.212	1	39
Rolling	$H = \hat{H}$	0.883	0.076	1.606	0.299	-0.763	14	24	0.900	0.136	1.918	0.362	0.307	6	30
Rolling	$H = 20$	0.869	0.078	1.635	0.317	-0.665	11	23	0.997	0.118	2.744	0.506	1.224	10	27
	$H = 30$	0.857	0.130	1.623	0.293	-0.837	6	30	0.958	0.158	2.064	0.384	0.667	10	30
Exponential	$\rho = 0.99$	0.937	0.606	1.078	0.093	-1.297	9	42	1.023	0.672	2.060	0.191	2.895	24	21
	$\rho = 0.95$	0.796	0.196	1.350	0.266	-0.784	2	38	0.899	0.164	1.738	0.302	-0.083	11	30
	$\rho = 0.90$	0.745	0.066	1.319	0.295	-0.673	1	39	0.818	0.071	1.786	0.380	0.047	5	34
	$\rho = 0.80$	0.689	0.036	1.201	0.323	-0.340	3	39	0.760	0.031	1.661	0.413	-0.168	5	36
	$\rho = 0.70$	0.661	0.028	1.308	0.356	-0.091	4	39	0.727	0.016	1.315	0.426	-0.195	4	37
	$\rho = 0.50$	0.658	0.020	1.563	0.422	0.176	4	38	0.729	0.007	1.589	0.497	0.152	5	39
Averaging		0.914	0.448	1.146	0.135	-1.183	11	40	1.070	0.459	3.322	0.389	3.410	24	20
Nonparametric		0.825	0.042	1.579	0.341	-0.522	2	22	1.105	0.050	4.805	0.917	2.181	16	30
Polynomial	$\alpha = \hat{\alpha}$	0.670	0.023	1.347	0.378	-0.220	2	38	0.729	0.002	1.176	0.330	-0.561	3	37
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.802	0.076	1.619	0.283	-0.659	2	40	0.817	0.102	1.339	0.301	-0.690	2	30