

2012/48



Forecasting long memory processes
subject to structural breaks

Shin-Huei Wang, Luc Bauwens
and Cheng Hsiao

The word "CORE" in a bold, black, sans-serif font. A thin, light blue curved line starts above the 'C', arches over the 'O' and 'R', and ends below the 'E'.

CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium

<http://www.uclouvain.be/core>

CORE DISCUSSION PAPER
2012/48

**Forecasting long memory processes
subject to structural breaks**

Shin-Huei WANG¹, Luc BAUWENS²
and Cheng HSIAO³

December 2012

Abstract

We develop an easy-to-implement method for forecasting a stationary autoregressive fractionally integrated moving average (ARFIMA) process subject to structural breaks with unknown break dates. We show that an ARFIMA process subject to a mean shift and a change in the long memory parameter can be well approximated by an autoregressive (AR) model and suggest using an information criterion (AIC or Mallows' Cp) to choose the order of the approximate AR model. Our method avoids the issue of estimation inaccuracy of the long memory parameter and the issue of spurious breaks in finite sample. Insights from our theoretical analysis are confirmed by Monte Carlo experiments, through which we also find that our method provides a substantial improvement over existing prediction methods. An empirical application to the realized volatility of three exchange rates illustrates the usefulness of our forecasting procedure. The empirical success of the HAR-RV model is explained, from an econometric perspective, by our theoretical and simulation results.

Keywords: forecasting, long memory process, structural break.

JEL Classification: C22, C53

¹ Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium; National TsingHwa University, Taiwan.

² Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: luc.bauwens@uclouvain.be
This author is also member of ECORE, the association of CORE and ECARES.

³ Department of Economics, University of Southern California, USA; City University of Hong Kong; WISE, Xiamen University, China.

1. Introduction

Macroeconomic and financial time series are subject to occasional structural breaks (e.g., Stock and Watson, 1996; Pesaran and Timmermann, 2002). Ignoring the presence of breaks can lead to seriously biased estimates and forecasts (see Clements and Hendry, 1998). Accordingly, a conventional approach to forecast a time series with breaks is first to determine when the most recent break occurred and then to use the post-break data to estimate the forecasting model. Nevertheless, Pesaran and Timmermann (2005, 2007) showed that such an approach is not necessarily yielding the optimal forecasting performance, especially in the case when the time series are subject to multiple breaks, due to the difficulty in estimating the timing of breaks. Moreover, they illustrated that pre-break data can be informative for forecasting the after-break outcomes, provided that the break is not too large.

Many researchers use the autoregressive fractionally integrated moving average process of order p, d, q , denoted as ARFIMA (p, d, q) , or $I(d)$ process, to model and forecast time series, where the differencing parameter d is a fractional number between -0.5 and 0.5 . The main feature of the stationary $I(d)$ process is that its autocovariance function declines at a hyperbolic rate, slower than the geometric rate of stationary ARMA processes. For example, Ding *et al.* (1993) and Bollerslev and Mikkelsen (1996) showed that the persistence in stock market volatility could be well described by a long memory process. These findings further induced Hidalgo and Robinson (1996) to consider the issue of structural stability of a regression model with a long memory error term. However, Kuan and Hsu (1998) point out that the Hidalgo and Robinson (1996) test could have large size distortions. Additionally, extending the results of Nunes *et al.* (1995), Kuan and Hsu (1998) show that the conventional break tests for stationary long memory processes may misleadingly infer a structural break when there is none. Because of the possibility of misleading inference by the existing structural change tests for long memory processes, there has been scant attention to suggesting optimal methods to forecast long memory processes in the presence of structural breaks.

The purpose of this paper is to propose an easy-to implement approach for forecasting a long memory process subject to structural breaks. The conventional forecasting method based on post-break data can be suboptimal because the break detection approach may lead to spurious conclusions concerning the number of breaks even there is none (e.g. Granger and Hyung, 2004). Furthermore, Diebold and Inoue (2001) showed that regime switching can generate spurious long memory behavior in observed time series, thus a shift in the mean may be mistaken for a change in the differencing parameter d from $d = 0$ to a different value between minus and plus 0.5 . Chen *et al.* (1999) and Granger and Terasvirta (1999) showed many nonlinear models can display long memory characteristics. Moreover, Granger and Hyung (2004)

and Choi *et al.* (2010) also showed that an increase of the number of mean breaks makes the memory of the process seemingly more persistent. Choi *et al.* (2010) use a break-adjusted forecast method to reduce the forecast error for several realized volatility series, which are modelled as long memory processes with breaks, however, their method still depends on the knowledge of the accurate break dates. To avoid the risk of misleading inference on the break date or imprecise estimation of the fractional parameter d , we note that first, Granger (1980) has shown that when a long memory process has a break in the parameter d , the complete time series can be represented by another long memory process with memory parameter d^* that is a linear combination of the pre-and the post-break memory parameters d_1 and d_2 . Second, Poskitt (2007) and Wang and Hsiao (2012) have shown that a stationary ARFIMA (p, d, q) process can be approximated by an autoregressive process with an order increasing with a certain rate of sample size T . We therefore suggest using an autoregressive approximation to predict the outcome for long memory processes with breaks. We justify the use of an autoregression approximation when the time series follows an ARFIMA (p, d, q) process, for two types of structural breaks: a change in the long memory parameter and/or a shift in mean when the shift size is within some magnitude of standard deviation of the random noise. Furthermore, an AR approximation approach also has the advantage of avoiding the inaccurate parameter estimation and the break locations.

An important issue in finite sample is to select the appropriate order of the AR model. We suggest using Mallows' C_p criterion to select the order of an $AR(k)$ model fitted to a long memory process with structural change. Our Monte Carlo experiments show that the lag length based on the Mallows' criterion to approximate the ARFIMA process with structural breaks is usually small in a sample of two hundred observations. and the residual variance estimate is very close to the true error variance. Our simulation experiments confirm the theoretical analysis clearly by demonstrating that an AR-approximation forecast method for forecasting a long memory process with structural breaks outperforms conventional methods, namely the two naive ARFIMA-based methods, the post-break and Tiao and Tsay (1994) adaptive forecasting methods, even in cases where the structure of an ARFIMA model, including its parameters and lag orders, changes dramatically after breaks. Furthermore, for the special case in which the structural breaks take place immediately prior to the forecast period, our AR-approximation also performs better. An empirical forecasting exercise of realized volatility for the DEM/USD, USD/YEN and GBP/USD spot exchange rates shows that our AR-approximation dominates the existing methods.

We present the basic model and theoretical results in Section 2. Section 3 provides the mean-squared prediction errors of forecasts generated by an AR approximation, post-break model, and naive forecast. Section 4 provides the finite sample simulation comparison. Section 5 provides the comparison of different methods for predicting volatilities. Concluding remarks are in Section 6. Proofs are in the Appendix.

2. Model and Theoretical Results

2.1. The Basic Model

Let

$$(1 - L)^{d_1}(\eta_t^{(1)} - \mu_1) = e_t \quad (1)$$

$$(1 - L)^{d_2}(\eta_t^{(2)} - \mu_2) = e_t \quad (2)$$

be two $ARFIMA(0, d, 0)$ processes where (i) $d_1, d_2 \in (-0.5, 0.5)$, $d_1 \neq 0$ and $d_2 \neq 0$ are differencing parameters; (ii) e_t is an independently and identically distributed process, with $E(e_t) = 0$, $E(e_t^2) = \sigma^2$, and $E(e_t^4) < \infty$. We suppose the T time series observations, η_t take the form

$$\eta_t = \eta_t^{(1)} \quad \text{for } t = 1, 2, \dots, T_1,$$

and

$$\eta_t = \eta_t^{(2)} \quad \text{for } t = T_1 + 1, \dots, T,$$

where $T_1 = \kappa T$, $\kappa \in (0, 1)$. We consider two scenarios:

Case I: Changes in the differencing parameter only, i.e., $d_1 \neq d_2$, $\mu_1 = \mu_2 = \mu$.

Case II: Changes in both differencing parameter and mean, i.e., $d_1 \neq d_2$, $\mu_1 \neq \mu_2$.

We focus on breaks in the mean and long memory parameters in the DGP. We could, of course, examine $ARFIMA(p, d, q)$ models with breaks in the coefficients of the AR and MA terms. However, this would substantially complicate the notations and derivations without gaining insight.

Lemma 1¹. *When the DGP satisfies the basic model, with $T_1 = \kappa T$, $\kappa \in (0, 1)$, the observed data can be represented by a long memory process with long memory parameter d^* that is a linear combination of the pre-and post-break parameters d_1 and d_2 , and with a mean μ^* that is a linear combination of the pre-and post-break means μ_1 and μ_2 , that is*

$$(1 - L)^{d^*}(\eta_t - \mu^*) = e_t, \quad (3)$$

where $\mu^* = \kappa\mu_1 + (1 - \kappa)\mu_2$, $d^* = \lambda d_1 + (1 - \lambda)d_2$, $0 \leq \lambda \leq 1$. When T is finite, $\kappa \rightarrow 0$, $\lambda \rightarrow 0$, and $\kappa \rightarrow 1$, $\lambda \rightarrow 1$. When $T \rightarrow \infty$, $d^* = \max(d_1, d_2)$ and $\mu^* = \kappa\mu_1 + (1 - \kappa)\mu_2$.

Brockwell and Davis (1991) have shown that a long memory process can be represented by an infinite order autoregressive process,

$$\eta_t = \mu + \sum_{j=1}^{\infty} \beta_j \eta_{t-j} + e_t, \quad (4)$$

¹ Lemma 1 is a generalization of that of Granger (1980) with less restrictive conditions on the distribution of the data generating process and also allows breaks in the mean.

where $\beta_j = \Gamma(j - d)/[\Gamma(j + 1)\Gamma(d)]$, $d \in (-0.5, 0.5)$, $d \neq 0$, and e_t is a zero mean white noise process with constant variance σ^2 . Poskitt (2007) and Wang and Hsiao (2012) have shown that η_t can be approximated by an ever increasing order autoregressive model, $\text{AR}(k)$, as T increases,

$$\eta_t = e_{t,k} + \sum_{j=1}^k \beta_{jk} \eta_{t-j} + \beta_{0k}, \quad (5)$$

where $e_{t,k}$ is the prediction error and $\beta_{jk}, j = 0, 1, 2, \dots, k$ are the coefficients of the minimum mean squared predictor of η_t based only on a constant term and the past observations $\eta_{t-1}, \eta_{t-2}, \dots, \eta_{t-k}$.

THEOREM 1². *When the DGP satisfies the basic model, with $T_1 = \kappa T$, $\kappa \in (0, 1)$, when $d^* \in (-0.5, 0.5)$ there exists an $\text{AR}(k)$ approximation of η_t , as $T \rightarrow \infty$, $k = O(T^r)$, $r > 2d^*/(1 + 2d^*)$, such that*

1. $\|\hat{\beta}(k) - \beta(k)\| = O_p(k^{-1/2}(\log T)^{d^* - 0.5} T^{-d^*})$,
2. $\hat{\sigma}_{t,k}^2 = \frac{1}{T-k} \sum_{t=k+1}^T \hat{e}_{t,k}^2 = \sigma^2 + O_p(k^{-2d^* - 1} T^{2d^*}) = \sigma^2 + o_p(1)$, where $\hat{\beta}(k)$ is the OLS estimator of $\beta(k)$ and $\hat{e}_{t,k}$ the OLS residual.

Given Lemma 1 and Theorem 1, we suggest using an $\text{AR}(k)$ model to approximate the data generating process of η_t and using it to generate post-sample predictions.

2.2. Mallows' C_p Criterion for Choosing k

Theorem 1 says that $k \rightarrow \infty$ as $T \rightarrow \infty$. As shown in Shibata (1980), for each T , there is an optimal k^* to balance the estimation error and bias of the finite $\text{AR}(k)$ approximation. In other words, the adequacy of an approximate model for the DGP depends on the choice of the lag order. Therefore, in this paper we propose using AIC (Akaike, 1973) or C_p criterion (Mallows, 1973) criterion to choose a suitable order k^{AIC} or k^{C_p} of an $\text{AR}(k)$ model to approximate a long memory process subject to a break of the kind defined for our basic model.

Mallows' criterion is based on choosing the lag length to minimize the expected mean-squared prediction error divided by the variance of the DGP. Because the complete time series can be represented by another long memory process with parameter d^* , when a long memory process has a break in the long memory parameter, by Poskitt (2007) and Wang and Hsiao (2012), we note the AIC and C_p criteria are simply

$$\text{AIC}(k) = \log \hat{s}_{t,k}^2 + \frac{2k}{T}$$

² The rate of convergence of $\|\hat{\beta}(k) - \beta(k)\|$ is shown faster than that of Poskitt (2007) or Wang and Hsiao (2012) because we use a different methodology.

and

$$C_p(k) = (T - k) \left(\frac{\hat{\sigma}_{t,k}^2}{\hat{\sigma}^2} - 1 \right) + 2k,$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2 (or the true value in a simulation setup).

It appears that a suitable procedure for choosing an appropriate k^* for an $AR(k)$ approximation to our basic model is to minimize $AIC(k)$ or $C_p(k)$ with respect to k . More specifically, following the analysis of Shibata (1980), Poskitt (2007) showed that $AR(k^{AIC})$ and $AR(k^{C_p})$ model are asymptotically efficient. We conducted Monte Carlo experiments about the choice of k using these two criteria for several DGPs and sample sizes. Because of nearly identical results between AIC and C_p (the performance is same as the results of Wang and Hsiao (2012)), we only show the results by C_p in our simulations. Those results are available upon request. The main conclusion is that when $T=200$, a value of k^{C_p} not larger than 4 often provides a good AR-approximation of long memory processes subject to one or two breaks, which are representative of empirical models reported in the literature.

