Regression Model Fitting with a Long Memory Covariate Process[1]

HIRA L. KOUL[2]
Michigan State University

RICHARD T. BAILLIE
Michigan State University and Queen Mary, University of London

DONATAS SURGAILIS
Institute of Mathematics & Informatics, Vilnius

Abstract

This paper proposes some tests for fitting a regression model with a long memory covariate process and with errors that form either a martingale difference sequence, or a long memory moving average process, independent of the covariate. The tests are based on a partial sum process of the residuals from the fitted regression. The asymptotic null distribution of this process is discussed in some detail under each set of these assumptions. The proposed tests are shown to have known asymptotic null distributions in the case of martingale difference errors, and in the case of fitting a polynomial of a known degree through the origin when the errors have long memory. The theory is then illustrated with some examples based on the forward premium anomaly where a squared interest rate differential proxies a time dependent risk premium. The paper also shows that the proposed test statistic converges weakly to non-standard distributions in some cases.

1 Introduction

A discrete time, stationary stochastic process, is said to have *long memory* if its autocorrelations tend to zero hyperbolically in the lag parameter, as the lag tends to infinity. The importance of long memory processes in econometrics, hydrology and various other physical sciences is abundantly demonstrated by Beran (1992, 1994), Baillie (1996), and the references therein.

[1] *AMS (1991) Subject classification.* Primary 62M15, Secondary 62P05, 62P20.
Key words. Forward premiums and spot returns. Partial sum residual process.
[2] Research of the first two authors partly supported by the NSF Grant DMS 00-71619.

This paper investigates the problem of fitting a parametric model to the regression function, when the explanatory variable has long memory, and when the errors either form a martingale differences, or a long memory process. Several previous studies have emphasized the importance of regression model fitting; not only in econometrics, but also in more general statistical applications; see, e.g., Bates and Watts (1989) and Seber and Wild (1989). However, the analysis is often limited to the least squares methodology when the explanatory variables are either non-random or i.i.d. and/or weakly dependent.

To proceed further, let q be a positive integer, X, Y be random variables, with X representing a covariate variable, and Y the response variable with $E|Y| < \infty$. Let $\mu(x) := E(Y|X = x)$ denote the regression function, h_j, $j = 1, \cdots, q$, be some known functions of X, and let $h' := (h_1, \cdots, h_q)$. The problem of interest here is to test

$$H_0 : \mu(x) = \beta' h(x), \qquad \forall\, x \in \mathbb{R}, \text{ for some } \beta \in \mathbb{R}^q ,$$

against the alternative H_0 is not true, based on the T observations (X_t, Y_t), $t = 1, \ldots, T$. This is a classical problem and literature is full of tests that have been investigated under various weak dependence assumptions, see, e.g., the review paper by MacKinnon (1992).

Here we are interested in investigating the large sample behavior of tests based on the partial sum process

$$\hat{\mathcal{V}}_T(x) := \sum_{t=1}^{T} \Big(Y_t - \hat{\beta}' h(X_t)\Big) I(X_t \le x), \qquad x \in \mathbb{R},$$

when $\{X_t\}$ forms a long memory moving average, and when the errors $\{\varepsilon_t := Y_t - \beta' h(X_t)\}$ either form a homogeneous martingale difference sequence or a long memory moving average process. In both cases the error process $\{\varepsilon_t\}$ is assumed to be independent of the covariate process $\{X_t\}$. For the convenience of the exposition we take $\hat{\beta}$ to be the least squares estimator of β under H_0 throughout this paper. The process $\hat{\mathcal{V}}_T$ has its roots in the CUSUM process of the one sample model. It is also known as a marked process with marks being the residuals $\{Y_t - \hat{\beta}' h(X_t)\}$. In the decade of the 1990's, several authors proposed tests based on this process for regression and autoregressive model checking. See, e.g., An and Cheng (1991), Su and Wei (1991), Stute (1997), Stute, González Manteiga and Presedo Quindimil (1998), Stute, Thies and Zhu (1998), Diebolt and Zuber (1999), Koul and Stute (1999), among others. Most of these papers deal with either an i.i.d. covariate and errors setup or an autoregressive framework with i.i.d. errors.

The errors $\{\varepsilon_t\}$ are said to form a homoskedastic martingale difference sequence if

(1.1) $\{\varepsilon_t\}$ are mean zero finite variance martingale differences with $E\Big(\varepsilon_t^2 | \mathcal{F}_{t-1}\Big) \equiv \sigma_e^2$, where $\mathcal{F}_t := \sigma - \text{field}\{\varepsilon_s, s \le t\}$ and σ_e^2 is a constant.

They are said to form a long memory moving average process if

$$(1.2)\varepsilon_t := \sum_{j \le t} b_j \zeta_{t-j}, \quad b_s = s^{-(1-d)} L(s), \quad 1 \le t \le T, \ s \ge 1, \ \text{for some } 0 < d < 1/2,$$

where L is a slowly varying function at infinity and $L(s)$ is positive for all large s, and where ζ_t, $t \in \mathbb{Z} := \{0, \pm 1, \pm 2, \cdots, \}$ are i.i.d. standardized r.v.'s, independent of X_t, $t \in \mathbb{Z}$.

Throughout this paper the covariate process X_t, $t \in \mathbb{Z}$, will be assumed to be a stationary long memory moving average of the form

$$(1.3)X_t = \theta + \sum_{s \le t} b_{1,t-s} \xi_s, \quad b_{1,s} := s^{-(1-d_1)} L_1(s), \quad s \ge 1, \quad \text{for some } 0 < d_1 < 1/2,$$

where L_1 is another slowly varying function at infinity, $L_1(k)$ positive for all large k, and where ξ_s, $s \in \mathbb{Z}$, are i.i.d. standardized r.v.'s, and independent of $\{\zeta_s, \ s \in \mathbb{Z}\}$; θ is some constant. Note that $E\varepsilon_0 = 0$, $EX_0 = \theta$ and

$$\sum_{s=0}^{\infty} b_s^2 < \infty, \quad \sum_{s=0}^{\infty} b_{1,s}^2 < \infty,$$
$$Cov(\varepsilon_t, \, \varepsilon_{t+k}) = k^{-(1-2d)} \tilde{L}(k), \quad Cov(X_t, X_{t+k}) = k^{-(1-2d_1)} \tilde{L}_1(k), \quad \forall \, t \in \mathbb{Z}, \, k \ge 1,$$

where $\tilde{L}(k)$, $\tilde{L}_1(k)$ are slowly varying functions, positive for all large k, and such that

$$\begin{aligned} \lim_{k \to \infty} \frac{\tilde{L}(k)}{L^2(k)} &= \int_0^{\infty} (u + u^2)^{-(1-d)} du = Beta(1 - 2d, d), \\ \lim_{k \to \infty} \frac{\tilde{L}_1(k)}{L_1^2(k)} &= \int_0^{\infty} (u + u^2)^{-(1-d_1)} du = Beta(1 - 2d_1, d_1). \end{aligned}$$

We shall additionally assume the following:

$$(1.4) \qquad \left| E e^{iu\xi_0} \right| \le C \, (1 + |u|)^{-\delta}, \quad \text{for some} \quad C, \, \delta > 0, \ \forall \, u \in \mathbb{R},$$

$$(1.5) \qquad E \left| \xi_0 \right|^r < \infty, \quad \text{for some } r > 2.$$

Throughout this paper, G will denote the distribution function of X_0.

In the case when both the errors and the covariates are i.i.d., Stute, Thies and Zhou (1998) have shown that under the null hypothesis, the sequence of processes $T^{-1/2}\hat{\mathcal{V}}_T$ converges weakly to a continuous Gaussian process with an unknown and complicated covariance function, which makes this asymptotic null distribution infeasible. They then proposed a transformation of the process along the lines of Khmaladze (1981) that converges weakly to $\sigma_e B(G)$ under H_0, where B is a standard Brownian motion on $[0, 1]$. In this paper, we find this continues to be the case when the errors form a homogeneous martingale difference sequence (1.1) having finite fourth moment, and when the explanatory variable is a long memory process (1.3), having finite third moment and independent of the errors, as indicated in Section 2 below.

Now consider the case when $\{X_t\}$ is as in (1.3), and the error process follows the long memory moving average model (1.2) under H_0. In this case the nature of the limiting null distribution of $\widehat{\mathcal{V}}_T$ depends on the nature of the vector function h and on whether $d+d_1 > 1/2$ and $d+d_1 < 1/2$. In particular, from Lemma 4.2 below, under a mild moment condition on $h(X_0)$, it follows that $(L(T)T^{d+1/2})^{-1}\hat{\mathcal{V}}_T(x)$ converges weakly to $J(x)Z$, where Z is $\mathcal{N}(0, \sigma^2(d))$ r.v., with

$$\begin{aligned} \sigma^2(d) &:= \frac{Beta(1-2d, d)}{d(2d+1)}, \quad H := Eh(X_0)h(X_0)', \quad \nu(x) := Eh(X_0)I(X_0 \le x), \\ J(x) &= G(x) - Eh(X_0)'\, H^{-1}\, \nu(x), \qquad x \in \mathbb{R}. \end{aligned} \tag{1.6}$$

Consequently, when $\sup_x |J(x)| \neq 0$, e.g., as in the case of fitting a polynomial regression of order two or higher through the origin, the test that rejects H_0 whenever

$$\frac{\sup_x |\hat{\mathcal{V}}_T(x)|}{L(T)T^{\hat{d}+1/2}\sigma(\hat{d}) \sup_x |\hat{J}_T(x)|} \ge z_{\alpha/2}, \tag{1.7}$$

will be of the asymptotic size α. Here, z_α is $100(1-\alpha)\%$ percentile of the $\mathcal{N}(0, 1)$ distribution,

$$\begin{aligned} \hat{J}_T(x) &:= G_T(x) - T^{-1}\sum_{t=1}^{T} h(X_t)'\overline{H}_T^{-1}\overline{\nu}_T(x), \quad G_T(x) := T^{-1}\sum_{t=1}^{T} I(X_t \le x), \\ \nu_T(x) &:= \sum_{t=1}^{T} h(X_t)\, I(X_t \le x), \quad \overline{\nu}_T(x) := T^{-1}\nu_T(x), \quad x \in \mathbb{R}; \\ H_T &:= \sum_{t=1}^{T} h(X_t)h(X_t)', \quad \overline{H}_T := T^{-1}H_T. \end{aligned} \tag{1.8}$$

Also note that if $Eh(X_0) = 0$, then $J(x) \equiv G(x)$, and the test that rejects H_0 whenever

$$\frac{1}{L(T)T^{\hat{d}+1/2}\sigma(\hat{d})} \sup_x |\hat{\mathcal{V}}_T(x)| \ge z_{\alpha/2} \tag{1.9}$$

would also be of the asymptotic size α.

