A Convergent t-statistic in Spurious Regressions*

Yixiao Sun
Department of Economics
University of California, San Diego

First Version: December 2002
This Version: July 2003

*I am very grateful to Bruce Hansen and two anonymous referees for helpful comments and suggestions. Correspondence to: Yixiao Sun, Department of Economics, 0508, University of California, San Diego, La Jolla, CA 92093-0508, USA. E-mail: yisun@ucsd.edu. Tel: (858) 534-4692. Fax: (858)534-7040.

Abstract

This paper proposes a convergent t-statistic for spurious regressions. The t-statistic is based on the heteroscedasiticity and autocorrelation consistent (HAC) standard error estimate with the bandwidth proportional to the sample size. Using autocovariances of large lags, the so-defined HAC estimator is capable of capturing the high persistence of the regressor and regression residuals. It is shown that the resulting t-statistic converges to a non-degenerate limiting distribution for all cases of spurious regressions considered in the literature. This finding suggests that inferences based on the new asymptotic theory developed in this paper will not result in the finding of a significant relationship that does not actually exist.

Keywords: Spurious Regression, Fractional Process, HAC Estimator.

JEL Classification Numbers: C22

1 Introduction

Since the first Monte Carlo study by Granger and Newbold (1974), much effort has been taken to understand the nature of spurious regressions. Phillips (1986) developed an asymptotic theory for a regression between $I(1)$ processes, showing that the usual t-statistic does not have a limiting distribution but diverges at the rate of $\sqrt{T}$ as the sample size T increases. Extending Phillips' (1986) approach, Durlauf and Phillips (1988) and Marmol (1995, 1998) found that the usual t-statistic diverges at the same rate in a regression between an $I(1)$ process and a linear trend and between two nonstationary $I(d)$ processes. More recently, Tsay and Chung (2000) found that the usual t-statistic diverges, albeit at a slower rate, in a regression between two stationary $I(d)$ processes, as long as their memory parameters sum up to a value greater than 0.5. The divergence of the usual t-statistic seems to be a defining characteristic of a spurious regression. In this paper, we show that the divergence of the usual t-statistic arises from the use of a standard error that underestimates the true variation of the OLS estimator. We propose an alternative estimator of the standard error and use it to construct the t-statistic. We show that the resulting t-statistic no longer diverges.

The standard error estimator we used is the heteroscedasiticity and autocorrelation consistent (HAC) standard error estimator with bandwidth (M) proportional to the sample size. Specifically, we set $M = bT$ for some $b \in (0, 1]$. This sharply contrasts with the usual HAC estimator in that the bandwidth is usually taken to grow at a slower rate than the sample size. The optimal rate of growth depends on the shape of the underlying spectral density. In a linear regression model in which the regressors and errors are independent AR(1) processes with the same autoregressive parameter γ, Andrews (1991) showed that the optimal bandwidth increases with γ. This result suggests that the bandwidth should be larger for more persistent processes. In a spurious regression, both the regressors and the regression residuals are highly persistent. It turns out that the bandwidth needs to grow at the same rate as the sample size to capture the high autocorrelation. In other words, for the t-statistic to converge, it is necessary to include autocovariances of large lags (up to lag $[bT]$) in the construction of the HAC estimator.

We show that when the OLS estimator is scaled by the new standard error, the resulting t-statistic converges to a well-defined distribution. This is true for regressions between two independent fractional processes, stationary or nonstationary, and between a fractional process and a linear trend. For all the cases considered, the limiting distributions depend on the kernel used and the persistence of the underlying processes. They are nonstandard and their probability densities can be estimated by simulations. Our findings suggest that inferences based on the new asymptotic theory and critical values obtained via simulations will not lead to the finding of a spurious relationship. This result sheds some new light on spurious regressions. The inevitable significance of the usual t-ratio can be attributed to the use of a small bandwidth ($M = o(T)$) in the HAC estimation.

The results of this paper are related to those of Kiefer and Vogelsang (2003). They considered the specification of $M = bT$ in hypothesis testing when the time series are weakly dependent. In their framework, different specifications ($M = bT$ or $M = o(T)$) simply give arise different approximations. In finite samples, any observed value of the pair (M, T) may be regarded as compatible with $M = bT$ or $M = o(T)$. We can thus use the approximation that appears to be more accurate. These discussions apply to spurious regressions as well. In the case of spurious regression, we can go one step further by arguing that it is natural to set $M = bT$. In other words, the specification of $M = bT$ should not be viewed as just another way to approximating the true distribution. Instead, we think it is natural to choose a large bandwidth from the beginning and then invoke the '$M = bT$' asymptotics to make inferences. First, the motivation for the conventional choice of $M = o(T)$ is lost when the time series are strongly dependent. The specification $M = o(T)$ is used in order to ensure the consistency of the HAC estimator. In the presence of strongly dependent processes, the HAC matrix is not well defined (the spectral density is unbounded at zero). Hence, it is not meaningful to use consistency as a criterion to choose the growth rate of M. Second, under the rate condition $M = bT$, the standard error estimate captures the variability of the coefficient estimate whereas under $M = o(T)$, the standard error estimate under-estimates the variability by an order of magnitude. A more natural standard estimate error should be the one that reflects the uncertainty in the parameter estimate.