3. Comparison of Forecasting Methods

The usual approaches to forecasting a long memory process with structural breaks are the “post-break (PB)” and “naive” ARFIMA-based forecast (NV) methods. The post-break method uses an estimated ARFIMA model after the last identified break date. The conventional naive method uses a structural break test to detect if there is a break. If there is a break in the mean, but no break in AR and/or MA operators, both pre-and post-break mean filtered data are used to estimate the post-break ARFIMA model (Hyung and Franses, 2001). The resulting model is then used to generate forecasts. Inspired by Lemma 1 and without concern for break detections, we also consider an alternative naive ARFIMA (d^*)-based method (NVNO). In this section, we show by a theoretical argument that h -step ahead forecasts of a long memory process subject to a structural break computed on the basis of our AR-approximation method dominate, in the mean-squared error sense, those of the post-break and naive ARFIMA-based forecast methods. Our analysis is based on selecting the minimum mean-squared error of the point forecast of η_{t+h} in the context of our basic model.

We assume the post-sample η_t is generated from $(1 - L)^{d_2}(\eta_t^{(2)} - \mu_2) = e_t$. When there is no structure break for the observed η_t , $t = 1, 2, 3, \dots, T$, $\kappa = 0$, $\mu^* = \mu_2$ and $d^* = d_2$. In deriving the mean square prediction error of different forecasting methods, we assume T_1 is known and the post-break window size, $T - T_1 = (1 - \kappa)T = T_2$, is not too large, because otherwise the forecast of η_{T+h} will just be dominated by post-break observations without loss of important information.

Following Lemma 1, η_t can be written as

$$\eta_t = (1 - L)^{d^*} \mu^* + \sum_{j=1}^{\infty} \beta_j \eta_{t-j} + e_t, \quad \text{where } \beta_j \sim C j^{-d^*-1}. \quad (6)$$

Approximating (7) by an AR(k) model yields

$$(1 - \sum_{j=1}^k \beta_{jk} L^j) \eta_t = (1 - \sum_{j=1}^k \beta_{jk} L^j) \mu^* + e_t = \beta_{0k} + e_t,$$

where $\beta_{0k} = (1 - \sum_{j=1}^k \beta_{jk}) \mu^*$ and $E(\eta_t) = \mu^*$. Hence, the h -step ahead forecast based on the AR(k) approximating model is

$$\hat{\eta}_{T+h} = \hat{\beta}_{0k} + \sum_{j=1}^k \hat{\beta}_{jk} \eta_{T+h-j}. \quad (7)$$

The h -step ahead predictor $\tilde{\eta}_{T+h}$ of the naive ARFIMA-based method can be computed recursively by using $\tilde{\pi}_1(L) = \sum_{j=0}^{\infty} \tilde{\pi}_{1,j} L^j = (1 - L)^{\tilde{d}_1}$, $\pi_{1,j} = \frac{\Gamma(j - \tilde{d}_1)}{\Gamma(j+1)\Gamma(-\tilde{d}_1)}$ and $\tilde{\pi}_2(L) = \sum_{j=0}^{\infty} \tilde{\pi}_{2,j} L^j = (1 - L)^{\tilde{d}_2}$, $\pi_{2,j} = \frac{\Gamma(j - \tilde{d}_2)}{\Gamma(j+1)\Gamma(-\tilde{d}_2)}$, where $\tilde{\pi}_{1,j} \sim j^{-\tilde{d}_1-1}/\Gamma(-\tilde{d}_1)$ and $\tilde{\pi}_{2,j} \sim j^{-\tilde{d}_2-1}/\Gamma(-\tilde{d}_2)$ denote the forecast coefficients based on the estimated memory parameter $\tilde{d}_1$ and $\tilde{d}_2$ before and after the break, respectively. In other words, we need to estimate d_1 and d_2 first and then plug into the formula of $\pi_{1,j}$ and $\pi_{2,j}$. For the conventional ARFIMA-based prediction approach, there could be two possibilities—a break is detected or not detected. Suppose there is a break at $T_1 = \kappa T$, $\kappa \in (0, 1)$ for the observed η_t . If we do not detect an existing break by a structural break test, we use a naive ARFIMA-based model, which is similar to the NVNO method³, with a long memory parameter d^* and μ^* to generate the forecast. Explicitly,

$$\tilde{\eta}_{T+h} = (1 - L)^{\tilde{d}^*} \tilde{\mu}^* + \sum_{j=1}^T \tilde{\pi}_j^* \eta_{T+h-j}, \quad (8)$$

where $\tilde{\pi}_j^* \sim j^{-\tilde{d}^*}$ is derived from the estimated parameter $\tilde{d}^*$.

On the other hand, if a break is detected at T_1 , then $\eta_t = \eta_t^{(2)}$ for data generated by the post-break model $e_t = (1 - L)^{d_2}(\eta_t^{(2)} - \mu_2)$, for $t \geq T_1 + 1$. Specifically,

$$\eta_{T+h} = (1 - L)^{d_2} \mu_2 + \sum_{j=1}^{T_2} \pi_{2,j} \eta_{T+h-j}^{(2)} + \sum_{j=T_2+1}^T \pi_{2,j} \eta_{T+h-j}^{(2)} + e_{T+h}.$$

³ For ease of exposition, we follow Brockwell and Davis (1991) to do forecast based on all observations (T). Here, we do not consider a truncated order k fitted to the observed time series.

However, $\eta_t^{(2)}$ for t before $T_1 + 1$ are unobservable. On the other hand, we do know e_t for $t \leq T_1$, which is equal to $(1 - L)^{d_1}(\eta_t^{(1)} - \mu_1) = e_t$, for $t = 1, 2, \dots, T_1$. Since $\eta_t^{(2)} = \mu_2 + (1 - L)^{-d_2}e_t$ and $d_1 - d_2 = d_3$, therefore, for t before $T_1 + 1$,

$$\eta_t^{(2)} = \mu_2 + (1 - L)^{-d_2}(1 - L)^{d_1}(\eta_{T+h-j}^{(1)} - \mu_1) = \mu_2 + (1 - L)^{d_3}(\eta_{T-j+h}^{(1)} - \mu_1).$$

Thus, η_{T+h} expressed in terms of $\eta_t, t = 1, 2, \dots, T$, becomes

$$\begin{aligned} \eta_{T+h} &= (1 - L)^{d_2}\mu_2 + \sum_{j=1}^{T_2} \pi_{2,j}\eta_{T+h-j} \\ &+ \sum_{j=T_2+1}^T \pi_{2,j} \left((1 - L)^{d_3}(\eta_{T+h-j} - \mu_1) + \mu_2 \right) + e_{T+h}. \end{aligned} \quad (9)$$

Let a and $1 - a$ be the probability of finding no break (type II error) and a break, respectively. Let $\pi_j^*, \pi_{1,j}, \pi_{2,j}$ denote the forecast coefficients derived from estimated d^*, d_1 and d_2 , i.e., $\tilde{d}^*, \tilde{d}_1, \tilde{d}_2$. Then the predictor for η_{T+h} based on the naive forecast approach becomes

$$\begin{aligned} \tilde{\eta}_{T+h}^{II} &= a \left((1 - L)^{\tilde{d}^*} \tilde{\mu}^* - \sum_{j=1}^T \tilde{\pi}_j^* \eta_{T+h-j} \right) + (1 - a) \left((1 - L)^{\tilde{d}_2} \tilde{\mu}_2 + \sum_{j=1}^{T_2} \tilde{\pi}_{2,j} \eta_{T+h-j} \right. \\ &+ \left. \sum_{j=T_2+1}^T (1 - L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \eta_{T+h-j} - \sum_{j=T_2+1}^T (1 - L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \mu_1 + \sum_{j=T_2+1}^T \tilde{\pi}_{2,j} \mu_2 \right). \end{aligned}$$

In the event there is no break in the observed η_t , i.e., $\kappa = 0, \eta_t = \eta_t^{(2)}$. Let b denote the probability of finding no break and $1 - b$ denote the probability of finding a break (type I error). Then the naive forecast for η_{T+h} is equal to

$$\begin{aligned} \tilde{\eta}_{T+h}^I &= b \left((1 - L)^{\tilde{d}_2^*} \tilde{\mu}_2^* + \sum_{j=1}^T \tilde{\pi}_{2,j}^* \eta_{T+h-j} \right) + (1 - b) \left((1 - L)^{\tilde{d}_2} \tilde{\mu}_2 + \sum_{j=1}^{T_2} \tilde{\pi}_{2,j} \eta_{T+h-j} \right. \\ &+ \left. \sum_{j=T_2+1}^T (1 - L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \eta_{T+h-j} - \sum_{j=T_2+1}^T (1 - L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \mu_1 + \sum_{j=T_2+1}^T \tilde{\pi}_{2,j} \mu_2 \right), \end{aligned}$$

where $(\tilde{d}_2^*, \tilde{\mu}_2^*)$ and $(\tilde{d}_2, \tilde{\mu}_2)$ are based on MLE estimates of μ_2 and d_2 based on T and $T_2 = (1 - k)T$ observations, respectively.

The h -step ahead forecast of post-break forecast is just

$$\hat{\eta}_{T+h}^* = \left((1 - L)^{\hat{d}_2} \hat{\mu}_2 + \sum_{j=1}^{T_2} \hat{\pi}_{2,j} \eta_{T+h-j} \right). \quad (10)$$

The following theorem shows the comparison of the h -step ahead mean squared error for four forecasting methodologies.

THEOREM 2. *If the DGP satisfies the basic model, (i) the AR(k) approximation-based forecast method (AR) yields no greater mean squared error than the post-break forecast and the two naive ARFIMA-based forecast (NV and NVNO) method if $\kappa|\mu_1 - \mu_2|$ is of smaller order than $\sigma_\eta/\sqrt{(1 - \kappa)T}$, where σ_η denotes the standard error of η ; (ii) when $k = O((\ln T)^{-1}(T/\ln T)^{1-2d^*})$, the AR outperforms ARFIMA(d^*) (NVNO).*

Tiao and Tsay (1994) and Baillie and Morana (2009) suggest an adaptive forecasting approach to account for both long memory and structural change in the conditional variance process. The basic idea of their approach is to obtain the estimates by minimizing the sum of squares of h -step-ahead forecast errors. More specifically, as each additional post-sample time series observation becomes available, the parameters of the forecasting model is recursively re-estimated by minimizing the sum of squares of h -step-ahead forecast errors based on the rolling window with a fixed window size where with each additional available observation, the observation at the beginning of the time period is dropped. This approach would be inefficient⁴ compared to the recursively updating the parameter estimates when each additional observation becomes available if there is no structural break. On the other hand, if there is indeed a structural break at time period T_1 , as more and more post-sample information becomes available, this is equivalent to pushing the break point k towards zero in Lemma 1 and Theorem 2. In other words, our AR approximation-based forecast approach under Tiao and Tsay (1994) and Baillie and Morana (2009)'s adaptive forecasting scheme, denote as AFAR, yields

Lemma 2. *If the DGP satisfies the basic model (with a fixed break data T_1), the adaptive AR approximation-based forecast method (AFAR) with the rolling window (with window size T_n) eventually yields no greater mean squared error than the post-break forecast and the two naive ARFIMA-based forecast (NV and NVNO) methods for Case I and for Case II.*

Theorem 2 and Lemma 2 illustrate the performance of the two AR approximation-based forecast method (AR and AFAR) that are limited to the case where the DGP has a single structural break. Although the analysis can be extended to the cases of more breaks, the extension is tedious to show. Additionally, Poskitt (2007) points out estimating the parameters of the AR approximation model by directly minimizing the observed mean squared error could lead unstable estimates. Thus, AR and AFAR could exhibit quite different finite sample behaviour. We therefore turn to resort to Monte Carlo experiments to provide intuition on these extensions, and also to illustrate the difference in performance of the different forecasting methods for the h -step forecasts under a single break.

⁴ Tiao and Tsay (1994) points out the adaptive forecasting would be inefficient if the model used is indeed the "correct" one.

4. Comparison of Forecasting Methods: Simulation Results

4.1 Simulation Design

In light of the empirical evidence provided by Christensen and Nielsen (2006), and Choi *et al.* (2010), we consider

$$\begin{aligned}\text{DGP (a): } & (1 - L)^d(\eta_t - \mu) = e_t, \\ \text{DGP (b): } & (1 - 0.7L)(1 - L)^d(\eta_t - \mu) = (1 + 0.5L)e_t, \\ \text{DGP (c): } & (1 - 0.7L - 0.2L^2)(1 - L)^d(\eta_t - \mu) = e_t.\end{aligned}$$

We let the above DGPs be subject to breaks in the mean and in the long memory parameter, each being considered with $d = (0.1, 0.2, 0.3, 0.4, 0.45)$ and $\sigma^2 = 1$, see Tables 1 and 2 for the values of the break point, means and differencing parameters. Simulations are performed for $T = 200$ and repeated 1000 times. Following McLeod and Hipel (1978) and Hosking (1984), we first generate T independent $N(0,1)$ values and form a $T \times 1$ vector e , then generate the vector y of the T realized values of the $I(d)$ process as $y = Ce$, where C is the Cholesky decomposition of prespecific $T \times T$ autocovariance matrix $\Sigma = CC'$. For each T , we generate $T + 200$ observations and discard the first 200 to reduce the effects of initial values.

For each experiment, we assume the break dates are either known or unknown. In the case of unknown break points, tests proposed by Bai and Perron (1998) and Beran and Terrin (1996) are used to detect the change points. We then use the Schwarz information criterion to select the number of breaks for each model. To estimate the differencing parameter d , we use the GPH and the approximate ML estimation methods. Both methods provide similar results confirming that the higher p in the ARFIMA (p, d, q) DGP, the larger the bias of the estimate of d .