In the above, $\hat{d}$ is an estimator of d based on the residuals $\hat{\varepsilon}_t$ such that $|\log(T)(\hat{d} - d)| = o_p(1)$, e.g., the local Whittle estimator. Theorem 1.2 of Koul and Surgailis (2000) shows that in the case of linear regression models with the errors and covariates as in (1.2) and (1.3), this estimator satisfies the required condition. A similar analysis shows this continues to be the case under H_0.

There are, however, some interesting cases when $\sup_x |J(x)| = 0$; for example when the regression model being fitted has a non-zero intercept. From Lemma 4.3 and Corollary 4.1, one obtains the following fact when fitting a simple linear regression model where $q = 2$, $h(x)' = (1, x)$. If $d + d_1 > 1/2$, then $\{L(T)L_1(T)T^{d+d_1}\}^{-1}\hat{\mathcal{V}}_T(x)$ converges weakly to $V(x)(Z\,Z_1 - U)$, where U is as in Corollary 4.1 below,

$$V(x) := g(x) + \frac{1}{\sigma^2}\Big(K(x) - \theta G(x)\Big), \quad \sigma^2 = Var(X_0), \quad K(x) := EX_0 I(X_0 \le x),\ x \in \mathbb{R},$$

Z, Z_1 are independent mean zero Gaussian r.v.'s with respective variances $\sigma^2(d)$, $\sigma^2(d_1)$, and where g is density of G whose existence is guaranteed by (1.4), cf. Giraitis, Koul and Surgailis (1996). It is interesting to note that $V(x) \equiv 0$, if and only if G is $\mathcal{N}(\theta, \sigma^2)$ distribution. The distribution of the r.v. $ZZ_1 - U$ is not easy to characterize, mainly because of the complicated nature of the r.v. U. This in turn makes implementation of tests based on $\widehat{\mathcal{V}}_T$ relatively difficult in these cases. The limiting null distribution is even harder to characterize in the case $d + d_1 < 1/2$. See Lemma 4.4 for some special cases. In general, the Stute, Thies and Zhu (1998) transformation is not available when the errors follow the long memory model (1.2).

The testing procedures given by equations (1.7) and (2.6) below are illustrated with an application to the well known forward premium anomaly in international finance. The anomaly refers to the fact that the regression of spot exchange rate returns on the lagged forward premium invariably produces a negative slope coefficient, instead of unity as implied by the theory of uncovered interest rate parity. Several authors including Baillie and Bollerslev (1994, 2000) and Maynard and Phillips (2001) have documented the long memory characteristics of the forward premium. While spot returns generally appear to have little serial correlation, Cheung (1993) has also found empirical evidence for long memory in some spot returns series. The occurrence of long memory in both spot returns and in the forward premium then gives rise to the possibility of a balanced regression and some of the motivation and evidence for this is further discussed in section 3 below. On using similar data to Cheung (1993) we find evidence for long memory behavior in both spot returns and the forward premium, and in one case for the Canada versus the US, there is also evidence of long memory in the regression residuals. We show how the tests for mis-specification can be formed for two of the currencies, which are found to have different regression features.

We now briefly address some aspects of the asymptotic power of the above proposed tests. Let γ be a real valued function on $\mathbb{R}$ satisfying

(1.10) $0 < E\gamma^2(X_0) < \infty$,

(1.11) $\Delta(x) := E\gamma(X_0)I(X_0 \le x) - \nu'(x)H^{-1}Eh(X_0)\gamma(X_0) \neq 0, \quad$ for some $x \in \mathbb{R}$.

Then the following claims hold. All of the above tests are consistent against any fixed alternative $\mu(x) = \beta' h(x) + \gamma(x)$. The test (2.6) that is valid for the martingale difference errors has nontrivial asymptotic power against the sequence of alternatives $\mu(x) = \beta' h(x) + T^{-1/2}\gamma(x)$. The test (1.7), valid under (1.2) and (1.3) when $\sup_x |J(x)| \neq 0$, has nontrivial asymptotic power against the sequence of alternatives $\mu(x) = \beta' h(x) + L(T)T^{d-1/2}\gamma(x)$. In fact the asymptotic power of this test against these alternatives is

$$P\Big(J(x)Z > z_{\alpha/2} \sup_z |J(z)| + \Delta(x), \text{ for some } x \in \mathbb{R}\Big)$$
$$+P\Big(J(x)Z < -z_{\alpha/2} \sup_z |J(z)| + \Delta(x), \text{ for some } x \in \mathbb{R}\Big).$$

An example where the conditions (1.10)-(1.11) hold is when X_0 has six finite moments, $h(x)' = (x, x^2)$, $\gamma(x) = \alpha\, x^3$, $\alpha \in \mathbb{R}$. Another class of examples is given by all those γ's that satisfy (1.10), have $Eh(X_0)\gamma(X_0) = 0$, and $\sup_x |E\gamma(X_0)I(X_0 \le x)| \neq 0$. An example where (1.10) does not hold is when $\gamma(x) = a'h(x)$, $a \in \mathbb{R}^q$.

In the sequel, X, ε stand for copies of X_0, ε_0, respectively and all limits are taken as $T \longrightarrow \infty$, unless specified otherwise. By $\Longrightarrow$ we denote the weak convergence in $D[-\infty, \infty]$ with respect to the uniform metric, and by $\longrightarrow_D$ the finite dimensional weak convergence.

2 The transformed $\widehat{\mathcal{V}}_T$ process under martingale difference errors

This section first discusses the weak convergence of the process $\widehat{\mathcal{V}}_T$ under H_0 when the covariate process has the long memory structure (1.3), and when the errors are the martingale differences (1.1), independent of the covariate process. The distribution of this limiting process depends on the null model and hence is unknown. Then a transformation $\widehat{\mathcal{T}}_T$ of $\widehat{\mathcal{V}}_T$ is given such that $\widehat{\mathcal{T}}_T$ converges weakly to $\sigma_e B(G)$, where B is the standard Brownian motion on $[0, 1]$. Consequently tests based on suitably normalized process $\widehat{\mathcal{T}}_T$ will be asymptotically distribution free. A computation formula for the specific test based on the supremum norm is also given.

Recall the notation from (1.6), (1.8) and let

$$\mathcal{V}_T(x) \;:=\; \sum_{t=1}^{T} \varepsilon_t \, I(X_t \le x), \qquad \mathcal{Z}_T := \sum_{t=1}^{T} h(X_t)\varepsilon_t.$$

Assume that

$$H_T \text{ is almost surely positive definite for all } T \ge q. \tag{2.1}$$

Then, almost surely, the least squares estimator $\hat{\beta}$ satisfies $\hat{\beta} - \beta = H_T^{-1}\, \mathcal{Z}_T,$ and we obtain

$$\begin{aligned}
\hat{\mathcal{V}}_T(x) &= \sum_{t=1}^{T} (Y_t - \hat{\beta}' h(X_t))\, I(X_t \le x) \\
&= \mathcal{V}_T(x) - (\hat{\beta} - \beta)' \sum_{t=1}^{T} h(X_t)\, I(X_t \le x) = \mathcal{V}_T(x) - \mathcal{Z}_T'\, H_T^{-1}\, \nu_T(x) \\
&= \mathcal{V}_T(x) - \mathcal{Z}_T'\, \overline{H}_T^{-1}\, \overline{\nu}_T(x). \qquad (2.2)
\end{aligned}$$

Under (1.1), (1.3), and the assumed independence between $\{X_t\}$ and $\{\varepsilon_t\}$, the covariance terms in the variances of $\mathcal{V}_T(x)$ and $\mathcal{Z}_T$ will be zero so that the long memory aspect of $\{X_t\}$ is dominated by the martingale structure of the errors. Hence,

$$Cov\Big(\mathcal{V}_T(x), \mathcal{V}_T(y)\Big) \;=\; T\, \sigma_e^2\, G(x \wedge y), \qquad E\mathcal{Z}_T\mathcal{Z}_T' = T\, \sigma_e^2\, H.$$

Moreover, from Lemma 5.4, proved in the last section of this paper, we obtain that $T^{-1/2}\hat{\mathcal{V}}_T$ converges weakly to $W_G,$ where W_G is a Gaussian process on $\mathbb{R}$ with mean zero and covariance function $\sigma_e^2[G(x \wedge y) - \nu(x)'\, H^{-1}\, \nu(y)],$ $x,\, y \in \mathbb{R}.$ Due to the complicated nature of this covariance function and the fact G is unknown, the close form of the distribution of this limiting process is unknown. To overcome this we now consider a transformation of $\hat{\mathcal{V}}_T(x)$ whose limiting null distribution is like that of $\sigma_e\, B(G)$. Let

$$H(x) := Eh(X)h(X)'I(X \ge x), \quad \overline{H}_T(x) := T^{-1} \sum_{t=1}^{T} h(X_t)h(X_t)'I(X_t \ge x), \quad x \in \mathbb{R}.$$

Assume

$$H(x_0) \text{ is positive definite for some } x_0 < \infty. \tag{2.3}$$

Then, $H(x)$ is positive definite for all $x \le x_0$.

For a real valued function f, define the transform

$$\mathcal{T} f(x) := f(x) - \int_{y\le x} h(y)' H^{-1}(y) \Big\{ \int h(z)\, I(z \ge y) f(dz) \Big\} G(dy),$$
$$\widehat{\mathcal{T}}_T(x) := \hat{\mathcal{V}}_T(x) - \int_{y\le x} h(y)' \overline{H}_T^{-1}(y) \Big\{ \int h(z)\, I(z \ge y)\, \hat{\mathcal{V}}_T(dz) \Big\} G_T(dy), \quad x \le x_0.$$

Whenever $\mathcal{T}$ is applied to a Brownian motion, the inner integral is to be interpreted as a stochastic integral. Since $\mathcal{T}$ is a linear functional, $\mathcal{T}B(G)$ is a centered Gaussian process. In fact Stute, Thies and Zhu (1998) observed that $\mathcal{T}B(G) = B(G)$, in distribution. They also proved that in the case of the i.i.d errors and i.i.d. design, $T^{-1/2}\widehat{\mathcal{T}}_T$ converges weakly to $\sigma_e B(G)$, on $D[-\infty, x_0]$, so that the asymptotic null distribution of the test based on $\sup_{x\le x_0} |\widehat{\mathcal{T}}_T(x)|/[T s_T^2\, G_T(x_0)]^{1/2}$ is known, where $s_T^2 := T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2$. A similar result is established when fitting an autoregressive model of order 1 in Koul and Stute (1999).

This continues to be true when the errors form martingale differences as in (1.1) and the process X_t has long memory as in (1.3), independent of the errors. In view of Lemma 5.4 below, the proof of this claim is similar to that in Stute, Thies and Zhu (1998).