Other related papers include Kiefer and Vogelsang (2002a, 2002b) who set $M = T$. Their motivation is to develop asymptotically valid tests that are free from the bandwidth selection and have good size and power properties. Other papers that use or investigate the HAC estimator without truncation include Jansson (2002), Phillips, Sun and Jin (2003) and Sun (2002).

The rest of the paper is organized as follows. Section 2 considers the spurious regressions with nonstationary fractional processes and linear trends. It establishes the asymptotic distributions when $M = bT$ for some positive b. Section 3 extends the results in Section 2 to stationary fractional processes. Section 4 provides kernel estimates of the probability densities of the limiting t-statistics in Sections 2 and 3. Section 5 concludes. All proofs are given in the appendix.

Throughout the paper, " $\Rightarrow$ " signifies convergence in the $D[0,1]^k$ space endowed with the Skorohod topology. $1\{\lambda = \mu\}$ is the indicator function. $[\cdot]$ signifies the integer part.

2 Spurious Regressions with Nonstationary Fractional Processes

Consider the following data generating process:

$$y_t = \alpha + x_t\beta + u_t, \quad t = 1, 2, ..., T, \tag{1}$$

where x_t and u_t are nonstationary $I(d)$ processes such that

$$(1 - L)^{d_x} x_t = \varepsilon_{xt} 1\{t > 0\}, \ (1 - L)^{d_u} u_t = \varepsilon_{ut} 1\{t > 0\} \tag{2}$$

where d_x, $d_u > 1/2$ and ε_{xt} and ε_{ut} are two weakly dependent processes.

We further assume that x_t and u_t satisfy the following functional central limit theorem (FCLT):

$$\begin{pmatrix} T^{-(2d_x-1)/2} & 0 \\ 0 & T^{-(2d_u-1)/2} \end{pmatrix} \begin{pmatrix} x_{[Tr]} \\ u_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} V_x(r) \\ V_u(r) \end{pmatrix} \tag{3}$$

where

$$V_x(r) = \frac{1}{\Gamma(d_x)} \int_0^r (r-s)^{d_x-1} dB^x(s), V_u(r) = \frac{1}{\Gamma(d_u)} \int_0^r (r-s)^{d_u-1} dB^u(s), \tag{4}$$

$B(s) = (B^x(s), B^u(s))$ is a bivariate Brownian motion with positive definite long run variance

$$\omega = \begin{pmatrix} \omega_x^2 & \omega_{xu} \\ \omega_{xu} & \omega_u^2 \end{pmatrix}. \tag{5}$$

The FCLT holds under a wide range of primitive conditions (e.g. Akonom and Gourieroux 1987; Marinucci and Robinson 2000). When x_t and u_t are unit root processes, the limiting process $V(r) = (V_x(r), V_u(r))'$ reduces to a scaled Brownian motion. For a general nonstationary fractional process, the limiting process is a type II fractional Brownian motion (Marinucci and Robinson 1999).

When $d_u \geq d_x$, there is no cointegrating relationship between y_t and x_t, and we have a spurious linear model. When $d_u < d_x$, y_t and x_t are fractionally cointegrated with cointegrating vector $(1, -\beta)$. However, since the long run equilibrium error u_t is nonstationary, Marmol (1998) referred to the second situation as a partially spurious case. For convenience, we will refer to both cases as spurious cases.

It is worth noting that in the above FCLT we implicitly define a nonstationary $I(d)$ process z_t as

$$z_t = (1-L)^{-d} w_t = \sum_{k=0}^{t-1} \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(k+1)} w_{t-k}, \tag{6}$$

for some weakly dependent process w_t. This definition was used in Robinson (1994a), Phillips (1999), Marmol and Velasco (2002), among others. There is another often-used definition, in which a nonstationary $I(d)$ process is defined as the partial sum of an $I(d-1)$ process and a stationary fractional process is defined as an infinite order moving average:

$$z_t = \sum_{k=0}^{\infty} \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(k+1)} w_{t-k}. \tag{7}$$

The second definition was used in a number of papers including Hurvich and Ray (1995) and Velasco (1999). Different definitions imply different functional central limit theorems (Marinucci and Robinson 2000). In this section we use the first definition. Our results hold with obvious modifications if the second definition were used.