We consider both AR forecasting and the AFAR forecasting procedure. The AFAR is the combination of the adaptive forecasting scheme and our AR approximation. If the AR model we employed here is indeed the "correct" one (the estimate of σ^2 converges to σ^2 asymptotically), then it would be inefficient to use the adaptive forecasting (see Tiao and Tsay, 1994). Additionally, we also compute forecasts using the naive ARFIMA (d^*) model (3) without concern for the break detection (NVNO). We compare the out-of-sample forecasting performance of the three forecast methods described in Section 3 and that of the AFAR and NVNO methods for two forecasting horizons: $h = 1, 5$.

The recursive forecast window scheme is computed as follows: given a simulated sample of size $T = 2\tilde{T}$, we estimate the model using the first half of the sample, and generate the forecasts of $\eta_{\tilde{T}+h}$ for $h = 1, 5$. Then we add one observation to the estimation sample, and generate the forecasts of $\eta_{\tilde{T}+1+h}$. We continue like this until

observation $T - h$, which enables to forecast η_T . We denote by $\eta_{\tilde{T}+f+h}^M$ the h -step ahead forecast of $\eta_{\tilde{T}+f+h}$ using model M and information up to date $\tilde{T} + f$. The relative RMSFE of model M with respect to the $AR(k)$ model, for $\tilde{T}$ forecasts at horizon h , is defined as

$$\frac{\left(\sum_{f=0}^{\tilde{T}-h} \left(\eta_{\tilde{T}+f+h} - \eta_{\tilde{T}+f+h}^{AM}\right)^2\right)^{1/2}}{\left(\sum_{f=0}^{\tilde{T}-h} \left(\eta_{\tilde{T}+f+h} - \eta_{\tilde{T}+f+h}^{AR}\right)^2\right)^{1/2}}. \quad (11)$$

The value of k for the $AR(k)$ model is chosen by Mallows' criterion. Table 3 reports the root mean squared forecast errors (RMSFE) of the naive, naive-no and post-break methods, relative to that of the AR-based method with recursive forecasting window. We further consider the rolling estimation window. The window size is based on the suggestion of Hansen and Timmermann (2012) to use half of the full sample for in-sample estimation (=100 observations), and the remaining for out-of-sample evaluation⁵. Those relevant results are reported in Table 4. If the relative RMSFE is smaller than one, the alternative model forecasts better than the AR model.

To further determine which model provides more accurate information about the future, we use the Mincer and Zarnowitz (1969) encompassing regression for the one-step-ahead forecasts of four forecast models with the rolling forecast window

$$\eta_{t+1} = \alpha_0 + \alpha_1 \eta_{t+1|t}^{AR} + \alpha_2 \eta_{t+1|t}^{AM} + \varepsilon_t,$$

where $\eta_{t+1|t}^{AR}$ denotes predictions by our benchmark AR approximation model, and $\eta_{t+1|t}^{AM}$ denotes predictions from four alternative models. The results show that the forecasts of the AR-approximation generally outperform the alternative methods. We find that :

- (1) Unknown or known break dates: the AR approximation ignoring the possibility of breaks almost always generates more accurate one-step ahead forecasts. For the multiperiod ahead forecasts, there are very few cases where post-break forecasts are more accurate than forecasts with known break dates. However, if break dates are unknown, which is the most relevant situation in empirical studies, the AR-based forecast method dominates other methods in 3 out of 15 experiments for all considered forecast horizons and forecast windows. The domination is substantial, with frequently occurring gains in RMSFE of more than 20 per cent. More precisely, although using the rolling window could reduce the RMSFE of

⁵ We thank the editors for this suggestion.

comparison models for several cases (experiment 1,2,6,7,8 and 12), the AR-based forecast method in general still performs the best.

- (2) (i) AR compared to NVNO: for most cases, AR dominates uniformly, except in experiment 11 for horizon $h = 5$. More interestingly, for the cases where the series properties shift (experiment 1, 2, 7,8 and 15), the NVNO even performs the worst among all forecast methods, although the rolling window could reduce its RMSFE. These results could be due to the inaccurate estimation of the NVNO model in finite samples. (ii). AR compared to the AFAR method: for most cases, AR performs much better than ARFA method, except for the experiment 3,4,5,9,10 and 13, where we only consider the change in long memory parameters and rolling window estimation. However, the AFAR performs better than the other five methodologies. We further calculate the average (AVE) of RMSFE for each h and forecast method with two different forecast windows. The results do not change.
- (3) The results in Table 5 show the superior forecasting ability of the AR model. In the encompassing regression that includes both the AR approximation-based forecast model and an alternative forecast, the estimates of α_1 are close to unity and the estimates for α_2 are close to zero whether the break date is known or unknown. We also use a F -test for the null hypothesis of $\alpha_0 = 0$, $\alpha_1 = 1$ and $\alpha_2 = 0$ at the 5% significant level. The P-value of the F -test ($P(F)$) are all above 15%. Furthermore, including an alternative forecast method has little contribution to increasing R^3 ⁶.

4.2 Comparison When the Break Occurs at the End of the Sample

Forecasting observations of long memory processes just after a structural break is a difficult task. The post-break forecast method cannot be implemented in this case, due to the absence of informative observations about the post-break period. The naive ARFIMA-based method can be used with estimates from the pre-break period, being in addition prone to the spurious break and inaccurate parameter estimation problems. Moreover, in empirical research, it is difficult to know exactly when structural breaks occur. Thus, the existing forecasting approaches are ill-suited for real-time forecasting. Given that a structural break occurs at T_1 , we present in Table 7 the performance of the AR-approximation (AR), AFAR and NV of observations $T_1 + h$ for $h = 1, 2, \dots, 5$, assuming we do not know the existence of the break to generate the forecasts. The results clearly indicate that the AR-approximation method produces the lowest RMSFEs.

⁶ For the encompassing regression, the first forecast is said to subsume information in other forecasts if these additional forecasts do not significantly increase the R^2 .

In closing this section, it is worth pointing out that the superior performance of our AR-approximation forecast method goes along with its ease of implementation in empirical analysis. Intuitively, the reason why the AR-based forecasting method produces the best performance is that it avoids the inaccurate estimation of the fractional parameters and the issue of spurious breaks induced by long memory processes

7

5. Forecasting Realized Volatilities

Volatility forecasting plays a central role in financial decision making, for example, fund managers adjust the portfolio allocations according to the volatility forecasts. Choi *et al.* (2010) use a break-adjusted forecast method to reduce the forecast error for several realized volatility series, which are modelled as long memory processes with breaks. In order to show the potential usefulness of our AR-approximation based forecasting method, we compare the forecasts of our method to the four competing methods considered in Section 3, the heterogeneous autoregressive (HAR) model of Corsi (2004), and the VAR-RV-Break model proposed by Choi *et al.* (2010). The HAR model is actually an AR model with only three coefficients for 21 lags, plus a constant term, which can be written

$$RV_t = \alpha_0 + \alpha_1 RV_{t-1} + \alpha_2 \sum_{i=1}^5 \frac{1}{5} RV_{t-i} + \alpha_3 \sum_{j=1}^{21} \frac{1}{21} RV_{t-j} + u_t.$$

Corsi's justification for this model was to capture the long memory property which he justifies by an economic argument. Many researchers have found empirically this model to fit and to forecast realized volatility series better than ARFIMA models. The HAR model can be seen as a particular case of our AR approximation, where the lag order is fixed to 21 whatever the sample size, and the coefficients are constrained. Our theoretical and simulation results in this paper explain to a large extent, from an econometric viewpoint, why the HAR model is a good model in the context of realized volatility modelling.

The VAR-RV-Break (VRB) model is a break-adjusted VAR, denoted as

$$\Phi(L)(Y_t^* - \mu) = \varepsilon_t,$$

where Y^* is the (3×1) vector of logarithmic realized exchange rate volatilities after mean break adjustments, μ is the unconditional mean, ε_t is a vector white noise

⁷ The reason of why the forecasting performance of ARFIMA (d^*) (NVNO) is worse than that of AR in finite samples might be due to the inaccurate parameter fractional d when d is close to 0.5 and sample size is not large $T = 100, 200$. Furthermore, it is also hard to distinguish the behaviour of long and short range components as data with both components in finite samples (See Crator and Ray, 1996).

process and the lag order is equal to five following Choi *et al.* (2010). The detailed procedure of the mean break adjustments is explained in Choi *et al.* (2010).

We perform the comparisons for the realized volatility series of the spot exchange rate returns of the German mark (DEM) and British pound (GBP) versus the US dollar, and the US dollar versus the Japanese Yen (YEN). The data (obtained from Olsen & Associates) span the period starting on January 6, 1987 and ending on June 30, 1999. They consist of fifteen-minute returns, which we square and aggregate to construct the daily realized volatility. We prepared the data in a way similar to Andersen *et al.* (2001) who use the same series but their sample ends on November 30, 1996 and Andersen *et al.* (2003) and Choi *et al.* (2010) who use the intra-day 30-minute log returns. We transform the realized volatility series in logarithm since this ensures the positivity of forecasts of realized volatility. For simplicity, we denote by RV_t the log of realized volatility for day t . The summary statistics are in Table 7. Unconditionally, the series have a positive skewness and a little too much excess kurtosis to be characterized as Gaussian.

To detect structural break dates, we use the procedure of Beran and Terrin (1996). Table 8 reports the estimated break dates. Four breaks are found for GBP and five for the other series. Thus, for example, there are six regimes for the DEM series, the first one spanning the period from January 6, 1987 to May 22, 1989. The break dates are almost identical for DEM and GBP except that the break of November 3, 1993 for the DEM is not found for the GBP. The Yen shares two breaks with the other series (May 22, 1989 and June 30/July 1, 1997).

To implement the post-break and out-of-sample NV and NVNO forecast methods, we need an ARFIMA(p, d, q) model for each series. We estimated several models using the full sample and each regime by Beran's approximate ML method, and we selected the best model according to Akaike's information criterion, defined as $\text{LN } \hat{\sigma}^2 + 2(p + q + 1)/T$. The results are available on request. The lag orders p and q of the selected models are all either 0 or 1, and the estimated fractional differencing parameters suggest that the series are stationary but have indeed a long memory component (the estimates are between 0.29 and 0.49).

Forecasts are obtained by the rolling estimation window. We follow the suggestion of Hansen and Timmermann (2012) to use half of the full sample for in-sample estimation, and the remaining for out-of-sample evaluation. For the AR-approximation based method, the naive ARFIMA with break method, the naive ARFIMA without break method and the VAR-RV-Break model, the in-sample period goes from January 6, 1987 to Jan 10, 1992 for DEM (1528 observations), Jan 13, 1992 for GBP (1532 observations), and Jan 16, 1992 (1536 observations) for YEN. We estimate the models over these samples and generate forecasts at horizon $h = 1$ and 3. Then we delete the first observation and add the next observation to the estimation sample to generate again the two forecasts. The model is re-estimated and repeated until the last observation of June 1999.

The lag orders k for the approximating $AR(k)$ models of the three series, selected by Mellow's C_p criterion, are equal to 18 for DEM, 19 for GBP, and 18 for Yen RV series. As these values remain the same for the full samples and for the in-sample periods, we kept them fixed throughout the prediction period. These values are close to the 21 lags suggested by Corsi. They are much larger than in our simulation experiments, because the sample size we are dealing with in this application is much higher (about 1522 versus 200).

To compare the predictive ability of the five methods at forecast horizons one and three, we use the RMSFE of the five methods, relative to the RMSFE of the AR-approximation based method. The definition is the same as in equation (11). Since all relative RMSFE are above 1, the results, reported in Table 9, indicate that our AR-approximation based forecast method outperforms the other methods. This is especially spectacular for the post-break method, the naive methods (NV and NVNO) and the VAR-RV-Break model (VRB). Corsi's HAR model involves a small efficiency loss (between 3 and 7 per cent).

We examine the explanatory power of the one period ahead forecasts for the six forecasting models by the Mincer and Zarnowitz encompassing regression. Table 10 shows that in the encompassing regression, the estimates of α_1 are close to unity and the estimates of α_2 are close to zero. Furthermore, these additional forecasts do not significantly increase the R^2 . Under the null hypothesis that $\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = 0$, the P-value of the F -test are all above 15%. Hence these results confirm the better performance of the forecasts based on the AR-approximation.

6. Conclusions

Forecasting is often a difficult task, especially in a rapidly changing world where structural changes occur. Given that some macroeconomic and financial time series display high persistence and signs of structural breaks, it is important to develop adequate forecasting methods. However, there could be differences between the structural analysis and the forecast. In the structural analysis, the aim is to identify the data generating process and obtain the efficient inference on unknown parameters. In forecasting, because of unknown breaks and the size of break, the focus should be on the robustness of forecasting procedures. In this paper, we have developed a simple to implement, and improved forecasting method based on an AR-approximation to forecast a long memory process subject to structural breaks. Additionally, we have proved theoretically and illustrated by Monte Carlo simulations that i) a stationary long memory ARFIMA(p, d, q) process subject to structural breaks can be well approximated by an $AR(k)$ model when k is chosen appropriately, and ii) our AR-based forecast method provides a better out-of-sample forecast performance than conventional methods for forecasting long memory processes under structural breaks. The root of the success of this new method is that it avoids the inaccurate estimation of parameters inherent to the other methods, and the problem of spurious breaks. The

method is shown useful for forecasting the daily realized volatilities of exchange rate series.