A computational formula for $\widehat{\mathcal{T}}_T$ is as follows. Let $X_{(j)}$, $1 \le j \le T$, denote the ordered X_t's in an ascending order and $\hat{\eta}_t$'s denote the corresponding $\hat{\varepsilon}_t$'s. Also, let, for $1 \le s,\, j \le T-1$,

$$H_{sT} \;:=\; H_T(X_{(s)}) := \sum_{t=s}^T h(X_{(t)}) h(X_{(t)})' \qquad \mathcal{J}_{jt} := \sum_{s=1}^j h(X_{(s)})' H_{sT}^{-1}\, h(X_{(t)}).$$

Then, for all $x < X_{(T)}$,

$$\text{(2.4)}\quad \widehat{\mathcal{T}}_T(x) \;=\; \sum_{t=1}^T \Big[I(X_t \le x) - T^{-1} \sum_{s=1}^T h(X_s)' \overline{H}_T^{-1}(X_s) h(X_t)\, I(X_s \le X_t \wedge x) \Big]\, \hat{\varepsilon}_t$$
$$=\; \sum_{t=1}^T \Big[I(X_{(t)} \le x) - \sum_{s=1}^T h(X_{(s)})' H_T^{-1}(X_{(s)}) h(X_{(t)})\, I(X_{(s)} \le X_{(t)} \wedge x) \Big]\, \hat{\eta}_t.$$

From (2.4), one sees that if for some $1 \le j \le T-1$, $X_{(j)} \le x < X_{(j+1)}$, then $\mathcal{T}_T(x) \equiv \mathcal{S}_j$, where

$$\mathcal{S}_j \;=\; \sum_{t=1}^j \Big[1 - \sum_{s=1}^t h(X_{(s)})' H_{sT}^{-1}\, h(X_{(t)}) \Big]\, \hat{\eta}_t - \sum_{t=j+1}^T \sum_{s=1}^j h(X_{(s)})' H_{sT}^{-1}\, h(X_{(t)}) \hat{\eta}_t$$
$$\text{(2.5)}\qquad =\; \sum_{t=1}^j \Big[1 - \mathcal{J}_{tt} \Big]\, \hat{\eta}_t - \sum_{t=j+1}^T \mathcal{J}_{jt} \hat{\eta}_t.$$

We shall now give a further simplification of this formula in the case of fitting a simple linear regression model: $q = 2$, $h_1(x) \equiv 1$, $h_2(x) \equiv x$. Let

$$\begin{aligned}
m_{1,s} &:= \sum_{t=s}^{T} X_{(t)}, \qquad \overline{m}_{1,s} := (T-s+1)^{-1} m_{1,s}, \\
m_{2,s} &:= \sum_{t=s}^{T} X_{(t)}^2, \qquad \overline{m}_{2,s} := (T-s+1)^{-1} m_{2,s}, \\
\tau_s^2 &:= m_{2,s} - \frac{m_{1,s}^2}{T-s+1}, \quad 1 \le s \le T; \quad s_T^2 := T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2.
\end{aligned}$$

Note that $\tau_T^2 = 0$. Thus, in calculating the above entities we must stay away from the last observation. We will compute our test statistics with $x_0 = X_{(T-1)}$. Then

$$\mathcal{J}_{jt} = \sum_{s=1}^{j} \frac{1}{\tau_s^2} \Big[\overline{m}_{2,s} - X_{(s)} \overline{m}_{1,s} + (X_{(s)} - \overline{m}_{1,s}) X_{(t)} \Big],$$

is well defined for $1 \le j \le T-1$. Thus the proposed test of the asymptotic level $0 < \alpha < 1$ is to reject the null hypothesis of a linear fit if the statistic

$$\mathcal{D}_T := T^{-1/2} \frac{\max_{1 \le j \le T-1} |\mathcal{S}_j|}{s_T [1 - (1/T)]^{1/2}} \tag{2.6}$$

is larger than b_α, the upper $100(1-\alpha)\%$ quantile of the $\sup\{|B(u)|;\, 0 \le u \le 1\}$.

3 Application to the Forward Premium Anomaly

This section considers some applications of the tests (1.7) and (2.6) to the well known forward premium anomaly in financial economics and international finance. The anomaly refers to the widespread empirical finding that the returns on most freely floating nominal exchange rates up until the early 1990's appear to be negatively correlated with the lagged forward premium or forward discount. This implies that a non-negative interest rate differential would on average, result in an appreciating currency for the country with the higher rate of interest, which violates the important theory of uncovered interest parity. This rejection has been found to create considerable problems for subsequent modeling and understanding in international finance. A good survey of the literature on the forward premium anomaly is provided by Engel (1996). The papers by Baillie and Bollerslev (1994, 2000) and Maynard and Phillips (2001) have emphasized the role of long memory processes in this context and how inappropriate inference that ignores the nature of the long memory aspects may have been partly responsible for the anomaly.

In the following, S_t denotes the spot exchange rate at time t, while F_t refers to the forward exchange rate at time t, for delivery at time $t+1$. Corresponding logarithmic values are denoted by the lower case variables, s_t, f_t, respectively; and all the rates are denominated with the U.S. dollar as the numeraire currency. The uncovered interest rate parity theory assumes rational expectations, risk neutrality, free capital mobility and the absence of taxes on capital transfers. A common test of the theory is to estimate the regression

$$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + u_{t+1}, \tag{3.1}$$

and to test the hypothesis that $\alpha = 0$, $\beta = 1$ and u_{t+1} be serially uncorrelated. However, the uncovered interest rate parity condition can be derived from a discrete time, consumption based asset pricing model, which includes a time dependent risk premium. The risk adjusted real returns over current and future consumption streams of the representative investor are given by $E_t[(F_t - S_{t+1})/P_{t+1}] \times U'(C_{t+1})/U'(C_t) = 0$, where $U'(C_{t+1})/U'(C_t)$ equals the marginal rate of substitution in terms of utility derived from current and future consumption and P_t represents the price level. A Taylor series approximation to second order terms implies the alternative regression function

$$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) - \frac{1}{2}Var_t(s_{t+1}) + Cov_t(s_{t+1}, p_{t+1}) + Cov_t(s_{t+1}, q_{t+1}) + u_{t+1}, \tag{3.2}$$

where q_{t+1} denotes the logarithm of the intertemporal marginal rate of substitution. Under the null hypothesis of uncovered interest parity, $\alpha = 0$, $\beta = 1$ and u_{t+1} is serially uncorrelated. Usually, the Jensen inequality terms $Var_t(s_{t+1})$ and $Cov_t(s_{t+1}, p_{t+1})$ are considered to be statistically negligible and are ignored. Importantly, the covariance term $Cov_t(s_{t+1}, q_{t+1})$ has the interpretation of being a time dependent risk premium and may have long memory.

To illustrate the above testing procedures, we now consider instances where both the forward premium and spot returns appear to have long memory characteristics. Baillie and Bollerslev (2000) and Maynard and Phillips (2001) discuss the case of an unbalanced regression when spot returns are close to being uncorrelated, while the forward premium is a long memory process. However, a number of authors including Booth, Kaen and Koveos (1982), Cheung (1993) and Lo (1991) have argued that asset and exchange rate returns in general, may exhibit long memory behavior, in which case the regression may not be unbalanced. In particular, Cheung (1993) provides convincing evidence for the presence of long memory in spot returns for certain currencies in the initial part of the post Bretton Woods era of floating exchange rates. He argues that slow adjustments to purchasing power parity can cause the long memory property of spot returns.

A more general motivation is to use a model where both the spot and forward rate have a common stochastic trend term $z_t = z_{t-1} + v_t$, where v_t is a long memory process. Then $s_t = z_t + u_{1t}$ and $f_t = z_t + u_{2t}$, where both u_{1t} and u_{2t} are short memory processes. Hence this type of model defines the spot exchange rate as being determined by fundamentals, or long run equilibrium value, z_t, that follow an integrated process with a long memory noise component. This is an empirically very reasonable model since it is well known that price levels and money supplies follow this type of process; e.g. see Baillie, Chung and Tieslau (1996). The model is also similar to the theoretical formulation of Mussa (1982), with the v_t process partly determining the state of disequilibrium in the goods market and speed of adjustment coefficient. Hai, Mark and Wu (1997) further consider this model in an unobserved components framework. The spot returns are then the sum of a short memory, possibly white noise component and a long memory component. It should be noted that if the variance of u_{1t} component dominates that of the v_t component, the spot returns may appear approximately uncorrelated in small sample sizes. This would explain the mixed findings on the order of integration of the spot returns that have been reported in the literature. However, there is very widespread empirical evidence that forward premiums are invariably long memory processes and Baillie and Bollerslev (1994, 2000) and Maynard and Phillips (2001) find them to be well described by ARFIMA(p, d, q) processes, so that $(1-L)^d(f_t - s_t)$ is a stationary and invertible $I(0)$ process. The presence of long memory components in both the spot returns and forward premium will then suggest the possibility of a balanced regression.

The regression in equation (3.1) was fitted to a very similar sample period and currencies to that of Cheung (1993) where some spot returns appeared to exhibit long memory. The first example uses monthly observations on the Canadian \$ - US \$ spot and the one-month forward rate from January 1974 through December 1991, which realizes a total of $T = 215$ observations. As expected there is strong evidence of long memory in the forward premium with the Ljung-Box statistic on the first 20 lags being $Q(20) = 803.52$; and the first twelve autocorrelation coefficients of the forward premium series are $0.89, 0.74, 0.63, 0.55, 0.52, 0.50, 0.49, 0.47, 0.43$, 0.36, and 0.28, respectively. The application of the local Whittle estimator resulted in an estimated value of .25 for the long memory parameter in the forward premium.[3]

The monthly Canadian \$ - US \$ spot returns over the same period also have evidence of long memory with the d estimate in an ARFIMA$(1, d, 0)$ model being 0.12 and the robust t

[3] An alternative approach of estimating a parametric ARFIMA model with GARCH errors was also pursued and the QMLE of the long memory parameter was found to be 0.20 and was statistically significant at the .05 level.

statistic being significant at the .05 level. We test the hypothesis H_0 with $h(x) = (1, x)'$, with x representing the lagged forward premium. The OLS estimates of β were $(-2.606, -1.432)'$ with standard errors of $(0.811, 0.407)'$. Hence the regression gives rise to the usual negative slope coefficient that is consistent with the anomaly. Ljung-Box tests on the autocorrelations of the standardized residuals and also their squares failed to reject the hypothesis that the residuals were serially uncorrelated and without ARCH effects. Hence it seemed reasonable to assume that the errors form a martingale difference sequence. This justifies the application of the test statistic $\mathcal{D}_T$ of (2.6) to fit the model (3.1) to this data. For the Canadian forward premium regression, $\mathcal{D}_T = 5.83$, which indicates a clear rejection of the null hypothesis.