**Table 1. Breakpoint Specifications by Experiments(EX)
(One Break)**

EX	$DGP(I)$	$DGP(II)$	τ	d_1	d_2	μ_1	μ_2
1	SD	(a)	0.5	0.0	0.1	0.0	0.0
2	SD	(b)	0.5	0.0	0.1	0.0	0.0
3	(a)	(a)	0.25	0.3	0.45	0.0	0.0
4	(a)	(a)	0.5	0.3	0.45	0.0	0.0
5	(a)	(a)	0.75	0.3	0.45	0.0	0.0
6	(a)	(a)	0.5	0.3	0.45	0.3	0.9
7	(a)	(b)	0.5	0.1	0.4	0.0	0.0
8	(a)	(b)	0.5	0.1	0.4	0.3	0.9
9	(b)	(b)	0.25	0.3	0.45	0.0	0.0
10	(b)	(b)	0.50	0.3	0.45	0.0	0.0
11	(b)	(b)	0.75	0.3	0.45	0.0	0.0
12	(b)	(b)	0.5	0.3	0.45	0.3	0.9

Notes: $SD \sim N(0, 1)$.

**Table 2. Breakpoint Specifications by Experiments (EX)
(Two Breaks)**

EX	$DGP(I)$	$DGP(II)$	$DGP(III)$	τ_1	τ_2	d_1	d_2	d_3	μ_1	μ_2	μ_3
13	(a)	(a)	(a)	0.3	0.6	0.0	0.4	0.2	0.0	0.0	0.0
14	(a)	(a)	(a)	0.3	0.6	0.1	0.3	0.45	0.3	0.9	0.75
15	(a)	(b)	(c)	0.3	0.6	0.1	0.3	0.45	0.3	0.9	0.75

Table 3. Relative Root Mean Squared Forecast Error (Recursive Window)

<i>EX</i>	<i>h</i>	<i>PBK</i>	<i>NVK</i>	<i>PBUK</i>	<i>NVUK</i>	<i>NVNO</i>	<i>AFAR</i>
1	1	1.172	1.383	1.258	1.422	1.479	1.138
	5	1.311	1.373	1.362	1.438	1.522	1.145
2	1	1.119	1.366	1.204	1.335	1.420	1.128
	5	1.212	1.185	1.126	1.347	1.475	1.108
3	1	1.037	0.992	1.065	1.164	1.127	1.062
	5	1.021	0.949	1.078	1.098	1.219	1.066
4	1	0.988	1.070	0.989	1.172	1.192	1.049
	5	1.011	0.978	1.017	1.162	1.208	1.041
5	1	1.109	1.108	1.126	1.136	1.201	1.088
	5	0.932	0.892	0.957	0.986	1.119	1.097
6	1	1.236	1.349	1.406	1.619	1.245	1.117
	5	1.663	1.446	1.390	1.601	1.271	1.145
7	1	1.216	1.253	1.321	1.372	1.877	1.194
	5	1.246	1.257	1.328	1.663	1.902	1.182
8	1	1.214	1.253	1.321	1.372	2.271	1.226
	5	1.246	1.257	1.329	1.663	2.412	1.279
9	1	0.994	1.226	1.181	1.467	1.304	1.101
	5	0.987	1.051	1.200	1.584	1.272	1.117
10	1	1.024	1.180	1.224	1.555	1.122	1.110
	5	0.960	1.100	1.191	1.676	1.008	1.119
11	1	1.346	0.954	1.511	1.117	1.084	1.090
	5	1.194	1.015	1.147	1.101	0.997	1.108
12	1	1.028	0.964	1.173	1.895	1.994	1.128
	5	0.955	0.863	1.509	1.801	1.867	1.148
13	1	1.357	1.010	1.371	1.189	1.277	1.089
	5	1.425	0.990	1.547	1.138	1.257	1.102
14	1	1.216	1.087	1.318	1.418	1.322	1.166
	5	1.143	0.980	1.153	1.375	1.307	1.109
15	1	1.461	1.230	1.687	1.402	2.527	1.161
	5	1.534	1.241	2.077	1.702	2.602	1.125
<i>AVE</i>	1	1.167	1.161	1.277	1.375	1.496	1.129
	5	1.189	1.105	1.294	1.422	1.495	1.126

Notes: PBK and PBUK represent the post-break method with the known and unknown break respectively. NVK and NVUK represent the Naive ARFIMA-based method with the known and unknown break respectively. NOVO represents the ARFIMA (d^*)-based method and ARFA means the additive AR forecast method. AVE means the average of RMSFEs.

Table 4. Relative Root Mean Squared Forecast Error (Rolling Window)

<i>EX</i>	<i>h</i>	<i>PBK</i>	<i>NVK</i>	<i>PBUK</i>	<i>NVUK</i>	<i>NVNO</i>	<i>AFAR</i>
1	1	1.024	1.207	1.151	1.372	1.466	1.122
	5	1.018	1.230	1.212	1.286	1.435	1.129
2	1	1.065	1.249	1.108	1.226	1.407	1.110
	5	1.091	1.221	1.002	1.247	1.426	1.092
3	1	1.057	1.108	1.085	1.204	1.086	1.046
	5	1.018	0.931	1.102	1.118	1.054	1.048
4	1	0.998	1.050	1.100	1.202	1.077	1.030
	5	1.001	0.958	1.027	1.182	1.046	1.036
5	1	1.155	1.104	1.131	1.144	1.098	1.075
	5	0.966	0.992	0.969	0.969	1.070	1.066
6	1	1.021	1.119	1.217	1.339	1.194	1.120
	5	0.993	1.046	1.090	1.201	1.171	1.152
7	1	1.154	1.133	1.219	1.272	1.621	1.180
	5	1.198	1.134	1.228	1.163	1.602	1.182
8	1	1.204	1.231	1.228	1.302	1.982	1.198
	5	1.162	1.229	1.202	1.321	1.902	1.180
9	1	0.992	1.265	1.198	1.512	1.205	1.054
	5	0.971	1.167	1.201	1.574	1.172	1.067
10	1	1.074	1.176	1.281	1.585	1.088	1.078
	5	1.002	1.116	1.297	1.606	1.001	1.060
11	1	1.431	1.014	1.539	1.318	1.037	1.082
	5	1.302	1.016	1.217	1.265	0.989	1.112
12	1	1.112	0.934	1.063	1.783	1.509	1.138
	5	0.962	0.822	1.406	1.393	1.487	1.166
13	1	1.369	1.019	1.307	1.229	1.201	1.084
	5	1.452	1.000	1.516	1.297	1.194	1.097
14	1	1.126	1.057	1.198	1.196	1.262	1.121
	5	0.993	0.940	1.076	1.210	1.216	1.156
15	1	1.341	1.236	1.881	1.359	2.217	1.137
	5	1.389	1.313	1.918	1.476	2.172	1.176
<i>AVE</i>	1	1.141	1.126	1.180	1.336	1.363	1.105
	5	1.101	1.074	1.230	1.287	1.329	1.114

Table 5. Forecast Evaluation of η_{T+1}

EX		α_0	α_1	α_2	R^2	$P(F)$
1	AR	0.010(0.054)	0.957(0.090)		0.266	
	AR+PBK	-0.022(0.079)	0.890(0.132)	0.169(0.100)	0.292	0.38
	AR+PBUK	0.039(0.087)	0.927(0.127)	0.102(0.136)	0.268	0.49
	AR+NVK	0.018(0.049)	0.900(0.099)	0.158(0.104)	0.281	0.40
	AR+NVUK	0.044(0.076)	0.918(0.116)	0.118(0.167)	0.267	0.44
	AR+NVNO	-0.107(0.129)	0.937(0.136)	0.087(0.126)	0.266	0.59
	AR+AFAR	-0.014(0.090)	0.857(0.147)	0.246(0.142)	0.293	0.26
2	AR	0.021(0.049)	0.927(0.110)		0.255	
	AR+PBK	-0.032(0.103)	0.897(0.105)	0.139(0.174)	0.276	0.41
	AR+PBUK	0.049(0.095)	0.908(0.119)	0.124(0.201)	0.257	0.45
	AR+NVK	0.027(0.117)	0.901(0.119)	0.147(0.202)	0.266	0.44
	AR+NVUK	0.097(0.187)	0.927(0.100)	0.109(0.216)	0.259	0.52
	AR+NVNO	-0.156(0.190)	0.912(0.187)	0.104(0.232)	0.256	0.48
	AR+AFAR	-0.075(0.102)	0.887(0.169)	0.194(0.168)	0.278	0.39
4	AR	0.017(0.049)	0.978(0.103)		0.353	
	AR+PBK	0.047(0.095)	0.898(0.129)	0.156(0.104)	0.365	0.32
	AR+PBUK	0.062(0.092)	0.809(0.148)	0.176(0.148)	0.355	0.36
	AR+NV	0.038(0.064)	0.857(0.132)	0.211(0.191)	0.360	0.34
	AR+NVUK	0.089(0.112)	0.882(0.147)	0.199(0.241)	0.354	0.36
	AR+NVNO	-0.022(0.054)	0.876(0.129)	0.201(0.154)	0.358	0.34
	AR+AFAR	-0.021(0.090)	0.789(0.156)	0.344(0.239)	0.386	0.20
6	AR(k)	-0.052(0.067)	0.928(0.103)		0.310	
	AR+PBK	0.027(0.085)	0.869(0.125)	0.157(0.217)	0.330	0.44
	AR+PBUK	0.093(0.101)	0.901(0.156)	0.103(0.165)	0.319	0.49
	AR+NVK	-0.049(0.076)	0.879(0.117)	0.201(0.189)	0.326	0.45
	AR+NVUK	-0.088(0.101)	0.889(0.137)	0.169(0.217)	0.310	0.50
	AR+NVNO	0.100(0.079)	0.899(0.119)	0.108(0.134)	0.312	0.49
	AR+AFAR	-0.029(0.149)	0.801(0.145)	0.349(0.289)	0.341	0.40
7	AR	0.032(0.096)	0.936(0.109)		0.322	
	AR+PBK	0.048(0.133)	0.902(0.155)	0.166(0.184)	0.332	0.53
	AR+PBUK	0.078(0.141)	0.908(0.177)	0.141(0.201)	0.328	0.58
	AR+NVK	0.039(0.093)	0.907(0.185)	0.199(0.169)	0.329	0.56
	AR+NVUK	0.109(0.101)	0.889(0.189)	0.121(0.176)	0.325	0.60
	AR+NVNO	-0.209(0.232)	1.044(0.129)	0.102(0.279)	0.322	0.64
	AR+AFAR	-0.011(0.097)	0.849(0.134)	0.243(0.203)	0.352	0.46

Table 5. Continued

EX	AM	α_0	α_1	α_2	R^2	$P(F)$
8	AR(k)	-0.092(0.111)	1.100(0.123)		0.296	
	AR+PBK	0.094(0.109)	0.942(0.169)	0.141(0.113)	0.320	0.49
	AR+PBUK	0.108(0.153)	0.877(0.173)	0.202(0.186)	0.306	0.58
	AR+NVK	0.100(0.121)	0.902(0.159)	0.167(0.173)	0.314	0.50
	AR+NVUK	0.126(0.123)	0.860(0.181)	0.190(0.202)	0.300	0.61
	AR+NVNO	-0.185(0.196)	1.022(0.129)	0.109(0.279)	0.305	0.58
	AR+AFAR	-0.096(0.165)	0.889(0.104)	0.161(0.224)	0.318	0.52
10	AR	0.031(0.082)	0.932(0.123)		0.346	
	AR+PBK	0.113(0.108)	0.869(0.167)	0.257(0.198)	0.362	0.34
	AR+PBUK	0.141(0.137)	0.841(0.139)	0.241(0.204)	0.352	0.39
	AR+NVK	0.126(0.143)	0.842(0.135)	0.214(0.210)	0.358	0.37
	AR+NVUK	0.152(0.167)	0.887(0.187)	0.165(0.167)	0.349	0.42
	AR+NVNO	-0.029(0.119)	0.857(0.091)	0.201(0.201)	0.359	0.37
	AR+AFAR	0.045(0.135)	0.821(0.114)	0.333(0.199)	0.376	0.28
12	AR	-0.101(0.177)	0.878(0.089)		0.259	
	AR+PBK	0.141(0.123)	0.802(0.177)	0.371(0.194)	0.271	0.63
	AR+PBUK	0.177(0.157)	0.866(0.119)	0.207(0.186)	0.263	0.68
	AR+NV	0.102(0.129)	0.849(0.165)	0.254(0.201)	0.278	0.60
	AR+NVUK	-0.196(0.178)	0.860(0.179)	0.288(0.171)	0.263	0.68
	AR+NVNO	0.149(0.281)	0.822(0.129)	0.171(0.288)	0.265	0.69
	AR+AFAR	-0.103(0.129)	0.843(0.192)	0.298(0.219)	0.284	0.55

Table 6. Root Mean Squared Forecast Error Value of η_{T_1+1}

Experiments	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
AR					
6	1.0092	1.0703	1.1096	1.1702	1.1813
8	1.0227	1.0295	1.0546	1.0909	1.0887
12	1.0095	1.0211	1.0842	1.1293	1.1403
NVNO					
6	1.8350	1.7950	1.7175	1.6747	1.7030
8	1.9345	2.0178	2.1541	2.1919	2.2887
12	1.0175	1.0241	1.1842	1.1593	1.2403
AFAR					
6	1.1068	1.0624	1.1112	1.1628	1.1800
8	1.0369	1.0478	1.0591	2.1221	1.1276
12	1.1075	1.0281	1.1139	1.1333	1.1479

Table 7. Statistics for the Log-Realized Volatilities

Statistic	<i>DEM/USD</i>	<i>USD/YEN</i>	<i>GBP/USD</i>
T	3056	3071	3063
Mean	-0.99951	-0.87484	-1.0810
Std.Devn	0.7165	0.80617	0.6799
Skewness	0.3198	0.2520	0.3775
Excess Kurtosis	0.5255	0.5606	0.5259
Normality $\chi^2(2)$	55.729	47.559	69.825
Minimum	-3.3947	-3.5447	-3.1705
Maximum	1.8967	3.6103	2.0873

Table 8. Numbers of Breaks and Break Dates

Series	#	T ₁	T ₂	T ₃	T ₄	T ₅
DEM/USD	5	1989.5.22	1991.3.21	1993.3.11	1995.9.1	1997.7.2
USD/YEN	5	1989.5.22	1991.7.8	1993.4.20	1997.6.30	1998.7.20
GBP/USD	4	1989.5.19	1991.4.3	1995.8.25	1997.7.1	

Table 9. Relative Root Mean Squared Forecast Error Values

Series	<i>h</i>	<i>PB</i>	<i>NV</i>	<i>NVNO</i>	<i>HAR</i>	<i>VRB</i>	<i>AFAR</i>
DEM	1	1.586	1.487	1.551	1.068	1.214	1.093
	3	1.578	1.502	1.526	1.062	1.208	1.108
YEN	1	1.412	1.689	1.231	1.046	1.181	1.063
	3	1.398	1.668	1.256	1.054	1.186	1.098
GBP	1	1.566	1.595	1.519	1.032	1.163	1.060
	3	1.612	1.584	1.511	1.030	1.158	1.047

Table 10. Out-of-Sample Forecast Evaluation

EX		α_0	α_1	α_2	R^2	$P(F)$
DEM	AR	0.088(0.074)	0.918(0.129)		0.338	
	AR+PB	0.112(0.189)	0.900(0.172)	0.179(0.130)	0.340	0.62
	AR+NV	-0.089(0.053)	0.887(0.190)	0.218(0.152)	0.345	0.52
	AR+NVNO	-0.118(0.109)	0.863(0.129)	0.243(0.134)	0.344	0.44
	AR+HAR	-0.038(0.089)	0.805(0.119)	0.292(0.097)	0.373	0.24
	AR+VRB	-0.091(0.165)	0.842(0.131)	0.263(0.126)	0.356	0.36
	AR+AFAR	-0.069(0.190)	0.798(0.149)	0.337(0.219)	0.377	0.26
YEN	AR	0.029(0.089)	0.945(0.129)		0.403	
	AR+PB	-0.068(0.116)	0.872(0.112)	0.216(0.174)	0.407	0.46
	AR+NV	-0.063(0.105)	0.890(0.109)	0.200(0.168)	0.408	0.48
	AR+NVNO	-0.056(0.111)	0.851(0.122)	0.281(0.154)	0.414	0.36
	AR+HAR	0.020(0.097)	0.775(0.190)	0.337(0.201)	0.445	0.20
	AR+VRB	0.046(0.110)	0.836(0.157)	0.289(0.232)	0.423	0.26
	AR+AFAR	-0.049(0.127)	0.801(0.149)	0.321(0.219)	0.377	0.24
GBP	AR	0.048(0.094)	0.906(0.145)		0.364	
	AR+PB	-0.090(0.107)	0.892(0.175)	0.199(0.134)	0.370	0.59
	AR+NV	-0.123(0.156)	0.879(0.148)	0.169(0.108)	0.367	0.54
	AR+NVNO	-0.138(0.184)	0.860(0.168)	0.234(0.178)	0.368	0.47
	AR+HAR	-0.028(0.192)	0.765(0.179)	0.361(0.214)	0.379	0.22
	AR+VRB	-0.061(0.104)	0.883(0.158)	0.257(0.140)	0.373	0.33
	AR+AFAR	-0.027(0.036)	0.801(0.169)	0.304(0.231)	0.386	0.18

Acknowledgments

We would like to thank Harrison Cheng for helpful decision. Cheng Hsiao wishes to thank the NSF of China grant # 71131008 for partial support of this research. Bauwens and Wang were supported by the contract "Projet d'Actions de Recherche Concertées" 07/12-002 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain". We also thank the referees and the editors, as well as the seminar participants at CORE, University of Aarhus, the 2011 conference in Honor of Professor Hashman Pesaran and the 32th International Symposium on Forecasting for insightful suggestions and comments.

References

- Berk, K. N. (1974): "Consistent Autoregressive Spectral Sstimates, " *Ann. Statist.* 2, 489-502.

- Brockwell, P. J. and R. A. Davis (1991): *Time Series: Theory and Methods*, 2nd Edn. New York: Springer-Verlag.
- Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.
- Baillie, R. T.(1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics* 73, 5-59.
- Baillie, R. T. and C. Morana (2009): "Modeling Long Memory and Structural Breaks in Conditional Variances: an Adaptive FIGARCH Approach," *Journal of Economic Dynamics and Control*, 33, 1577-1592.
- Beran, J. and N. Terrin. 1996. "Testing for a change of the long-memory parameter," *Biometrika*, 83, 627-638.
- Bollerslev, T. and H. O. Mikkelsen (1996):" Modeling and Pricing Long memory In Stock Market Volatility," *Journal of Econometrics*, 73, 151-184.
- Chen, X., Hansen, L.P., Carrasco, M., (1999): "Nonlinearity and Temporal Dependence, Manuscript, Department of Economics, University of Chicago.
- Clements, M.P. and D.F. Hendry (1998): "Forecasting Economic Time Series". Cambridge University Press, Cambridge.
- Choi, K., Yu, W.C., Zivot, E., (2010): "Long Memory versus Structural Breaks in Modeling and forecasting Realized Volatility", *Journal of International Money and Finance*, 29, 857-875.
- Crator, N. and Ray, B. (1996): "Model Selection and Forecasting for Long-Range Dependent Processes," *Journal of Forecasting*, 15, 107-125.
- Ding, Z., Granger, C. W. J. and Engle, R. F. (1993): "A Long Memory Property of Stock Mark Returns and a New Model," *Journal of Empirical Finance* 1, 83-106.
- Diebold, F. X. and A. Inoue (2001): "Long Memory and Regime Switching," *Journal of Econometrics* 105, 131-159.
- Granger, C.W.J., (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics* 14, 227-238.
- Granger, C.W.J. and Terasvirta, T. (1999): "A Simple Nonlinear Time Series Model with Misleading Linear Properties," *Economics Letters* 62, 161-165.
- Granger, C.W.J. and N. Hyung (2004): "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns." *Journal of Empirical Finance*, 11, 399-421.

- Hosking, J. R. M. (1996): " Asymptotic Distributions of the Sample Mean, Autocovariances, and Autocorrelations of Long-Memory Time Series." *Journal of Econometrics* 73, 261-284.
- Hyung, N. and Franses, P. (2001): " Structural breaks and long memory in US inflation rates: Do they matter for forecasting? " *Econometric Institute Research Report*, EI-2001-13.
- Hansen, P.H. and A. Timmermann (2012): " Choice of Sample Split in Out-of-Sample Forecast Evaluation. " *EUI Working Paper ECO 2012/10*.
- Kuan, C.-M. and C.-C. Hsu (1998): " Change-Point Estimation of Fractionally Integrated Processes." *Journal of Time Series Analysis* 19, 693-708.
- Mallows, C. L. (1973): " Come Comments on C_p , " *Technometrics*, 15, 661-675.
- Mincer, J. and V. Zarnowitz. (1969): " The Evaluation of Economic Forecasts, " in *Economic Forecasts and Expectations*, edited by W.F.Sharpe and C.M. Cooter. New Jersey: Prentice-Hall.
- Nunes, L. C., C.-M. Kuan and P. Newbold (1995): " Spurious break." *Econometric Theory* 11, 736-749.
- Pesaran, M.H. and A. Timmermann (2005): " Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks, " *Journal of Econometrics* 129, 183-217.
- Pesaran, M.H. and A. Timmermann (2007): " Selection of Estimation Window in the Presence of breaks, " *Journal of Econometrics* 129, 183-217.
- Stock, J. H. and M. W. Watson (1996): " Evidence on Structural Instability in Macroeconomic time Series Reactions." *Journal of Business and Economics Statistics* 14, 11-30.
- Tiao, G. C. and R. S. Tsay (1994): " Some Advances in Non Linear and Adaptive Modelling in Time Series." *Journal of Forecasting* 13(2), 109-148.
- Wang, S. H. and C. Hsiao (2012): " Testing Uncorrelatedness Between Two Memory Processes. " 3rd Revision .

APPENDIX ⁸

A.1. Proof of Lemma 1

We follow Granger (1980) and Diebold and Inoue (2001) to define a process η_t as long memory if the rate of growth of variances of partial sums $Var(S_T) = O(T^{2d+1})$, where $-0.5 < d < 0.5, d \neq 0$, and $S_T = \sum_{t=1}^T \eta_t$.

⁸ The detailed proofs are available on request.

By Theorem 13.2.1 of Brockwell and Davis (1991), we know that

$$\eta_t = \mu_1 + (1 - L)^{-d_1} e_t = \mu_1 + \sum_{v=0}^{\infty} \psi_v e_{t-v}, \quad E\eta_t = \mu_1, \quad \text{as } 1 \leq t \leq T_1,$$

$$\text{and } \eta_t = \mu_2 + (1 - L)^{-d_2} e_t = \mu_2 + \sum_{s=0}^{\infty} \psi_s^* e_{t-s}, \quad E\eta_t = \mu_2, \quad \text{as } T_1 + 1 \leq t \leq T_2$$

where $\psi_v \sim v^{d_1-1}$ and $\psi_s^* \sim s^{d_2-1}$.

$$\begin{aligned} \text{Var}\left(\sum_{t=1}^T \eta_t\right) &= \text{Var}\left(\sum_{t=1}^{T_1} \eta_t + \sum_{t=T_1+1}^T \eta_t\right) = \text{Var}\left(\sum_{t=1}^{T_1} \eta_t\right) + \text{Var}\left(\sum_{t=T_1+1}^T \eta_t\right) \\ &+ 2 \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T \text{Cov}(\eta_t, \eta_s) = (A) + (B) + (C). \end{aligned} \quad (\text{A.1})$$

To prove the asymptotic property of (A.1), we investigate the property of (A) on the right hand side of (A.1). We first let $t - v = h$ and note that

$$\begin{aligned} (A) &= \text{Var}\left(\sum_{t=1}^{T_1} \left(\mu_1 + \sum_{v=0}^{\infty} \psi_v e_{t-v}\right)\right) = \text{Var}\left(\sum_{t=1}^{T_1} \sum_{v=0}^{\infty} \psi_v e_{t-v}\right) = E\left(\sum_{t=1}^{T_1} \sum_{v=0}^{\infty} \psi_v e_{t-v}\right)^2 \\ &= \sum_{t=1}^{T_1} E\left(\sum_{v=0}^{\infty} \psi_v e_{t-v}\right)^2 + \sum_{t=1}^{T_1-1} \sum_{s=t+1}^{T_1} E\left(\left(\sum_{v=0}^{\infty} \psi_v e_{t-v}\right)\left(\sum_{v=0}^{\infty} \psi_v e_{s-v}\right)\right) \\ &= O(T_1^{2d_1+1}) = O(\kappa^{2d_1+1} T^{2d_1+1}). \end{aligned} \quad (\text{A.2}).$$

Similarly, it follows that

$$(B) = \text{Var}\left(\sum_{t=T_1+1}^T \eta_t\right) = O_p(T_2^{2d_2+1}) = O((1 - \kappa)^{2d_2+1} T^{2d_2+1}) \quad (\text{A.3})$$

Let $s' = s/T$ and $t' = t/T$, we further show that

$$\begin{aligned} \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T \text{Cov}(\eta_t, \eta_s) &= \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T \text{Cov}\left(\mu_1 + \sum_{v=0}^{\infty} \psi_v e_{t-v}, \mu_2 + \sum_{l=0}^{\infty} \psi_l^* e_{s-l}\right) \\ &= \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T \text{Cov}\left(\sum_{v=0}^{\infty} \psi_v e_{t-v}, \sum_{l=0}^{\infty} \psi_l^* e_{s-l}\right) = \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T E\left(\left(\sum_{v=0}^{\infty} \psi_v e_{t-v}\right)\left(\sum_{l=0}^{\infty} \psi_l^* e_{s-l}\right)\right) \\ &= \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T \sum_{v=0}^{\infty} \psi_{(s-t)+v} \psi_v^* \sigma^2 \sim \sigma^2 \sum_{t=1}^{T_1} \sum_{s=T_1+1}^T O((s-t)^{d_1+d_2-1}) \\ &\sim \sigma^2 T^{d_1+d_2+1} \int_0^\kappa \int_\kappa^1 (s' - t')^{d_1+d_2-1} ds' dt' \\ &= O\left(T^{d_1+d_2+1} ((1 - \kappa)^{d_1+d_2+1} - \kappa^{d_1+d_2+1})\right) \end{aligned} \quad (\text{A.4})$$

It follows from (A.2), (A.3) and (A.4) that

$$O(\text{Var}(S_T)) = O((\kappa T)^{2d_1+1} + ((1-\kappa)T)^{2d_2+1}) = O(T^{2d^*+1}) \quad (\text{A.5}).$$

where $d^* = \lambda d_1 + (1-\lambda)d_2$, $0 \leq \lambda \leq 1$ as T is finite. When $T \rightarrow \infty$, $O(\text{Var}(S_T))$ is dominated by $T^{\max(d_1, d_2)}$. Therefore, $d^* \rightarrow \max(d_1, d_2)$.

(A.5) shows that our basic model can be represented by a long memory with a long memory parameter d^* with mean

$$E(\eta_t) = \kappa\mu_1 + (1-\kappa)\mu_2 = \mu^* \quad (\text{A.6})$$

By (A.5) and (A.6), the Lemma 1 can be established.

A.2. Proof of Theorem 1

To prove Theorem 1, by Lemma 1, we first consider the constant term, β_0 , in our AR approximation procedure. From Lemma 1, we have

$$(1-L)^{d^*}(\eta_t - \mu^*) = e_t \text{ and } \eta_t = (1-L)^{d^*}\mu^* + \sum_{j=1}^{\infty} \beta_j \eta_{t-j} + e_t.$$

where $\beta_j \sim Cj^{-d^*-1}$ and $d^* \leq \max\{d_1, d_2\}$.