A further example of the methodology is provided by the monthly British pound - US dollar over the same time period. The MLE of the long memory parameter from an ARFIMA model on spot returns gives a value of 0.293 with standard error of 0.080, which indicates long memory in spot returns. Consistent with previous studies of Baillie and Bollerslev (1994) and Maynard and Phillips (2001) and the evidence for Canada above; the British forward premium also exhibits evidence of long memory, the MLE of the long memory parameter being 0.45 with a standard error of 0.124. Estimation of the traditional forward premium anomaly regression (3.1) confirmed evidence of the standard anomalous negative slope coefficient. Moreover, there is also evidence of substantial, persistent autocorrelation in the residuals, which can also be represented as a long memory process. An interpretation of this finding can be seen from equation (3.2) where the last covariance term $Cov_t(s_{t+1}, q_{t+1})$ is associated with a time dependent risk premium. Hence variables associated with the risk premium may give rise to the long memory process in the residuals. While there are many possible models for a risk premium, for the sake of illustration of the methodology in this paper, it is worth noting that several authors such as Giovannini and Jorion (1987) and Hodrick (1989) have used the lagged squared forward premium to represent the risk premium.

This motivates fitting the second degree polynomial to the British - pound - US dollar data with long memory errors, i.e., testing for H_0 with $h(x) = (x, x^2)'$, where again x represents the lagged forward premium. The OLS estimates of β were $(-2.414, -0.244)'$ with standard errors of $(0.835, 0.129)'$. To test for further mis-specification of this regression, the statistic in (1.7), with $L(T) \equiv 1$, was calculated to be 6.72, based on the MLE of the long memory parameter d in the error process equal to 0.262. Thus, again regression also appears to be misspecified. A possible implication of this result is that further economic variables terms associated with a risk premium should be included. The value of the statistic in (1.7) appeared to be relatively stable to alternative semi-parametric estimate of d obtained from local Whittle estimation, and still led to a clear rejection of the null hypothesis.

4 Some general results about $\widehat{\mathcal{V}}_T$ - process under long memory moving average errors

Now, we shall discuss the asymptotic null distribution of $\widehat{\mathcal{V}}_T$ when (1.2) and (1.3) hold. Towards this goal, we first need to determine the magnitude of $\hat{\mathcal{V}}_T(x)$. It turns out this depends on the nature of the vector function h and on whether $d+d_1 > 1/2$ or $d+d_1 < 1/2$. To make things a bit more precise, we first state some general rate results, then use them to determine the magnitude of $\hat{\mathcal{V}}_T(x)$. We need to introduce some more notation:

$$\mathcal{W}_T := \sum_{t=1}^{T} (h(X_t) - Eh(X_t))\,\varepsilon_t, \quad S_T := \sum_{t=1}^{T} \varepsilon_t, \quad \mathcal{U}_T := \sum_{t=1}^{T} (X_t - \theta)\varepsilon_t, \quad \mu := Eh(X).$$

For a given positive sequence $a_T \longrightarrow \infty$, by $U_p(a_T^{-1})$ $(u_p(a_T^{-1}))$, we mean a sequence of stochastic processes $\mathcal{Y}_T(x)$, $x \in \mathbb{R}$ such that $a_T \sup_{x\in\mathbb{R}} |\mathcal{Y}_T(x)|$ is bounded (converges to zero), in probability.

With the above notation, (2.2) can be rewritten as

$$\hat{\mathcal{V}}_T(x) \;=\; \mathcal{V}_T(x) - (\mu S_T + \mathcal{W}_T)' \overline{H}_T^{-1} \overline{\nu}_T(x). \tag{4.1}$$

Recall from Koul and Surgailis (1997) that under (1.4) and (1.5), the distribution function G and the density g are infinitely differentiable. The last fact together with (4.2) below implies the infinite differentiability of the functions

$$\begin{aligned} \mu(u) \;&:=\; Eh(X+u) = \int_{\mathbb{R}} h(x)g(x-u)dx, \qquad \mu := \mu(0), \\ \nu(u;x) \;&:=\; Eh(X+u)I(X+u \le x) = \int_{-\infty}^{x} h(y)g(y-u)dy, \quad x \in \mathbb{R},\, u \in \mathbb{R}. \end{aligned}$$

Note that

$$\dot{\nu}(x) := \frac{\partial \nu(u;x)}{\partial u}\Big|_{u=0} = -\int_{-\infty}^{x} h(y)\dot{g}(y)dy, \qquad \dot{\mu} := \frac{\partial \mu(u)}{\partial u}\Big|_{u=0} = -\int_{\mathbb{R}} h(x)\dot{g}(x)dx.$$

We need the following lemma whose proof can be deduced from the results of Koul and Surgailis (2002) (see also Ho and Hsing (1996, 1997), Koul and Surgailis (1997), and Giraitis and Surgailis (1999)).

Lemma 4.1 *Assume conditions (1.2), (1.3), (1.4), and (1.5). Let $h(x)$, $x \in \mathbb{R}$, be any function with values in $\mathbb{R}^q$ such that*

$$\|h(x)\| \le C(1+|x|)^{\lambda}, \quad \text{for some} \quad 0 \le \lambda < (r-2)/2. \tag{4.2}$$

Then, there exists a $\kappa > 0$ such that

$$\begin{aligned} \mathcal{W}_T &= \dot{\mu}\,\mathcal{U}_T + O_p(T^{d+d_1-\kappa}), && \text{if } d+d_1 > 1/2, \\ &= O_p(T^{1/2}), && \text{if } d+d_1 < 1/2. \end{aligned} \tag{4.3}$$

$$\overline{\nu}_T(x) = \nu(x) + \dot{\nu}(x)(\overline{X}_T - \theta) + U_p(T^{d_1-1/2-\kappa}). \tag{4.4}$$

$$\begin{aligned} \mathcal{V}_T(x) &= G(x)S_T - g(x)\mathcal{U}_T + U_p(T^{d+d_1-\kappa}), && \text{if } d+d_1 > 1/2, \\ &= G(x)S_T + U_p(T^{1/2}), && \text{if } d+d_1 < 1/2. \end{aligned} \tag{4.5}$$

Moreover,

$$S_T = O_p(|L(T)|T^{d+1/2}), \qquad \overline{X}_T - \theta = O_p(|L_1(T)|T^{d_1-1/2}). \tag{4.6}$$

$$\begin{aligned} \mathcal{U}_T &= O_p(|L(T)L_1(T)|T^{d+d_1}), && \text{if } d+d_1 > 1/2, \\ &= O_p(T^{1/2}), && \text{if } d+d_1 < 1/2. \end{aligned} \tag{4.7}$$

It is also well-known that under conditions (1.2) and (1.3) alone, one has the following facts:

$$L(T)^{-1}T^{-d-1/2}S_T \longrightarrow_D Z, \qquad L_1(T)^{-1}T^{-d_1+1/2}(\overline{X}_T - \theta) \longrightarrow_D Z_1, \tag{4.8}$$

where Z, Z_1 are stochastic integrals w.r.t. to independent standard Gaussian white noises W, W_1, respectively:

$$\begin{aligned} Z &:= \int_{-\infty}^{1}\Big\{\int_0^1 (\tau - x)_+^{-(1-d)}d\tau\Big\}W(dx), \\ Z_1 &:= \int_{-\infty}^{1}\Big\{\int_0^1 (\tau - x)_+^{-(1-d_1)}d\tau\Big\}W_1(dx), \end{aligned} \tag{4.9}$$

and therefore have independent normal distributions with $EZ = EZ_1 = 0$ and the respective variances $\sigma^2(d)$, $\sigma^2(d_1)$, with $\sigma^2(d)$ as in (1.6).

Lemma 4.2 *Assume the conditions of Lemma 4.1. Then,*

$$L(T)^{-1}T^{-d-1/2}\hat{\mathcal{V}}_T(x) \Longrightarrow J(x)\,Z, \qquad J(x) := G(x) - \mu' H^{-1}\nu(x),\ x \in \mathbb{R}. \tag{4.10}$$

Proof. Assume first $d+d_1 > 1/2$. Use (4.1) and Lemma 4.1 (4.3), (4.4), (4.5), (4.7) together with well-known properties of slowly varying functions to obtain

$$\hat{\mathcal{V}}_T(x) = S_T(G(x) - \mu'\overline{H}_T^{-1}\nu(x)) + U_p(T^{d+d_1+\gamma}), \tag{4.11}$$

for any $\gamma > 0$. Clearly, by the Ergodic Theorem (ET), $\overline{H}_T \longrightarrow H$, a.s., and in view of (2.1), we thus have

$$\overline{H}_T^{-1} \to H^{-1}, \quad \text{a.s.} \tag{4.12}$$

Now (4.10) clearly follows from (4.11), (4.8), (4.12), as $d + 1/2 > d + d_1 + \gamma$ provided $\gamma > 0$ is sufficiently small. A similar argument with small changes applies also in the case $d + d_1 = 1/2$.

Next, let $d + d_1 < 1/2$. In this case we have the exact representation

$$\begin{aligned} \hat{\mathcal{V}}_T(x) \;&=\; S_T(G(x) - \mu' H^{-1}\nu(x)) \\ &\quad + S_T\Big[\mu' H^{-1}\nu(x) - \mu'\overline{H}_T^{-1}\overline{\nu}_T(x)\Big] \\ &\quad + \sum_{t=1}^{T} \varepsilon_t \Big[I(X_t \le x) - G(x) - (h(X_t) - \mu)' H^{-1}\nu(x)\Big] \\ &\quad - \Big(\sum_{t=1}^{T} \varepsilon_t (h(X_t) - \mu)\Big)' \Big[\overline{H}_T^{-1}\overline{\nu}_T(x) - H^{-1}\nu(x)\Big]. \end{aligned} \tag{4.13}$$

The Glivenko-Cantelli type of argument along with the ET implies

$$\mu' H^{-1}\nu(x) - \mu'\overline{H}_T^{-1}\overline{\nu}_T(x) = u_p(1). \tag{4.14}$$

By (4.14), the second term on the r.h.s. of (4.13) is $u_p(|S_T|)$, and the fourth term is $u_p(|\mathcal{W}_T|) = u_p(T^{1/2}) = u_p(|S_T|)$, see (4.4), (4.6). Finally, for the third term one has the uniform bound

$$\begin{aligned} &\sum_{t=1}^{T} \varepsilon_t (I(X_t \le x) - G(x) - (h(X_t) - \mu)' H^{-1}\nu(x)) \\ &= \nu_T(x) - G(x) S_T - \mathcal{W}_T' H^{-1}\nu(x) = U_p(|\mathcal{U}_T|) + U_p(T^{1/2}) = U_p(T^{1/2}) = u_p(|S_T|), \end{aligned}$$

according to (4.5), (4.3). □

As noted in the Introduction, Lemma 4.2 enables to implement the tests based on $\hat{\mathcal{V}}_T$ as soon as we can estimate J and as long as $\sup_x |J(x)|$ is positive, e.g. the asymptotically distribution free test (1.7). However, in some cases J is identically zero, as we shall now show.