We define $\beta(k) = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)$, $\beta(k+1, \infty) = (\beta_{k+1}, \beta_{k+2}, \dots)$

$$\zeta_k = \begin{bmatrix} 1 & \eta_{(k+1)-1} & \eta_{(k+1)-2} & \cdots & \eta_{(k+1)-k} \\ 1 & \eta_{(k+2)-1} & \eta_{(k+2)-2} & \cdots & \eta_{(k+2)-k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \eta_{T_1-1} & \eta_{T_2-2} & \cdots & \eta_{T_1-k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \eta_{T-1} & \eta_{T-2} & \cdots & \eta_{T-k} \end{bmatrix}_{(T-k) \times (k+1)} = (\tilde{\eta}_0, \tilde{\eta}_1, \tilde{\eta}_2, \dots, \tilde{\eta}_k),$$

$$\zeta_{k+} = \begin{bmatrix} \eta_{(k+1)-(k+1)} & \eta_{(k+1)-(k+2)} & \cdots & \cdots \\ \eta_{(k+2)-(k+1)} & \eta_{(k+2)-(k+2)} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{T_1-(k+1)} & \eta_{T_1-(k+2)} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{T-(k+1)} & \eta_{T-2} & \cdots & \cdots \end{bmatrix} = (\tilde{\eta}_{k+1}, \tilde{\eta}_{k+2}, \dots),$$

$$\varepsilon_{T-k} = (e_{k+1}, e_{T+2}, \dots, e_T)', \tilde{\eta} = (\eta_{k+1}, \eta_{k+2}, \dots, \eta_T)'.$$

where ε_{T-k} , $\tilde{\eta}$, and all $\tilde{\eta}_1, \tilde{\eta}_2, \dots, \tilde{\eta}_{k+1}, \dots$ are all $(T-k) \times 1$ vectors respectively.

The true model is $\tilde{\eta} = \zeta_k \beta(k)' + \zeta_{k+} \beta(k+1, \infty)' + \varepsilon_{T-k}$. Because we cannot estimate an infinite number of parameters from a finite sample of T observations, we then consider the following k th order autoregressive model approximation:

$$\tilde{\eta} = \zeta_k \beta(k)' + \varepsilon_k. \quad (A.7)$$

where $\varepsilon_k = (e_{k+1,k}, e_{k+2,k}, \dots, e_{T,k})'$, $e_{t,k}, t = k+1, k+2, \dots, T$ is the residual derived from fitting η_t by an AR(k) approximation.

The adequacy of the approximate model for the true model depends on the choice of k . Let $M_k = I - \zeta_k (\zeta_k' \zeta_k)^{-1} \zeta_k'$ be an $(T-k) \times (T-k)$ symmetric, idempotent matrix with the property that $M_k \zeta_k = 0$, then the OLS estimate of the residual of (A.7) $\hat{\varepsilon}_k$ can be written as,

$$\hat{\varepsilon}_k = M_k \left(\varepsilon_{T-k} + \sum_{h=1}^k \tilde{\eta}_h \beta_h + \sum_{h=k+1}^{\infty} \tilde{\eta}_h \beta_h \right) = M_k \zeta_{k+} \beta(k+1, \infty)' + M_k \varepsilon_{T-k}.$$

We wish to show that when $-0.5 < d^* < 0.5$,

$$\begin{aligned} \hat{Q} &= T^{-1} \hat{\varepsilon}_k' \hat{\varepsilon}_k = T^{-1} (\beta(k+1, \infty) \zeta_{k+}' M_k \zeta_{k+} \beta(k+1, \infty)' \\ &\quad + \beta(k+1, \infty) \zeta_{k+}' M_k \varepsilon_{T-k} + \varepsilon_{T-k}' M_k \zeta_{k+} \beta(k+1, \infty)' + \varepsilon_{T-k}' M_k \varepsilon_{T-k}) \\ &= T^{-1} \beta(k+1, \infty) \zeta_{k+}' M_k \zeta_{k+} \beta(k+1, \infty)' + T^{-1} \beta(k+1, \infty) \zeta_{k+}' M_k \varepsilon_{T-k} \\ &\quad + T^{-1} \varepsilon_{T-k}' M_k \zeta_{k+} \beta(k+1, \infty)' + T^{-1} \varepsilon_{T-k}' \varepsilon_{T-k} - T^{-1} \varepsilon_{T-k}' (\zeta_k (\zeta_k' \zeta_k)^{-1} \zeta_k') \varepsilon_{T-k} \\ &= (I) + (II) + (III) + (IV) + (V) = T^{-1} \varepsilon_{T-k}' \varepsilon_{T-k} + o_p(1). \end{aligned} \quad (A.8)$$

That is, we need to find out a condition on growth rate of lag order k to ensure that the first term (I), the second term (II), the third term (III) and the fifth term (V) of (A.8) converge to zero when $-0.5 < d^* < 0.5, d^* \neq 0$.

We first investigate the asymptotic properties of (A.8) for the case of $-0.5 < d < 0$. The first term (I) on the right hand side of (A.8) can be written as

$$\begin{aligned} (I) &= T^{-1} (\beta(k+1, \infty) \zeta_{k+}' M_k \zeta_{k+} \beta(k+1, \infty)') \\ &= T^{-1} \{ \beta(k+1, \infty) \zeta_{k+}' I \zeta_{k+} \beta(k+1, \infty)' \} \\ &\quad - T^{-1} \{ \beta(k+1, \infty) \zeta_{k+}' \zeta_k (\zeta_k' \zeta_k)^{-1} \zeta_k' \zeta_{k+} \beta(k+1, \infty)' \} = A - B, \end{aligned} \quad (A.9)$$

where A and B denote the first and second term of the righthand side of (A.9), respectively. It can be shown that

$$\begin{aligned} \|A\|^2 &= \|T^{-1} \beta(k+1, \infty) \zeta_{k+}' I \zeta_{k+} \beta(k+1, \infty)'\|^2 = O_p(k^{-4d^* - 2} T^{4d^*}) \\ \text{and } \|A\| &= O_p(k^{-2d^* - 1} T^{2d^*}). \end{aligned} \quad (A.10)$$

To show (A.10), we note that for any conformable matrix A^* and B^* ,

$$\|A^*B^*\|^2 \leq \|A^*\|^2 \|B^*\|^2.$$

$$\begin{aligned} \text{therefore } \|A\|^2 &= \|T^{-1}\beta(k+1, \infty)\zeta'_{k+}I\zeta_{k+}\beta(k+1, \infty)'\|^2 \\ &\leq T^{-2}\|\beta(k+1, \infty)\zeta'_{k+}\|^2\|\zeta_{k+}\beta(k+1, \infty)'\|^2 \\ &= T^{-2}\left(\sum_{t=k+1}^T\left(\sum_{j=k+1}^{\infty}\beta_j\eta_{t-j}\right)^2\right)^2 = O_p(k^{-4d^*-2}T^{4d^*}) \end{aligned} \quad (\text{A.11})$$

We note that

$$\begin{aligned} \mathbf{E}\left(\sum_{j=k+1}^{\infty}\beta_j\eta_{t-j}\right)^2 &= \sum_{j=k+1}^{\infty}\beta_j^2\mathbf{E}(\eta_{t-j}^2) + \sum_{j=k+1}^{\infty}\sum_{v=j+1}^{\infty}\beta_j\beta_v\mathbf{E}(\eta_{t-v,i}\eta_{t-j,i}) \\ &\leq \sum_{j=k+1}^{\infty}\beta_j^2\gamma_0 + \sum_{j=k+1}^{\infty}\sum_{v=j+1}^{\infty}\beta_j\beta_v|\gamma_{v-j}|. \end{aligned} \quad (\text{A.12})$$

Since Lemma 1 indicates that the complete time series can be represented by another long memory process with the memory parameter d^* , we note that $\beta_j^2 \sim C^2j^{-2d^*-2}$, C and C_2 are finite constants. Following the proof of Hosking (1996, p.277) and letting $x = \frac{s}{m}$, we can show that (A.12) is bounded by

$$\begin{aligned} &\sum_{j=k+1}^{\infty}\beta_j^2\gamma_0 + |C_2|\sum_{j=k+1}^{\infty}\sum_{v=j+1}^{\infty}|\beta_j||\beta_v||v-j|^{2d^*-1} \\ &\leq O_p(k^{-2d^*-1}) + k^{-d^*-1}\sum_{j=k+1}^{\infty}\sum_{v=j+1}^{\infty}|v-j|^{2d^*-1}v^{-d^*-1} \\ &= O_p(k^{-2d^*-1}) + k^{-d^*-1}\sum_{j=k+1}^{\infty}\sum_{s=1}^{\infty}s^{2d^*-1}(s+j)^{-d^*-1} \\ &\leq O_p(k^{-2d^*-1}) + k^{-2d^*-1}\sum_{s=1}^{\infty}s^{2d^*-1} \\ &\sim O_p(k^{-2d^*-1}) + k^{-2d^*-1}T^{2d^*}\int_0^1x^{2d^*-1}dx \quad \text{as } T \rightarrow \infty \\ &= O_p(k^{-2d^*-1}), \end{aligned} \quad (\text{A.13})$$

where $\gamma_\tau = E(\eta_t\eta_{t+\tau}) \sim C_2\tau^{2d^*-1}$, C_2 are finite constants.

It follows that,

$$\sum_{j=k+1}^{\infty}\beta_j\eta_{t-j} = O_p(k^{-d^*-0.5}). \quad (\text{A.14})$$

Similarly, we can show that

$$\|B\| \leq \|T^{-1}\beta(k+1, \infty)\zeta'_{k+}\| \|\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_{k+}\beta(k+1, \infty)'\|, \quad (\text{A.15})$$

where by (A.13) and (A.14),

$$\begin{aligned} \|T^{-1}\beta(k+1, \infty)\zeta'_{k+}\|^2 &= T^{-2} \sum_{t=k+1}^T \left(\sum_{j=k+1}^{\infty} \beta_j \eta_{t-j} \right)^2 = O_p(k^{-2d^*-1}T^{-1}) \\ \text{and } \|T^{-1}\beta(k+1, \infty)\zeta'_{k+}\| &= O_p(k^{-d^*-0.5}T^{-0.5}). \end{aligned} \quad (\text{A.16})$$

$$\begin{aligned} &\|\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_{k+}\beta(k+1, \infty)'\|^2 \\ &= \text{tr} \left(\beta(k+1, \infty)\zeta'_{k+}\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_{k+}\beta(k+1, \infty)' \right) \\ &= \sum_{t=k+1}^T \left(\sum_{j=k+1}^{\infty} \beta_j \eta_{t-j} \right)^2 = O_p(T^{-1}k^{-2d^*-1}). \end{aligned} \quad (\text{A.17})$$

$$\text{and } \|\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_{k+}\beta(k+1, \infty)'\| = O_p(T^{0.5}k^{-d^*-0.5}). \quad (\text{A.18})$$

Thus, by (A.16) and (A.18), $B = O_p(k^{-2d^*-1})$. (A.19)

Furthermore, (A.11) and (A.19) imply that

$$(I) = T^{-1}(\beta(k+1, \infty)\zeta'_{k+}M_k\zeta_{k+}\beta(k+1, \infty)') = O_p(k^{-2d^*-1}). \quad (\text{A.20})$$

The second term (II) on the right side of (A.8),

$$\begin{aligned} (II) &= T^{-1}\beta(k+1, \infty)\zeta'_{k+}M_k\varepsilon_{T-k} = T^{-1}\beta(k+1, \infty)\zeta'_{k+}\varepsilon_{T-k} \\ &\quad - T^{-1}\beta(k+1, \infty)\zeta_{k+}\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\varepsilon_{T-k} = C - D, \end{aligned} \quad (\text{A.21})$$

where C, D denote the first and second term of the righthand side of (A.21), respectively. Noting that

$$\begin{aligned} \|T^{-1}\beta(k+1, \infty)\zeta'_{k+}\varepsilon_{T-k}\|^2 &= T^{-2}\|\beta(k+1, \infty)\zeta'_{k+}\|^2\|\varepsilon_{T-k}\|^2 \\ &= O_p(T^{-1}k^{-2d^*-1}), \end{aligned} \quad (\text{A.22})$$

it follows that $C = O_p(T^{-0.5}k^{-d^*-0.5})$. (A.23)

Similarly, by (A.21), we can show that

$$\begin{aligned} \|D\|^2 &= \|T^{-1}\beta(k+1, \infty)\zeta_{k+}\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\varepsilon_{T-k}\|^2 \\ &= T^{-2}\text{tr} \left(\varepsilon_{T-k}\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta'_{k+}\beta(k+1, \infty)'\beta(k+1, \infty)\zeta_{k+}\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\varepsilon_{T-k} \right) \\ &= O_p(T^{-1}k^{-2d^*-1}). \end{aligned}$$

$$\text{Thus, } (II) = O_p(T^{-0.5}k^{-d^*-0.5}). \quad (\text{A.24})$$

Likewise, the third term (III) on the right side of (A.7),

$$(III) = \varepsilon'_{T-k} M_k \zeta_{k+\beta}(k+1, \infty) = O_p(T^{-0.5}k^{-d^*-0.5}). \quad (\text{A.25})$$

To calculate the asymptotic properties of the fifth term (V) on the right side of (A.8), we further denote that

$$\begin{aligned} \widehat{\Omega}_k &= \frac{\zeta'_k \zeta_k}{T-k} \\ &= \begin{bmatrix} \frac{1}{T-k} \sum_{i=k+1}^T \eta_{i-1}^2 & \frac{1}{T-k} \sum_{i=k+1}^T \eta_{i-1} \eta_{i-2} & \cdots & \frac{1}{T-k} \sum_{i=k+1}^T y_{i-1} y_{i-k} \\ \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{T-k} \sum_{i=k+1}^T y_{i-k} y_{i-1} & \cdots & \cdots & \frac{1}{T-k} \sum_{i=k+1}^T \eta_{i-k}^2 \end{bmatrix} \end{aligned}$$

and

$$\Omega_k = \mathbf{E}(\zeta'_k \zeta_k) = \begin{bmatrix} \mathbf{E} \eta_{i-1}^2 & \mathbf{E} \eta_{i-1} \eta_{i-2} & \cdots & \mathbf{E} \eta_{i-1} \eta_{i-k} \\ \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{E} \eta_{i-k} \eta_{i-1} & \cdots & \cdots & \mathbf{E} \eta_{i-k}^2 \end{bmatrix}.$$

By the Theorem 3 of Hosking (1996) and the Theorem 4.1 of Poskitt (2007), we can show that

$$\|\widehat{\Omega}_k - \Omega_k\| = O_p(k(\log T/T)^{0.5-d^*}), \quad d^* \in (-0.5, 0.5), d^* \neq 0, \quad (\text{A.26})$$

$$\begin{aligned} \text{Thus, } \|\widehat{\Omega}_k^{-1}\| &\leq \|\widehat{\Omega}_k^{-1} - \Omega_k^{-1}\| + \|\Omega_k^{-1}\| = \|\widehat{\Omega}_k^{-1}(\Omega_k - \widehat{\Omega}_k)\Omega_k^{-1}\| + \|\Omega_k^{-1}\| \\ &= \|\widehat{\Omega}_k^{-1}\| \|\Omega_k - \widehat{\Omega}_k\| \|\Omega_k^{-1}\| + \|\Omega_k^{-1}\| \end{aligned}$$

$$\text{and } \|\Omega_k^{-1}\| < \infty,$$

because of a nonsingular matrix Ω_k and by Lemma 5.7 of Poskitt (2007) and Proposition 4.5.3. of Brockwell and Davis (1991). Hence, by (A.26),

$$\|\widehat{\Omega}_k^{-1}\| \leq \frac{\|\Omega_k^{-1}\|}{(1 - \|(\Omega_k - \widehat{\Omega}_k)\| \|\Omega_k^{-1}\|)} = O_p(k^{-1}(\log T/T)^{d^*-0.5}). \quad (\text{A.27})$$

Thus, by (A.27) the fifth term (V) could be

$$\begin{aligned} \|T^{-1} \varepsilon'_{T-k} \zeta_k (\zeta'_k \zeta_k)^{-1} \zeta'_k \varepsilon_{T-k}\|^2 &= \|T^{-1} \varepsilon'_{T-k} \left(\zeta_k \left(\frac{\zeta'_k \zeta_k}{T-k} \right)^{-1} \zeta'_k \right)' \frac{1}{T-k} \varepsilon_{T-k}\|^2 \\ &\leq \|T^{-1} \varepsilon_{T-k} \zeta_k\|^2 \left\| \left(\frac{\zeta'_k \zeta_k}{T} \right)^{-1} \right\|^2 \left\| \frac{1}{T} \varepsilon_T \zeta'_k \right\|^2 \\ &= O_p(kT^{-1}) O_p(k^{-2}(\log T/T)^{-1+2d^*}) O_p(kT^{-1}) \\ &= O_p((\log T)^{-1+2d^*} T^{-2d^*-1}). \end{aligned} \quad (\text{A.28})$$

Thus by (A.28), we note that

$$\begin{aligned}\frac{1}{T}\varepsilon'_{T-k}M_k\varepsilon_{T-k} &= \frac{1}{T}\varepsilon'_{T-k}\varepsilon_{T-k} - \frac{1}{T}\varepsilon'_{T-k}(\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\varepsilon_{T-k}) \\ &= \frac{1}{T}\varepsilon'_{T-k}\varepsilon_{T-k} - O_p((\log T)^{d^*-0.5}T^{-d^*-0.5}).\end{aligned}\quad (\text{A.29})$$

For the case of $0 < d < 0.5$, we reprove the (A.12) and (A.13) as

$$\mathbf{E} \left(\sum_{j=k+1}^{\infty} \beta_j \eta_{t-j} \right)^2 = O_p(k^{-2d^*-1}T^{2d^*}). \quad (\text{A.30})$$

Therefore, by (A.30) and following the similar proof procedure for the case of $-0.5 < d < 0$, when $0 < d < 0.5$,

$$(I) = T^{-1}(\beta(k+1, \infty)\zeta'_{k+}M_k\zeta_{k+}\beta(k+1, \infty)') = O_p(k^{-2d^*-1}T^{2d^*}), \quad (\text{A.31})$$

$$(II) = T^{-1}\beta(k+1, \infty)\zeta'_{k+}M_k\varepsilon_{T-k} = O_p(k^{-d^*-0.5}T^{d^*-0.5}), \quad (\text{A.32})$$

$$(III) = \varepsilon'_{T-k}M_k\zeta_{k+}\beta(k+1, \infty) = O_p(T^{d^*-0.5}k^{-d^*-0.5}). \quad (\text{A.33})$$

Therefore, when we restrict k on the growth rate of $T^r, r > 2d^*/(1+2d^*)$, i.e. $k = O(T^r), r > 2d^*/(1+2d^*)$, (A.20), (A.24), (A.25), (A.28), (A.31), (A.32) and (A.33) converge to zero. It follows that (I), (II), (III) and (V) converge to zero when $-0.5 < d^* < 0.5$. Hence, when $k = O(T^r), r > 2d^*/(1+2d^*)$ and $-0.5 < d^* < 0.5$,

$$\widehat{Q} = T^{-1}\varepsilon'_{T-k}\varepsilon_{T-k} + o_p(1). \quad (\text{A.34})$$

Furthermore, by (A.16), (A.27), (A.28) and (A.30), we note that when

$$\begin{aligned}\|\beta(k) - \beta\| &\leq \|\zeta_k(\zeta'_k\zeta_k)^{-1}\zeta'_k\zeta_{k+}\beta(k+1, \infty)'\| \\ &+ \|(\zeta'_k\zeta_k)^{-1}\zeta'_k\varepsilon_{T-k}\| = O_p(k^{-1/2}(\log T)^{d^*-0.5}T^{-d^*}).\end{aligned}\quad (\text{A.35})$$

A.3. Proof of Theorem 2

By Lemma 1, Brockwell and Davis (1991) and Poskitt (2007), we denote the h -step best linear prediction of η_t as

$$\eta_{t+h}^* = \mu^* + \sum_{j=h}^{\infty} \psi_j e_{t+h-j}, \quad \text{where } \psi \sim j^{d^*-1}. \quad (\text{A.36})$$

$$\text{or } \eta_{t+h}^* = (1-L)^{d^*}\mu^* + \sum_{j=1}^k \beta_{jk}\eta_{t+h-j}, \quad \text{where } \pi \sim j^{-d^*-1}. \quad (\text{A.37})$$

Thus, by (7), Theorem 1, (A.36) and (A.37), we note that the h -step ahead mean squared error of the forecast $\hat{\eta}_{T+h}$ based on the $AR(k)$ approximation model is

$$\begin{aligned} MSE(h)^{AR} &= E(\eta_{T+h} - \hat{\eta}_{T+h})^2 = E(\eta_{T+h} - \eta_{T+h}^* + \eta_{T+h}^* - \hat{\eta}_{T+h})^2 \\ &= E(\eta_{T+h} - \eta_{T+h}^*)^2 + E(\eta_{T+h}^* - \hat{\eta}_{T+h})^2 + 2E((\eta_{T+h} - \eta_{T+h}^*)(\eta_{T+h}^* - \hat{\eta}_{T+h})) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \\ &\quad + (1-L)^{d^*} \kappa^2 (\mu_1 - \mu_2)^2 + O_p(k^{-1}T^{-2d^*}(\log T)^{2d^*-1}) + O_p(k^{-1/2}T^{-d^*}(\log T)^{d^*-0.5}). \end{aligned} \quad (A.38)$$

Because $\pi_{i,j} = \frac{\Gamma(j-d_i)}{\Gamma(j+1)\Gamma(d_i)}$, $i = 1, 2$ and $\pi_j^* = \frac{\Gamma(j-d^*)}{\Gamma(j+1)\Gamma(d^*)}$. By the Stirling's formula, we note that as $j \rightarrow \infty$,

$$\pi_{i,j} \sim j^{-d_i-1}/\Gamma(-d_i), i = 1, 2, \quad \text{and} \quad \pi_j^* \sim j^{-d^*-1}/\Gamma(-d^*), \quad (A.39)$$

thus

$$\begin{aligned} \tilde{\pi}_j^* - \pi_j^* &= \frac{\Gamma(j - \tilde{d}^*)}{\Gamma(j+1)\Gamma(\tilde{d}^*)} - \frac{\Gamma(j - d^*)}{\Gamma(j+1)\Gamma(d^*)} \\ &\sim j^{-\tilde{d}^*-1}/\Gamma(-\tilde{d}^*) - j^{-d^*-1}/\Gamma(-d^*) \end{aligned} \quad (A.40)$$

Furthermore, it is well known that $T^{1/2}(\tilde{d}_i - d_i) \Rightarrow N(0, 6/\pi^2)$, $i = 1, 2$. (See Baillie 1996). Because $d^* \leq \max\{d_1, d_2\}$. For ease of notation, we suppose that $d^* = d_2$ and $\tilde{\pi}_j^* = \tilde{\pi}_{2,j}$. By Slutsky Theorem, $\tilde{\pi}_{2,j}$ converges to $\pi_{2,j}$. Then, by Taylor's Theorem, we note that

$$\tilde{\pi}_{2,j} - \pi_{2,j} \simeq c j^{-d_2-1} \ln j (\tilde{d}_2 - d_2) = O_p(j^{-d_2-1} \ln j T^{-1/2}) O_p(T^{-1}). \quad (A.41)$$

and letting $i - j = s$ and $T - j = u$,

$$\begin{aligned} &\mathbf{E} \left(\sum_{j=1}^T (\tilde{\pi}_{2,j} - \pi_{2,j}) \eta_{T+h-j} \right)^2 \\ &= \sum_{j=1}^T (\tilde{\pi}_{2,j} - \pi_{2,j})^2 E(\eta_{T+h-j})^2 + \sum_{j=1}^{T-1} \sum_{i=j+1}^T (\tilde{\pi}_{2,j} - \pi_{2,j})(\tilde{\pi}_{2,i} - \pi_{2,i}) \gamma_{i-j} \\ &\cong (\tilde{d}_2 - d_2)^2 \sigma_\eta^2 \sum_{j=1}^T (j^{-d_2-1} \ln j)^2 + \sum_{j=1}^{T-1} \sum_{i=j+1}^T (\tilde{d}_2 - d_2)^2 (j^{-d_2-1} \ln j) (i^{-d_2-1} \ln i) (i-j)^{2d_2-1} \\ &= C(6/\pi^2) \sigma_\eta^2 \frac{(\ln T)^2}{T} + \frac{C(6/\pi^2)}{T} \sum_{j=1}^{T-1} j^{-d_2-1} \ln j ((T-j)^{d_2-1} \ln(T-j)) \\ &= C(6/\pi^2) \sigma_\eta^2 \frac{(\ln T)^2}{T} + O_p\left(\frac{(\ln T)^2}{T}\right) \end{aligned} \quad (A.42)$$

Thus, by (A.37), (A.38) and (A.42), the h -step mean squared error of the NVNO forecast $\tilde{\eta}_{T+h}$ given by (8) is

$$\begin{aligned} MSE(h)^{NVNO} &= E(\eta_{T+h} - \tilde{\eta}_{T+h})^2 = E(\eta_{T+h} - \eta_{T+h}^* + \eta_{T+h}^* - \tilde{\eta}_{T+h})^2 \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + (1-L)^{d^*} \kappa^2 (\mu_1 - \mu_2)^2 + C(6/\pi^2) \sigma_\eta^2 \frac{(\ln T)^2}{T} + O_p\left(\frac{(\ln T)^2}{T}\right) \\ &\quad + C(6/\pi^2)^{1/2} \sigma_\eta \frac{\ln T}{T^{1/2}} + O_p\left(\frac{\ln T}{T^{1/2}}\right). \end{aligned} \quad (A.43)$$

and that of the post-break forecast $\hat{\eta}_{T+h}^*$ given by (10) is

$$\begin{aligned} MSE(h)^{PB} &= E(\eta_{T+h} - \hat{\eta}_{T+h}^*)^2 = E(\eta_{T+h} - \eta_{T+h}^* + \eta_{T+h}^* - \hat{\eta}_{T+h}^*)^2 \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + C(6/\pi^2) \sigma_\eta^2 \frac{(\ln T_2)^2}{T_2} + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) \\ &\quad + C(6/\pi^2)^{1/2} \sigma_\eta \frac{\ln T_2}{T_2^{1/2}} + O_p\left(\frac{\ln T_2}{T_2^{1/2}}\right). \end{aligned} \quad (A.44)$$

Furthermore, to prove the h -step mean squared error of the NV method, We first obtain that as $\pi_j^* = \pi_{2,j}$,

$$\begin{aligned} &\mathbf{E}\left(\sum_{j=T_2+1}^T (\tilde{\pi}_j^* - \pi_{2,j}) \eta_{T+h-j}\right)^2 \\ &\leq C(6/\pi^2) \sigma_y^2 \frac{(\ln T_2)^2}{T_2} + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) = O_p((\ln T_2)^2 T_2^{-1}) \end{aligned} \quad (A.45)$$