Case $J(x) \equiv 0$. Note $J(x)$ can be written as

$$J(x) = \int_{-\infty}^{x} (1 - \mu' H^{-1} h(y)) dG(y).$$

Clearly, $J(x) \equiv 0$ implies $\mu' H^{-1} h(y) = 1, G -$ a.e. Assume also $H^{-1}\mu \neq 0$. These two relations imply that the functions $\{1, h_1(x), \ldots, h_q(x)\}$ are *linearly dependent G-a.e.*. In other words, there exist some constants $k_0, k_1, \ldots, k_q$ not all of them identically zero and such that

$$k_0 + \sum_{i=1}^{q} k_i h_i(x) = 0 \qquad G - \text{a.e.} \tag{4.15}$$

If we assume $h_1, \ldots, h_q$ linearly independent, (4.15) implies that one of the functions $h_1, \ldots,$ h_q is a constant $G-$a.e. In other words, we may assume in the sequel $h_1(x) \equiv 1$. Then, according to Lemma 4.1, we obtain

$$\begin{aligned}
\hat{\mathcal{V}}_T(x) &= G(x) S_T - g(x)\mathcal{U}_T + U_p(T^{d+d_1-\kappa}) \\
&\quad - S_T \mu' \Big[H^{-1} + (\overline{H}_T^{-1} - H^{-1})\Big] \Big[\nu(x) + \dot{\nu}(x)(\overline{X}_T - \theta) + U_p(T^{d_1 - 1/2 - \kappa})\Big] \\
&\quad - \Big[\dot{\mu}\mathcal{U}_T + O_p(T^{d+d_1-\kappa})\Big]' \Big[H^{-1} + o_p(1)\Big] \Big[\nu(x) + u_p(1)\Big].
\end{aligned}$$

After cancellation of the main term $(G(x) - \mu' H^{-1}\nu(x))S_T = J(x)S_T$, this yields

$$\begin{aligned}
\hat{\mathcal{V}}_T(x) &= -g(x)\mathcal{U}_T - S_T\mu'(\overline{H}_T^{-1} - H^{-1})\nu(x) - S_T(\overline{X}_T - \theta)\mu' H^{-1}\dot{\nu}(x) \\
&\quad - \mathcal{U}_T \dot{\mu}' H^{-1}\nu(x) + O_p(T^{d+d_1-\kappa}) + o_p(|\mathcal{U}_T|) \\
&= S_T \mu' \left[H^{-1} - \overline{H}_T^{-1}\right] \nu(x) - S_T(\overline{X}_T - \theta)\mu' H^{-1}\dot{\nu}(x) \\
&\quad - \mathcal{U}_T \left[g(x) + \dot{\mu}' H^{-1}\nu(x)\right] + O_p(T^{d+d_1-\kappa}) + o_p(|\mathcal{U}_T|).
\end{aligned} \tag{4.16}$$

The next lemma discusses the particular case $q = 2$. It is in particular useful in arriving at the null distribution of tests based on $\hat{\mathcal{V}}_T$ when fitting a simple linear regression model with a non-zero intercept, provided X_t is a non-Gaussian process. Let ℓ be a real valued function, and let

$$\begin{aligned}
K(x) &:= E\ell(X)I(X \le x), \qquad \mu_1 := E\ell(X), \qquad \sigma_\ell^2 := Var(\ell(X)), \\
\dot{\mu}_1 &:= \partial E\ell(X+u)/\partial u\big|_{u=0}, \qquad V(x) := g(x) + \frac{\dot{\mu}_1}{\sigma_\ell^2}\Big(K(x) - \mu_1 G(x)\Big), \quad x \in \mathbb{R}.
\end{aligned}$$

Lemma 4.3 *Assume* $d + d_1 > 1/2$. *Let* ℓ *be an arbitrary function such that*

$$\sigma_\ell^2 > 0, \qquad |\ell(x)| \le C(1+|x|)^\lambda, \quad \text{for some } \lambda < (r-2)/4. \tag{4.17}$$

Assume $q = 2$, *and let*

$$h(x)' = (1,\, \ell(x)), \qquad x \in \mathbb{R}. \tag{4.18}$$

Then, there exists a $\kappa > 0$ *such that*

$$\hat{\mathcal{V}}_T(x) = (S_T(\overline{X}_T - \theta) - \mathcal{U}_T)V(x) + U_p(T^{d+d_1-\kappa}). \tag{4.19}$$

Remark 4.1 In the case $\ell(x) = x$, the statement of Lemma 4.3 is valid under the condition (4.2) instead of (4.17), i.e. provided the moment condition (1.5) holds for some $r > 4$.

Proof of Lemma 4.3. Return to relation (4.16). Let us first identify the expressions $\mu' H^{-1}\dot{\nu}(x)$ and $\dot{\mu}' H^{-1}\nu(x)$ for the case (4.18). Here,

$$\mu' = (1, \mu_1), \quad \mu_2 := E\ell^2(X), \quad H = \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}, \quad H^{-1} = \sigma_\ell^{-2}\begin{pmatrix} \mu_2 & -\mu_1 \\ -\mu_1 & 1 \end{pmatrix}.$$

Let $\dot{K}(x) \equiv -\int_{-\infty}^x \ell(y)\dot{g}(y)dy$. Then, $\dot{\nu}(x)' = (-g(x), \dot{K}(x))$, implying

$$\mu' H^{-1}\dot{\nu}(x) = -\sigma_\ell^{-2}\Big(\mu_2 g(x) - \mu_1^2 g(x) + \mu_1\dot{K}(x) - \mu_1\dot{K}(x)\Big) = -g(x). \tag{4.20}$$

Also, in the case of (4.18),

$$\dot{\mu}' = \Big(0, -\int_{\mathbb{R}} \ell(x)\dot{g}(x)dx\Big) = (0, \dot{\mu}_1), \qquad \nu(x)' = (G(x), K(x)), \tag{4.21}$$

implying

$$\dot{\mu}' H^{-1}\nu(x) = \frac{1}{\sigma_\ell^2}\Big(-\dot{\mu}_1\mu_1 G(x) + \dot{\mu}_1 K(x)\Big) = \frac{\dot{\mu}_1}{\sigma_\ell^2}[K(x) - \mu_1 G(x)]. \tag{4.22}$$

To complete the proof of the lemma, we need to examine the behavior of $\overline{H}_T^{-1} - H^{-1}$. We have

$$\begin{aligned} \overline{H}_T^{-1} &= \frac{1}{s_T^2}\begin{pmatrix} \overline{\ell^2} & -\overline{\ell} \\ -\overline{\ell} & 1 \end{pmatrix}, \qquad \overline{\ell} := T^{-1}\sum_{t=1}^T \ell(X_t), \quad \overline{\ell^2} := T^{-1}\sum_{t=1}^T \ell^2(X_t), \\ s_T^2 &= \det(\overline{H}_T) = \overline{\ell^2} - (\overline{\ell})^2. \end{aligned}$$

Then

$$\overline{H}_T^{-1} - H^{-1} = Q_{1T} + Q_{2T},$$

where

$$Q_{1T} = \frac{1}{s_T^2} \begin{pmatrix} \overline{\ell^2} - \mu_2 & -\overline{\ell} + \mu_1 \\ -\overline{\ell} + \mu_1 & 0 \end{pmatrix}, \qquad Q_{2T} = \frac{\sigma_\ell^2 - s_T^2}{s_T^2} H^{-1}.$$

The limiting behavior of $\overline{\ell^2} - \mu_2$ will depend on the first Appell coefficient

$$\dot{\mu}_2 := \frac{\partial E\ell^2(X+u)}{\partial u}\Big|_{u=0} = -\int_{\mathbb{R}} \ell^2(x)\dot{g}(x)dx.$$

According to Lemma 4.1, under condition (4.17), we have

$$\overline{\ell^2} - \mu_2 = T^{-1}\sum_{t-1}^{T}(\ell^2(X_t) - E\ell^2(X_t)) = \dot{\mu}_2(\overline{X}_T - \theta) + O_p(T^{d_1-1/2-\kappa}), \qquad (\exists \kappa > 0).$$

Similarly, under the same condition,

$$\overline{\ell} - \mu_1 = T^{-1}\sum_{t=1}^{T}(\ell(X_t) - E\ell(X_t)) = \dot{\mu}_1(\overline{X}_T - \theta) + O_p(T^{d_1-1/2-\kappa}), \qquad (\exists \kappa > 0).$$

Hence we obtain

$$Q_{1T} = \frac{\overline{X_T - \theta}}{s_T^2} \begin{pmatrix} \dot{\mu}_2 & -\dot{\mu}_1 \\ -\dot{\mu}_1 & 0 \end{pmatrix} + O_p(T^{d_1-1/2-\kappa}), \quad (\exists \kappa > 0).$$

We also have

$$\begin{aligned} s_T^2 - \sigma_\ell^2 &= T^{-1}\sum_{t=1}^{T}\ell^2(X_t) - \Big(T^{-1}\sum_{t=1}^{T}\ell(X_t)\Big)^2 - \sigma_\ell^2 \\ &= T^{-1}\sum_{t=1}^{T}(\ell^2(X_t) - E\ell^2(X_t)) - (\mu_1 + (\overline{\ell} - \mu_1))^2 + \mu_1^2 \\ &= \dot{\mu}_2(\overline{X}_T - \theta) - 2\mu_1\dot{\mu}_1(\overline{X}_T - \theta) + O_p(T^{d_1-1/2-\kappa}) \qquad (\exists \kappa > 0). \end{aligned}$$

Clearly, this implies

$$\frac{1}{s_T^2} - \frac{1}{\sigma_\ell^2} = \frac{\sigma^2 - s_T^2}{\sigma_\ell^2 s_T^2} = \frac{2\mu_1\dot{\mu}_1 - \dot{\mu}_2}{\sigma_\ell^4}(\overline{X}_T - \theta) + O_p(T^{d_1-1/2-\kappa}) \qquad (\exists \kappa > 0).$$

Hence we obtain, for some $\kappa > 0$,

$$\begin{aligned} &\overline{H}_T^{-1} - H^{-1} \\ &= \frac{(\overline{X}_T - \theta)}{\sigma_\ell^2}\left\{\begin{pmatrix} \dot{\mu}_2 & -\dot{\mu}_1 \\ -\dot{\mu}_1 & 0 \end{pmatrix} + \frac{2\mu_1\dot{\mu}_1 - \dot{\mu}_2}{\sigma_\ell^2}\begin{pmatrix} \mu_2 & -\mu_1 \\ -\mu_1 & 1 \end{pmatrix}\right\} + O_p(T^{d_1-1/2-\kappa}). \end{aligned}$$

Using the last result together with (4.21) and $\mu' = (1, \mu_1)$, after some algebra and numerous cancellations, we obtain, for some $\kappa > 0$,

$$
\begin{aligned}
(4.23) \qquad & S_T \mu'(H^{-1} - \overline{H}_T^{-1})\nu(x) \\
&= \frac{S_T(\overline{X}_T - \theta)}{\sigma_\ell^2} \Big\{ \dot{\mu}_1 [K(x) - \mu_1 G(x)] \Big\} + U_p(T^{d+d_1-\kappa}).
\end{aligned}
$$

Clearly, the lemma follows from (4.16), (4.20), (4.22), and (4.23). □

The following corollary describes the limiting distribution of the empirical process $\hat{\mathcal{V}}_T$ under the conditions of Lemma 4.3.

Corollary 4.1 *Under the conditions of Lemma 4.3,*

$$(L(T)L_1(T)T^{d+d_1})^{-1}\hat{\mathcal{V}}_T(x) \Longrightarrow V(x)\,(Z\,Z_1 - U),$$

where Z, Z_1 are the same as in (4.9) and where U is the stochastic integral (defined on the same probability space as Z, Z_1):

$$U \;:=\; \int_{-\infty}^{1} \int_{-\infty}^{1} \Big\{ \int_0^1 (\tau - x)_+^{-(1-d)} (\tau - x_1)_+^{-(1-d_1)} d\tau \Big\} W(dx) W_1(dx_1),$$

with $W(dx)$, $W_1(dx)$ being mutually independent Gaussian white noises with zero mean and variance $E(W(dx))^2 = E(W_1(dx))^2 = dx$.

Remark 4.2 Recall that Z and Z_1 are independent Gaussian r.v.'s, while the double Ito-Wiener integral U is non-Gaussian and is well-defined iff $d + d_1 > 1/2$. See Avram and Taqqu (1987) for the weak convergence to the multiple Ito-Wiener integrals of the type U and Dobrushin and Major (1979) for their characteristic functions.

Lemma 4.4 *Assume h is of the form (4.18) and satisfies the conditions of Lemma 4.3, and let $d + d_1 < 1/2$. Then*

$$(4.24) \qquad T^{-1/2}\hat{\mathcal{V}}_T(x) \longrightarrow_D \mathcal{G}(x),$$

where $\mathcal{G}(x), x \in \mathbb{R}$, is a Gaussian process with zero mean and covariance

$$
\begin{aligned}
(4.25) \quad & Cov(\mathcal{G}(x), \mathcal{G}(y)) \\
&= \sum_{t=-\infty}^{\infty} Cov(\varepsilon_0, \varepsilon_t) Cov\Big(I(X_0 \le x) - G(x) - (h(X_0) - \mu)' H^{-1} \nu(x), \\
&\qquad\qquad I(X_t \le y) - G(y) - (h(X_t) - \mu)' H^{-1} \nu(y) \Big) \\
&= \sum_{t=-\infty}^{\infty} Cov(\varepsilon_0, \varepsilon_t) Cov\Big(I(X_0 \le x) - G(x) - \sigma_\ell^{-2}(\ell(X_0) - \mu_1)(K(x) - \mu_1 G(x)), \\
&\qquad\qquad I(X_t \le y) - G(y) - \sigma_\ell^{-2}(\ell(X_t) - \mu_1)(K(y) - \mu_1 G(y)) \Big),
\end{aligned}
$$

where ℓ, μ_1, σ^2, $K(x)$ are the same as in Lemma 4.3.

Sketch of a Proof. The lemma follows from (4.13), provided one shows

$$
\begin{aligned}
&(4.26) \quad & T^{-1/2}\sum_{t=1}^{T}\varepsilon_t(I(X_t \le x) - G(x) - (h(X_t)-\mu)'H^{-1}\nu(x)) &\implies \mathcal{G}(x),\\
&(4.27) \quad & S_T\mu'(H^{-1}\nu(x) - \overline{H}_T^{-1}\overline{\nu}_T(x)) &= u_p(T^{1/2}),\\
&(4.28) \quad & \Big(\sum_{t=1}^{T}\varepsilon_t(h(X_t)-\mu)\Big)'(\overline{H}_T^{-1}\overline{\nu}_T(x) - H^{-1}\nu(x)) &= u_p(T^{1/2}).
\end{aligned}
$$

From Lemma 4.1 (4.4) and the proof of Lemma 4.3, we have

$$
\begin{aligned}
(4.29) \qquad & \overline{H}_T^{-1}\overline{\nu}_T(x) - H^{-1}\nu(x)\\
&= \Big[H^{-1} + (\overline{H}_T^{-1} - H^{-1})\Big)\Big(\nu(x) + (\overline{\nu}_T(x) - \nu(x)\Big) - H^{-1}\nu(x)\\
&= O_p\Big(\|\overline{H}_T^{-1} - H^{-1}\|\Big) + O_p(\sup_x|\overline{\nu}_T(x) - \nu(x)|)\\
&= U_p(|\overline{X}_T - \theta|).
\end{aligned}
$$

Therefore, by (4.6), and the property of slowly varying functions,

$$
S_T\mu'(H^{-1}\nu(x) - \overline{H}_T^{-1}\overline{\nu}_T(x)) \;=\; U_p(|S_T(\overline{X}_T-\theta)|) = u_p(T^{d+d_1+\kappa}), \qquad (\forall\kappa > 0).
$$

As $d + d_1 < 1/2$, this proves (4.27). In a similar way, from (4.3), (4.6) and (4.29), we obtain

$$
\begin{aligned}
&\Big(\sum_{t=1}^{T}\varepsilon_t(h(X_t)-\mu)\Big)'\Big[\overline{H}_T^{-1}\overline{\nu}_T(x) - H^{-1}\nu(x)\Big]\\
&= U_p(|\mathcal{W}_T(\overline{X}_T - \theta)|) = U_p(|L_1(T)|T^{1/2+d_1-1/2})\\
&= U_p(T^{d_1+\kappa}), \qquad\qquad (\forall\kappa > 0).
\end{aligned}
$$

This proves (4.28), as $d_1 < 1/2$.

The proof of the finite dimensional distributional convergence of the sum in (4.26) can be carried out using the ideas of Koul and Surgailis (1997) or Ho and Hsing (1997). The proof of the tightness is carried out along the lines of Dehling and Taqqu (1989) and Koul and Surgailis (1997). It is important to note that under conditions of Lemma 4.4, the sum of the covariances in (4.25) is absolutely convergent, for any x, $y \in \mathbb{R}$, due to $Cov(\varepsilon_0, \varepsilon_t) = O(L^2(t)t^{-(1-2d)})$ and $Cov(I(X_0 \le x) + \tilde{c}\ell(X_0), I(X_t \le y)) + c\ell(X_t)) = O(L_1^2(t)t^{-(1-2d_1)})$, for any constants $\tilde{c}, c$. □

Case $V(x) \equiv 0$. We shall consider the linear regression case $h(x)' = (1, x)$ only. Then, $\dot{\mu}_1 = 1$, $\mu_1 = \theta$, and the condition $V(x) \equiv 0$ is equivalent to X being normally distributed, i.e. that

$$g(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-\theta)^2/2\sigma^2}.$$

Thus, assume in the rest of this sub-section that the process X_t itself is Gaussian. Here, we must consider the following two sub-cases:

Sub-case 1: $d_1 + d_2 > 1/2$, $d_1 + 2d_2 < 1$:

In this case, the leading term in the approximation of the process $\hat{\mathcal{V}}_T(x)$ has the form:

$$\hat{\mathcal{V}}_T(x) = \sum_{t=1}^{T} \varepsilon_t (I(X_t \leq x) - G(x) + g(x)(X_t - \theta)) + u_p(T^{1/2}).$$

Using standard Hermite expansions and the ideas in Csörgö and Mielniczuk (1996), one obtains in this case

$$T^{-1/2} \hat{\mathcal{V}}_T(x) \Longrightarrow \Psi(x),$$

where $\Psi(x)$ is a Gaussian process with mean zero and covariance

$$\begin{aligned}
&\mathrm{Cov}(\Psi(x), \Psi(y)) \\
&= \sum_{t=-\infty}^{\infty} E(\varepsilon_0 \varepsilon_t) E\Big([I(X_0 \leq x) - G(x) + g(x)X_0][I(X_t \leq y) - G(y) + g(y)X_t]\Big).
\end{aligned}$$

Sub-case 2: $d_1 + d_2 > 1/2$, $d_1 + 2d_2 > 1$:

In this case, similar calculations based on Hermite expansions indicate that

$$(L(T)L_1(T))T^{d_1+2d_2-1/2})^{-1} \hat{\mathcal{V}}_T(x) \Longrightarrow (1/2)\dot{g}(x)(U_{1,2} - ZZ_2 + 2ZZ_1^2 - 2UZ_1),$$

where the stochastic integrals Z, Z_1, U are the same as in Corollary 4.1 and $U_{1,2}, Z_2$ are all defined on the same probability space, and the last two are given by

$$\begin{aligned}
U_{1,2} &:= \int_{-\infty}^{1} \int_{-\infty}^{1} \int_{-\infty}^{1} \Big\{ \int_0^1 (\tau - x)_+^{-(1-d)} (\tau - x_1)_+^{-(1-d_1)} (\tau - x_2)_+^{-(1-d_1)} d\tau \Big\} \\
&\qquad \times W(dx) W_1(dx_1) W_1(dx_2), \\
Z_2 &:= \int_{-\infty}^{1} \int_{-\infty}^{1} \Big\{ \int_0^1 (\tau - x_1)_+^{-(1-d_1)} (\tau - x_2)_+^{-(1-d_1)} d\tau \Big\} W_1(dx_1) W_1(dx_2).
\end{aligned}$$

5 Proof of the weak convergence of $T^{-1/2}\hat{\mathcal{V}}_T$ under martingale difference errors

This section contains the proof of the weak convergence of $T^{-1/2}\hat{\mathcal{V}}_T$ to a Gaussian process in the case of the martingale difference errors. The result is stated and proved in Lemma 5.4. The proof of this lemma is facilitated by Lemmas 5.1-5.3 below.

Recall (1.3), (1.4) and (1.5). Let

$$\begin{aligned} X_{il} &:= \sum_{k=0}^{l-1} b_{1,k}\xi_{i-k}, \qquad \tilde{X}_{il} := \sum_{k=l}^{\infty} b_{1,k}\xi_{i-k}, \qquad \tilde{X}_{i0} = X_i, \\ G_l(x) &:= P(X_{il} \le x), \quad g_l^{(p)}(x) := \frac{d^p G_l(x)}{dx^p}, \quad l \ge 1,\, p \ge 0; \quad g_l(x) := \frac{dG_l(x)}{dx}. \end{aligned}$$

The following two lemmas are analogous to Lemma 5.1 and 5.2 of Koul and Surgailis (2002), thus their proofs can be deduced from there and Lemma 5.1 of Koul and Surgailis (2001). Here we only give the proof of the second part of Lemma 5.2.