By Hosking (1981), we note that $(1-L)^{d_3} = C$ as $d_3 > 0$ and $(1-L)^{d_3} = O(T^{-d_3})$ as $d_3 < 0$, where $d_3 = d_2 - d_1$. Then following the analysis of (A.42), we can get

$$\begin{aligned} &E\left(\sum_{j=T_2+1}^T (1-L)^{d_3} \pi_{2,j} \mu_1\right)^2 \leq \mu_1^2 (1-L)^{2d_3} \mathbf{E}\left(\sum_{j=T_2+1}^T \pi_{2,j}\right)^2 \\ &\leq \mu_2^2 \left(\sum_{j=T_2+1}^T \mathbf{E}(j^{-d_2-1})^2 + \sum_{j=T_2+1}^T \sum_{i=j+1}^{T-1} \mathbf{E}(j^{-d_2-1} i^{-d_2-1})\right) \\ &= C\mu_1^2 O(T_2^{-2d_3-2d_2-1}) \end{aligned} \quad (A.46)$$

$$\text{and } E\left(\sum_{j=T_2+1}^T \pi_{2,j} \mu_2\right)^2 \leq \mu_2^2 \mathbf{E}\left(\sum_{j=T_2+1}^T \pi_{2,j}\right)^2 = C\mu_1^2 O(T_2^{-2d_2-1}) \quad (A.47)$$

Following the analysis of (A.42) and (A.46),

$$\mathbf{E}(\sum_{j=T_2+1}^T (\tilde{\pi}_j^* - (1-L)^{d_3} \pi_{2,j}) \eta_{T+h-j})^2 = O_p((\ln T)^2 T^{-1}). \quad (\text{A.48})$$

Moreover, by Baillie (1996), we note that $\tilde{\mu}_i - \mu_i = O_p(T^{-1/2})$, $i = 1, 2$ (A.49). Because $\mu^* \neq \mu_2$, when $d^* = d_2$, we note that as $T \rightarrow \infty$,

$$\begin{aligned} & \mathbf{E}((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}^*} \tilde{\mu}^*)^2 \\ &= \mathbf{E}((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}^*} (\kappa \tilde{\mu}_1 + (1-\kappa) \tilde{\mu}_2))^2 \\ &= \kappa^2 (1-L)^{2d_2} (\mu_2 - \mu_1)^2. \end{aligned} \quad (\text{A.50})$$

Then by (A.36), (A.42), (A.46), (A.47), (A.48), (A.49) and (A.50), when $a = 1$, $(1-a) = 0$ (i.e., type II error), the mean square error of $\tilde{\eta}_{T+h}^{II}$ is

$$\begin{aligned} MSE(h)^{NA^{II}} &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + E(((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}^*} \tilde{\mu}^* - \sum_{j=1}^{T_2} (\tilde{\pi}_j^* - \pi_{2,j}) \eta_{T+1-j}) \\ &\quad - \sum_{j=T_2+1}^T (\tilde{\pi}_j^* - (1-L)^{d_3} \pi_{2,j}) \eta_{T+1-j} + \sum_{j=T_2+1}^T (1-L)^{d_3} \pi_{2,j} \mu_1 - \sum_{j=T_2+1}^T \pi_{2,j} \mu_2)^2 \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) + \kappa^2 (1-L)^{d_2} (\mu_2 - \mu_1)^2 + O_p(T_2^{-2d_2-2d_3-1}) \\ &\quad + O_p\left(\frac{(\ln T_2)}{T_2^{1/2}}\right). \end{aligned} \quad (\text{A.51})$$

Moreover, when $a = 0$, $(1-a) = 1$ (i.e., finding a break), then we use a naive ARFIMA model with the pre-and post break estimated parameters $\tilde{d}_1$ and $\tilde{d}_2$ to do forecast. Using (A.36), (A.42), (A.46), (A.47), (A.48), (A.50), the mean square error of $\tilde{\eta}_{T+h}^{II}$ is

$$\begin{aligned} MSE(h)^{NA^I} &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + E(((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}_2} \tilde{\mu}_2 - \sum_{j=1}^{T_2} (\tilde{\pi}_{2,j} - \pi_{2,j}) \eta_{T+1-j}) \\ &\quad - \sum_{j=T_2+1}^T ((1-L)^{\tilde{d}_3} \tilde{\pi}_{2,j} - (1-L)^{d_3} \pi_{2,j}) \eta_{T+1-j}) \\ &\quad + \sum_{j=T_2+1}^T ((1-L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \tilde{\mu}_1 - (1-L)^{d_3} \pi_{2,j} \mu_1) - \sum_{j=T_2+1}^T (\tilde{\pi}_{2,j} \tilde{\mu}_2 - \pi_{2,j} \mu_2))^2 \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) + O_p(T_2^{-2d_2-2d_3-1}) + O_p\left(\frac{(\ln T_2)}{T_2^{1/2}}\right). \end{aligned} \quad (\text{A.52})$$

Similarly, by (A.51), when $b = 1$, the mean square error of $\hat{\eta}_{T+1}^h$ is

$$\begin{aligned}
MSE(h)^{NA^I} &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + E(((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}_2} \tilde{\mu}_2 - \sum_{j=1}^{T_2} (\tilde{\pi}_{1,j} - \pi_{2,j}) \eta_{T+1-j}) \\
&\quad + \sum_{j=T_2+1}^T (1-L)^{d_3} \pi_{2,j} \eta_{T+1-j} - \sum_{j=T_2+1}^T (1-L)^{d_3} \pi_{2,j} \mu_1 - \sum_{j=T_2+1}^T \pi_{2,j} \mu_2)^2 \\
&= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) + O_p(T_2^{-2d_2-2d_3-1}) + O_p\left(\frac{(\ln T_2)}{T_2^{1/2}}\right)
\end{aligned} \tag{A.53}$$

and when $b = 0$ and $(1-b) = 1$, by (A.52),

$$\begin{aligned}
MSE(h)^{NA^I} &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + E(((1-L)^{d_2} \mu_2 - (1-L)^{\tilde{d}_2} \tilde{\mu}_2 - \sum_{j=1}^{T_2} (\tilde{\pi}_{2,j} - \pi_{2,j}) \eta_{T+1-j}) \\
&\quad - \sum_{j=T_2+1}^T ((1-L)^{\tilde{d}_3} \tilde{\pi}_{2,j} - (1-L)^{d_3} \pi_{2,j}) \eta_{T+1-j}) \\
&\quad + \sum_{j=T_2+1}^T ((1-L)^{\tilde{d}_3} \tilde{\pi}_{2,j} \tilde{\mu}_1 - (1-L)^{d_3} \pi_{2,j} \mu_1) - \sum_{j=T_2+1}^T (\tilde{\pi}_{2,j} \tilde{\mu}_2 - \pi_{2,j} \mu_2)^2 \\
&= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + O_p\left(\frac{(\ln T_2)^2}{T_2}\right) + O_p(T_2^{-2d_2-2d_3-1}) + O_p\left(\frac{(\ln T_2)}{T_2^{1/2}}\right).
\end{aligned} \tag{A.54}$$

By (A.38) and (A.43), it follows that when $k = O((\ln T)^{-1}(T/\ln T)^{1-2d^*})$, the convergent rate of the term of $O_p(\frac{(\ln T)^2}{T})$ in equation (A.43) is slower than that of the term of $O_p(k^{-1}T^{-2d^*}(\log T)^{2d^*-1})$ in equation (A.38), thus our AR-approximation based forecast method (AR) outperforms $ARFIMA(d^*)$ (NVNO) for Case I and Case II. Furthermore, by (A.38), (A.43), (A.51), (A.52), (A.53) and (A.54), we know that Theorem 2 can be established immediately.

A.4. Proof of Lemma 2

By Tiao and Tsay (1994), we note that for the adaptive forecasting, the maximum likelihood estimates (MLE) are asymptotically equivalent to estimate obtained by minimizing the sum of squares of h -step-ahead forecast errors. Furthermore, by Brockwell and Davis (1991) and Poskitt (2007), we also note that MLE and least square AR estimators are asymptotically equivalent. Thus, following the proof of Theorem 2, Lemma 2 can be established immediately.

Recent titles

CORE Discussion Papers

- 2012/10. Pierre PESTIEAU and Grégory PONTIERE. Optimal lifecycle fertility in a Barro-Becker economy.
- 2012/11. Catherine KRIER, Michel MOUCHART and Abderrahim OULHAJ. Neural modelling of ranking data with an application to stated preference data.
- 2012/12. Matthew O. JACKSON and Dunia LOPEZ-PINTADO. Diffusion and contagion in networks with heterogeneous agents and homophily.
- 2012/13. Claude D'ASPREMONT, Rodolphe DOS SANTOS FERREIRA and Jacques THEPOT. Hawks and doves in segmented markets: A formal approach to competitive aggressiveness.
- 2012/14. Claude D'ASPREMONT and Rodolphe DOS SANTOS FERREIRA. Household behavior and individual autonomy: An extended Lindahl mechanism.
- 2012/15. Dirk VAN DE GAER, Joost VANDENBOSSCHE and José Luis FIGUEROA. Children's health opportunities and project evaluation: Mexico's *Oportunidades* program.
- 2012/16. Giacomo VALLETTA. Health, fairness and taxation.
- 2012/17. Chiara CANTA and Pierre PESTIEAU. Long term care insurance and family norms.
- 2012/18. David DE LA CROIX and Fabio MARIANI. From polygyny to serial monogamy: a unified theory of marriage institutions.
- 2012/19. Carl GAIGNE, Stéphane RIOU and Jacques-François THISSE. Are compact cities environmentally friendly?
- 2012/20. Jean-François CARPANTIER and Besik SAMKHARADZE. The asymmetric commodity inventory effect on the optimal hedge ratio.
- 2012/21. Concetta MENDOLICCHIO, Dimitri PAOLINI and Tito PIETRA. Asymmetric information and overeducation.
- 2012/22. Tom TRUYTS. Stochastic signaling: Information substitutes and complements.
- 2012/23. Pierre DEHEZ and Samuel FEREY. How to share joint liability: A cooperative game approach.
- 2012/24. Pilar GARCIA-GOMEZ, Erik SCHOKKAERT, Tom VAN OURTI and Teresa BAGO D'UVA. Inequity in the face of death.
- 2012/25. Christian HAEDO and Michel MOUCHART. A stochastic independence approach for different measures of concentration and specialization.
- 2012/26. Xavier RAMOS and Dirk VAN DE GAER. Empirical approaches to inequality of opportunity: principles, measures, and evidence.
- 2012/27. Jacques H. DRÈZE and Erik SCHOKKAERT. Arrow's theorem of the deductible : moral hazard and stop-loss in health insurance.
- 2012/28. Luc BAUWENS and Giuseppe STORTI. Computationally efficient inference procedures for vast dimensional realized covariance models.
- 2012/29. Pierre DEHEZ. Incomplete-markets economies: The seminar work of Diamond, Drèze and Radner.
- 2012/30. Helmuth CREMER, Pierre PESTIEAU and Grégory PONTIÈRE. The economics of long-term care: a survey.
- 2012/31. Mathieu VAN VYVE, Laurence A. WOLSEY and Hande YAMAN. Relaxations for two-level multi-item lot-sizing problem.
- 2012/32. Jean-François MERTENS and Anna RUBINCHIK. Equilibria in an overlapping generations model with transfer policies and exogenous growth.
- 2012/33. Jean-François MERTENS and Anna RUBINCHIK. Pareto optimality of the golden rule equilibrium in an overlapping generations model with production and transfers.
- 2012/34. Hélène LATZER and Florian MAYNERIS. Income distribution and vertical comparative advantage. Theory and evidence.
- 2012/35. Pascal MOSSAY and Takatoshi TABUCHI. Preferential trade agreements harm third countries.
- 2012/36. Aitor CALO-BLANCO. The compensation problem with fresh starts.
- 2012/37. Jean-François CARPANTIER and Arnaud DUFAYS. Commodities volatility and the theory of storage.

Recent titles

CORE Discussion Papers - continued

- 2012/38 Jean-François CARPANTIER and Christelle SAPATA. Unfair inequalities in France: A regional comparison.
- 2012/39 Shui-Ki WAN, Shin-Huei WANG and Chi-Keung WOO. Total tourist arrival forecast: aggregation vs. disaggregation.
- 2012/40 Thierry BRECHET, Yann MENIERE and Pierre M. PICARD. The clean development mechanism in a global carbon market.
- 2012/41 Nikolaos GEORGANTZIS, Rafael MONER-COLONQUES, Vicente ORTS and José J. SEMPERE-MONERRIS. Theoretical and experimental insights on firms' internationalization decisions under uncertainty.
- 2012/42 Berardino CESI and Dimitri PAOLINI. Peer group and distance: when widening university participation is better.
- 2012/43 Arnaud DUFAYS. Infinite-state Markov-switching for dynamic volatility and correlation models.
- 2012/44 Miguel JARA, Dimitri PAOLINI and J.D. TENA. Management efficiency in football: An empirical analysis of two extreme cases.
- 2012/45 Guillaume ROELS, Philippe CHEVALIER and Ying WEI. United we stand? Coordinating capacity investment and allocation in joint ventures.
- 2012/46 Mikel BEDAYO, Ana MAULEON and Vincent VANNETELBOSCH. Bargaining and delay in trading networks.
- 2012/47 Pierre M. PICARD and Ridwan D. RUSLI. State owned firms: Private debt, cost revelation and welfare.
- 2012/48 Shin-Huei WANG, Luc BAUWENS and Cheng HSIAO. Forecasting long memory processes subject to structural breaks.

Books

- G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
- J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en héritage*. Itinera Institute.
- M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.
- V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
- I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
- W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
- L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
- J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
- M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
- J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

- R. AMIR (2002), Supermodularity and complementarity in economics.
- R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
- A. SHAPIRO (2010), Stochastic programming: modeling and theory.