Lemma 5.1 *Under assumptions of (1.4) and (1.5) with $r = 3$, there exist a positive integer $l_0 \ge 1$ and a constant C such that for any $l \ge l_0$, $x \in \mathbb{R}$,*

$$\begin{aligned} |g^{(p)}(x)| + |g_l^{(p)}(x)| &\le C(1+|x|^3)^{-1}, \quad p = 0, 1, 2, \\ |g_l(x) - g_{l-1}(x)| &\le C b_l^2 (1+|x|^3)^{-1}. \end{aligned}$$

Lemma 5.2 *Let $\gamma(x) := (1+|x|^3)^{-1}$ and $f(x)$, $x \in \mathbb{R}$ be a real valued function such that*

$$|f(x)| \le C\gamma(x), \quad |f(x) - f(y)| \le C|x-y|\gamma(x), \tag{5.1}$$

hold for any $x,\ y \in \mathbb{R}$, $|x-y| \le 1$. Then there exists a constant C_1 depending only on C in (5.1), such that for any $x,\ y \in \mathbb{R}$,

$$|f(x+y)| \le C_1\gamma(x)(1 \vee |y|^3). \tag{5.2}$$

Moreover, for any $x_1 < x_2$,

$$\left| \int_{x_1}^{x_2} [f(u+v+w) - f(u+w)]du \right| \le C_1(|v| \vee |v|^3)(1 \vee |w|^3) \int_{x_1}^{x_2} [1+|u|^3]^{-1} du. \tag{5.3}$$

Proof. We only prove the (5.3). First, consider $|v| \le 1$, then by (5.1) and (5.2), the LHS of (5.3) does not exceed

$$C|v| \int_{x_1}^{x_2} (1+|u+w|^3)^{-1} du \le C|v|(1 \vee |w|^3) \int_{x_1}^{x_2} (1+|u|^3)^{-1} du.$$

Next, consider $|v| > 1$. Then the LHS of (5.3) does not exceed

$$C\int_{x_1}^{x_2}(1+|u+v+w|^3)^{-1}du + C\int_{x_1}^{x_2}(1+|u+w|^3)^{-1}du. \tag{5.4}$$

By (5.2), the first term of this bound does not exceed

$$C(1\vee|v+w|^3)\int_{x_1}^{x_2}(1+|u|^3)^{-1}du \le C|v|^3(1\vee|w|^3)\int_{x_1}^{x_2}(1+|u|^3)^{-1}du.$$

The second term of (5.4) follows similarly. This proves the lemma. □

To state the next important result we now let

$$\eta \;:=\; 1-2d_1, \qquad \mu(x,y] := \int_x^y \frac{1}{1+|u|^3}\,du, \quad -\infty \le x \le y \le \infty.$$

Lemma 5.3 *Under the assumptions (1.3), (1.4) and (1.5) with $r = 3$, there exists a constant C such that*

$$\Big|Cov(I(x_1 < X_0 \le x_2),\ I(x_2 < X_t \le x_3))\Big| \;\le\; Ct^{-\eta}\,\mu^{1/2}(x_1,x_2]\,\mu^{1/2}(x_2,x_3], \tag{5.5}$$

for all positive integers t and for all $-\infty \le x_1 \le x_2 \le x_3 \le \infty$.

Proof. Fix a positive integer i. Let $G(x,y) := G(y) - G(x)$, $G_0(x); = I(x \ge 0)$, and

$$\begin{aligned}
U_{i,l}(x_2,x_3) &= G_{l-1}(x_2-\tilde X_{i,l-1}, x_3-\tilde X_{i,l-1}) - G_l(x_2-\tilde X_{i,l}, x_3-\tilde X_{i,l})\\
&= U_{i,l}^{(1)}(x_2,x_3) + U_{i,l}^{(2)}(x_2,x_3),\\
U_{i,l}^{(1)}(x_2,x_3) &= G_l(x_2-\tilde X_{i,l-1}, x_3-\tilde X_{i,l-1}) - G_l(x_2-\tilde X_{i,l}, x_3-\tilde X_{i,l}),\\
U_{i,l}^{(2)}(x_2,x_3) &= G_{l-1}(x_2-\tilde X_{i,l-1}, x_3-\tilde X_{i,l-1}) - G_l(x_2-\tilde X_{i,l-1}, x_3-\tilde X_{i,l-1}).
\end{aligned} \tag{5.6}$$

Then one has the telescoping identity:

$$I(x_2 < X_i \le x_3) - G(x_2,x_3) = \sum_{l=1}^{\infty} U_{i,l}(x_2,x_3). \tag{5.7}$$

It thus suffices to show the following where l_0 is as in Lemma 5.1:

$$E\Big[U_{i,l}(x_2,x_3)\Big]^2 \;\le\; C\mu(x_2,x_3], \qquad l = 1,2,\cdots,l_0, \tag{5.8}$$

$$E\Big[U_{i,l}^{(q)}(x_2,x_3)\Big]^2 \;\le\; C\,l^{-1-\eta}\,\mu(x_2,x_3], \qquad l > l_0, \quad q = 1,2. \tag{5.9}$$

These bounds together with the orthogonality of (5.7) imply

$$
\begin{aligned}
\Big|Cov(I(x_1 < X_0 \le x_2),\, I(x_2 < X_t \le x_3))\Big| &= \Big|\sum_{l=1}^{\infty} EU_{t,t+l}(x_2,x_3)U_{0,l}(x_1,x_2)\Big| \\
\le\ & \sum_{l=1}^{\infty} E^{1/2}[U_{t,t+l}(x_2,x_3)]^2\, E^{1/2}[U_{0,l}(x_1,x_2)]^2 \\
\le\ & C\sum_{l=1}^{\infty}(i+l)^{-(1+\eta)/2}\, l^{-(1+\eta)/2}\mu^{1/2}(x_1,x_2]\mu^{1/2}(x_2,x_3], \\
\le\ & C\, t^{-\eta}\,\mu^{1/2}(x_1,x_2]\mu^{1/2}(x_2,x_3].
\end{aligned}
$$

Proof of (5.8). According to the definition, we have

$$
\begin{aligned}
E\Big[U_{i,l}(x_2,x_3)\Big]^2 &\le 2[EG^2_{l-1}(x_2-\tilde{X}_{i,l-1},x_3-\tilde{X}_{i,l-1}) + EG^2_l(x_2-\tilde{X}_{i,l},x_3-\tilde{X}_{i,l})] \\
&\le 2[EG_{l-1}(x_2-\tilde{X}_{i,l-1},x_3-\tilde{X}_{i,l-1}) + EG_l(x_2-\tilde{X}_{i,l},x_3-\tilde{X}_{i,l})] \\
&\le 4G(x_2,x_3) \le C\int_{x_2}^{x_3}[1+|u|^3]^{-1}du = C\,\mu(x_2,x_3].
\end{aligned}
$$

Proof of (5.9). For $q=1$,

$$
U^{(1)}_{i,l}(x_2,x_3) = \int_{x_2}^{x_3}[g_l(u-b_l\xi_{i-l}-\tilde{X}_{i,l}) - g_l(u-\tilde{X}_{i,l})]\,du.
$$

Apply Lemmas 5.1 and 5.2, (5.3), to obtain the following inequality

$$
|U^{(1)}_{i,l}(x_2,x_3)| \le C(|b_l\xi_{i-l}| \vee |b_l\xi_{i-l}|^3)(1+|\tilde{X}_{i,l}|^3)\int_{x_2}^{x_3}[1+|u|^3]^{-1}du.
$$

On the other hand, (5.6), the integrability of g_l, g_l', which in turn follows from Lemma 5.1, one also obtains

$$
|U^{(1)}_{i,l}(x_2,x_3)| \le C(|b_l\xi_{i-l}| \wedge 1).
$$

These bounds together with the fact that $(x\vee x^3)(x\wedge 1) \le x^2+x^3$, for any $x>0$, and the independence of $\{\xi_i\}$ imply that

$$
\begin{aligned}
E\Big[U^{(1)}_{i,l}(x_2,x_3)\Big]^2 &\le C(E|b_l\xi_{i-l}|^2 + E|b_l\xi_{i-l}|^3)(1+E|\tilde{X}_{i,l}|^3)\int_{x_2}^{x_3}[1+|u|^3]^{-1}du \\
&\le Cb_l^2\mu(x_2,x_3] \le C\,l^{-1-\eta}\mu(x_2,x_3],
\end{aligned}
$$

Note that here the second inequality follows from $E|\tilde{X}_{i,l}|^3 \le C$, which in turn follows from the Rosenthal inequality and (1.5) with $r=3$:

$$
E\Big|\sum_{k=\ell}^{\infty} b_k\xi_k\Big|^3 \le C\sum_{k=\ell}^{\infty}E|b_k\xi_k|^3 + C\Big(\sum_{k=\ell}^{\infty}E|b_k\xi_k|^2\Big)^{3/2} \le C.
$$

This proves (5.9) for $q = 1$.

For $q = 2$, apply Lemmas 5.1 and 5.2 to obtain

$$\begin{aligned} |U_{i,l}^{(2)}(x_2, x_3)| &\leq \Big|\int_{x_2}^{x_3} [g_l(u - \tilde{X}_{i,l-1}) - g_{l-1}(u - \tilde{X}_{i,l-1})]\, du\Big| \\ &\leq C \int_{x_2}^{x_3} b_l^2 (1 + |u - \tilde{X}_{i,l-1}|^3)^{-1}\, du \\ &\leq C b_l^2 (1 + |\tilde{X}_{i,l-1}|^3)\, \mu(x_2, x_3] \end{aligned}$$

Again, as $|U_{i,l}^{(2)}(x_2, x_3)| \leq 2$, we obtain (5.9) for $q = 2$. This completes the proof of (5.5). $\square$

Now consider the process $\mathcal{V}_T$ and assume that (1.1) holds. Note that $\mathcal{V}_T(-\infty) = 0$, $\mathcal{V}_T(\infty) = \sum_{t=1}^T \varepsilon_t$. Thus the process $T^{-1/2}\mathcal{V}_T$ is well defined in $D[-\infty, \infty]$. The next lemma gives the weak convergence of the process $T^{-1/2}\mathcal{V}_T$ to a continuous Gaussian limit in $D[-\infty, \infty]$. Recall that G stands for the d.f. of X_0.

Lemma 5.4 *Assume (1.1) with* $\sup_t E\varepsilon_t^4 < \infty$*, (1.3), (1.4) and (1.5) with* $r = 3$ *hold. Then,*

$$T^{-1/2}\mathcal{V}_T \Longrightarrow \sigma_e\, B \circ G \ \text{ in the space } \ D[-\infty, \infty], \tag{5.10}$$

$$T^{-1/2}\hat{\mathcal{V}}_T \Longrightarrow W_G \ \text{ in the space } \ D[-\infty, \infty], \tag{5.11}$$

where $B \circ G$ *is a continuous Brownian motion on* $\mathbb{R}$ *with respect to time* G*, and* W_G *is a continuous mean zero Gausian process on* $\mathbb{R}$ *with* $W_G(-\infty) = 0$*, and the covariance function*

$$K(x, y) := \sigma_e^2 [G(x \wedge y) - \nu(x)' H^{-1} \nu(y)], \qquad x,\, y \in \mathbb{R}.$$

Proof. Apply the CLT for martingales (Hall and Heyde: 1980, Corollary 3.1) to show that the finite dimensional distributions converge weakly to the right limit, under the assumed conditions.

To prove the tightness, fix $-\infty < x_1 < x_2 < x_3 < \infty$. Let $V_T \equiv T^{-1/2}\mathcal{V}_T$. Then

$$\begin{aligned} &[V_T(x_3) - V_T(x_2)]^2 [V_T(x_2) - V_T(x_1)]^2 \\ &= T^{-2} \left[\sum_{t=1}^T \varepsilon_t I(x_2 < X_t \leq x_3)\right]^2 \left[\sum_{t=1}^T \varepsilon_t I(x_1 < X_t \leq x_2)\right]^2 \\ &= T^{-2} \sum_{s,t,k,l} Y_s Y_j W_k W_l \end{aligned}$$

where $Y_s = \varepsilon_s I(x_2 < X_s \leq x_3)$ and $W_s = \varepsilon_s I(x_1 < X_s \leq x_2)$.

Now, if the largest index among s, t, k, l is not matched by any other, then $E\{Y_s Y_t W_k W_l\} = 0$. Also, since the two intervals $(x_2, x_3]$ and $(x_1, x_2]$ are disjoint, $Y_s W_s \equiv 0$ for all s, and because the errors are homoscedastic martingale differences, see (1.1), independent of X_t's, we have $E\{Y_s W_t Y_k^2\} = 0$, for all $s < k$, $t < k$. Hence,

$$E\left\{T^{-2}\sum_{s,t,k,l} Y_s Y_t W_k W_l\right\} = T^{-2}\sum_{s,t<k}\Big[E\{Y_s Y_t W_k^2\} + E\{W_s W_t Y_k^2\}\Big].$$

But, for a k fixed,

$$\begin{aligned}
E\{Y_s Y_t W_k^2\} &= E(\varepsilon_s\varepsilon_t\varepsilon_k^2)\, EI(x_2 < X_s \le x_3,\, x_2 < X_t \le x_3) I(x_1 < X_k \le x_2) \\
&= 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad s \ne t < k, \\
&= \sigma_e^4\, EI(x_2 < X_s \le x_3) I(x_1 < X_k \le x_2), \quad s = t < k.
\end{aligned}$$

A similar fact is true for $E\{W_s W_t Y_k^2\}$. Consequently, summing over k from 1 to T, in view of the inequality (5.5), we have

$$\begin{aligned}
&E\left\{[V_T(x_3) - V_T(x_2)]^2 [V_T(x_2) - V_T(x_1)]^2\right\} \\
&\le\; C\,T^{-2}\sum_{1\le s<k\le T}\Big\{(k-s)^{-\eta}\mu^{1/2}(x_1,x_2]\mu^{1/2}(x_2,x_3] + [G(x_2)-G(x_1)][G(x_3)-G(x_2)]\Big\} \\
&\le\; C\Big\{T^{-\eta}\mu(x_1,x_3] + [G(x_3)-G(x_1)]^2\Big\}.
\end{aligned}$$

This together with Theorem 12.1, equations (12.5), (12.10) of Billingsley (1968) and a chaining argument similar to that of Dehling and Taqqu (1989) and Koul and Mukherjee (1993) yields the tightness of the process V_T. Details are left out for the sake of brevity. This completes the proof of (5.10).

To prove (5.11), first we note that from (2.2) and in view of the Ergodic Theorem, we obtain that

$$\begin{aligned}
T^{-1/2}\hat{\mathcal{V}}_T(x) &= T^{-1/2}\Big[\mathcal{V}_T(x) - \mathcal{Z}_T' H^{-1}\nu(x)\Big] - T^{-1/2}\mathcal{Z}_T'\,[\overline{H}_T^{-1}\,\overline{\nu}_T(x) - H^{-1}\nu(x)] \\
&= T^{-1/2}\sum_{t=1}^{T}\varepsilon_t[I(X_t\le x) - h(X_t)' H^{-1}\nu(x)] + u_p(1),
\end{aligned}$$

where $u_p(1)$ is a sequence of stochastic processes on $\mathbb{R}$ tending to zero uniformly, in probability. This together with (5.10) and the uniform continuity of G and $\nu(x)$ completes the proof of (5.11). $\square$

Acknowledgment: The authors gratefully acknowledge the helpful comments of the Co-Editor Don Andrews and two anonymous referees.

REFERENCES

An, H. Z. and B. Cheng (1991) A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review* **59**, 287-307.

Avram, F. and M.S. Taqqu (1987) Noncentral limit theorems and Appell polynomials. *Annals Probability* **15**, 767-775.

Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**, 5-59.

Baillie, R.T. and T. Bollerslev (1994) The long memory of the forward premium. *Journal of International Money and Finance* **13**, 309-324.

Baillie, R.T. and T. Bollerslev (2000) The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* **19**, 471-488.

Baillie, R., C-F. Chung, and M.A. Tieslau (1996) Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal Applied Econometrics* **11**, 23-40.

Bates, D.M and D.G. Watts (1988) *Nonlinear analysis and its applications.* J. Wiley and Sons, N.Y.

Beran, J. (1992) Statistical methods for data with long-range dependence. (With discussion). *Statistical Science* **7**, 404-427.

Beran, J. (1994) *Statistics for Long-Memory Processes.* Monographs on Statistics and Applied Probability, **61**. Chapman and Hall. N.Y.

Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley and Sons, New York.

Booth, G.G., R.R. Kaen, and P.E. Koves (1982) R/S analysis of foreign exchange rates under two international money regimes. *Journal of Monetary Economics* **10**, 407-415.

Cheung, Y.-W. (1993) Long memory in foreign exchange rates. *Journal of Business and Economic Statistics* **11**, 93-101.

Csörgö, M. and J. Mielniczuk (1996) The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probability Theory and Related Fields* **104**, 15-25.

Dehling, H. and M.S. Taqqu (1989) The empirical process of some long-range dependent sequences with an application to U-statistics. *Annals of Statistics* **17**, 1767-1783.

Diebolt, J. and J. Zuber (1999) Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statistics and Probability Letters* **42(1)**, 53–60.

Dobrushin, R.L. and P. Major (1979) Non-central limit theorems for non-linear functionals of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 27-52.

Engel, C. (1996) The forward discount anomaly and the risk premium: a survey of recent evidence. *Journal of Empirical Finance* **3**, 123-192.

Fama, E.F. (1984) Spot and forward exchange rates. *Journal of Monetary Economics* **14**, 319-338.

Giovannini, A. and P. Jorion (1987) Interest rates and risk premium in the stock market and in the foreign exchange market. *Journal of International Money and Finance* **6**, 107-124.

Giraitis, L. and D. Surgailis (1999) Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference* **80**, 81-93.

Hai, W., N.C. Mark, and Y. Wu (1987) Understanding spot and forward exchange regressions. *Journal of Applied Econometrics*, **12**, 715-734.

Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and Its Application.* Academic Press. New York.

Hodrick, R.J. (1989) Risk, uncertainty and exchange rates. *Journal of Monetary Economics* **23**, 433-459.

Ho, H.C. and T. Hsing (1996) On the asymptotic expansion of the empirical process of long memory moving averages. *Annals of Statistics* **24**, 992-1024.

Ho, H.C. and T. Hsing (1997) Limit theorems for functionals of moving averages. *Annals of Probability* **25**, 1636-1669.

Khmaladze, E.V. (1981) Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and Applications* **26**, 240-257.

Koul, H.L. and K. Mukherjee (1993) Asymptotics of R-, MD- and LAD-estimators in linear regression models with long-range dependent errors. *Probability Theory and Related Fields* **95**, 535-553.

Koul, H.L. and W. Stute (1999) Nonparametric model checks for time series. *Annals of Statistics* **27**, 204-236.

Koul, H.L. and D. Surgailis (1997) Asymptotic expansion of M-estimators with long memory errors. *Annals of Statistics* **25**, 818-850.

Koul, H. L. and D. Surgailis (2000) Asymptotic normality of the Whittle estimator in linear regression models with long memory errors. *Statistical Inference for Stochastic Processes* **3**, 129-147.

Koul, H.L. and D. Surgailis (2001) Asymptotics of the empirical process of long memory moving averages with infinite variance. *Stochastic Processes and their Applications* **91**, 309-336.

Koul, H.L. and D. Surgailis (2002) Asymptotic expansion of the empirical process of long memory moving averages. In *Empirical process techniques for dependent data*. pp 213-239. **Eds.** H. Dehling, T. Mikosch, and M. Sørenson. Birkhauser, Boston, USA.

Lo, A. (1991) Long term memory in stock market prices. *Econometrica* **59**, 1279-1313.

MacKinnon, J.G. (1992) Model specification tests and artificial regression. *Journal Econometric Literature* **XXX**, 102-146.

Maynard, A. and P.C.B. Phillips (2001) Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* **16**, 671-708.

Mussa, M.L. (1982) A model of exchange rate dynamics. *Journal of Political Economy*, **90**, 74-104.

Seber, G.A.F. and C.J. Wild (1989) *Nonlinear regression.* J. Wiley and Sons. N.Y.

Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* **25**, 613-641.

Stute, W., W. González Manteiga, and M. Presedo Quindimil (1998) Bootstrap approximations in model checks for regression. *Journal of American Statistical Association* **93**, 141-149.

Stute, W., S. Thies, and L.X. Zhu (1998) Model checks for regression: An innovation process approach. *Annals of Statistics* **26**, 1916-1934.

Su, J. Q. and L. J. Wei (1991) A lack-of-fit test for the mean function in a generalized linear model. *Journal of American Statistical Association* **86**, 420-426.

Hira L. Koul
Department of Statistics & Probability
Michigan State University
East Lansing MI 48824-1027.
U.S.A.

Richard T. Baillie
Department of Economics
Michigan State University
East Lansing MI 48824-1027.
U.S.A.
Queen Mary
University of London
U.K.

Donatas Surgailis
Institute of Mathematics & Informatics
A kademijos, 4
2600 Vilnius
LITHUANIA