Consider regressing y_t on a constant and x_t,

$$y_t = \widehat{\alpha} + \widehat{\beta} x_t + \widehat{u}_t, t = 1, ..., T. \tag{8}$$

The ordinary least squares estimate of β is given by

$$\widehat{\beta} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \tag{9}$$

where $\bar{x} = \sum_{t=1}^T x_t/T$ and $\bar{y} = \sum_{t=1}^T y_t/T$. The heteoscedasiticity and autocorrelation consistent t-statistic is $\widehat{t}_\beta = (\widehat{\beta} - \beta)/\widehat{\sigma}_\beta$, where $\widehat{\sigma}_\beta$ is the HAC estimator defined as

$$\widehat{\sigma}_\beta^2 = \left(\sum_{t=1}^T (x_t - \bar{x})^2 \right)^{-1} T\widehat{\Omega} \left(\sum_{t=1}^T (x_t - \bar{x})^2 \right)^{-1}, \tag{10}$$

where

$$\widehat{\Omega} = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\widehat{\Gamma}(j), \tag{11}$$

$$\widehat{\Gamma}(j) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-j} (x_{t+j} - \bar{x})\, \widehat{u}_{t+j} \widehat{u}_t \,(x_t - \bar{x}) & \text{for } j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} (x_{t+j} - \bar{x})\, \widehat{u}_{t+j} \widehat{u}_t \,(x_t - \bar{x}) & \text{for } j < 0 \end{cases} \tag{12}$$

and $k(\cdot)$ is a kernel function and M is the bandwidth parameter.

In order get a consistent estimate of the long run variance of $(x_t - \bar{x})\, \widehat{u}_t$, conventional asymptotic theory assumes that $M \to \infty$ such that $M = o(T)$. However, $(x_t - \bar{x})\, \widehat{u}_t$ is nonstationary and the variance of $\sum_{t=1}^T (x_t - \bar{x})\, \widehat{u}_t/\sqrt{T}$ does not converge. In other words, the sum $\sum_{j=-\infty}^{\infty} ||\widehat{\Gamma}(j)||$ is infinite with the probability approaching one as $T \to \infty$. The infiniteness of this sum invalidates the usual truncation argument. As it becomes clear later on, when M is taken to grow at a slower rate than T, the estimate $\widehat{\Omega}$ underestimates the variability of $\sum_{t=1}^T (x_t - \bar{x})\, \widehat{u}_t/\sqrt{T}$ by an order of magnitude. The resulting t-statistic will diverge, leading to the spurious rejection. A natural question is: what if M grows at the same rate as T. In the rest of the paper, we set $M = bT$ for some $b \in (0, 1]$ and provide an answer to this question. For convenience, we call the asymptotics under the assumption of $M = bT$ the large M asymptotics.

To ensure the positive definiteness of $\widehat{\Omega}$, we assume that the kernel function belongs to the following class:

$$\mathcal{K} = \{k(\cdot) : [-1, 1] \to [0, 1] \mid k(x) = k(-x),\ k(0) = 1,\ \text{and } K(\lambda) \geq 0, \forall \lambda \in \mathbb{R}\}, \tag{13}$$

where

$$K(\lambda) = \int_{-\infty}^{\infty} k(x) \exp(-i\lambda x) dx. \tag{14}$$

For a kernel function $k(x) \in \mathcal{K}$, we have $\int_{-1}^1 \int_{-1}^1 k(r - s) f(r) f(s) dr ds \geq 0$ for any square integrable function $f(x)$. In other words, the functions in $\mathcal{K}$ are positive semi-definite.

The following theorem establishes the asymptotic distributions of $\widehat{\beta}$, $\widehat{\sigma}_\beta$ and the resulting t-statistic $\widehat{t}_\beta$. The theorem uses the following notation:

$$\widetilde{V}_x(r) = V_x(r) - \int_0^1 V_x(\tau)d\tau, \tag{15}$$

$$\widetilde{V}_u(r) = V_u(r) - \int_0^1 V_u(\tau)d\tau, \tag{16}$$

and

$$\widetilde{V}_{u,x}(r) = \widetilde{V}_u(r) - \left(\int_0^1 \widetilde{V}_x(\tau)\widetilde{V}_x(\tau)d\tau\right)^{-1} \left(\int_0^1 \widetilde{V}_x(\tau)\widetilde{V}_u(\tau)d\tau\right) \widetilde{V}_x(r). \tag{17}$$

Theorem 1 *Assume that x_t and u_t satisfy the functional central limit theorem in (3) and y_t is generated by (1). Let $k(x)$ be a continuous function in $\mathcal{K}$ and $M = bT$ for $b \in (0,1]$. Then, as $T \to \infty$,*

$$\begin{aligned}
&T^{d_x-d_u}\left(\widehat{\beta}-\beta\right) \Rightarrow \left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_u(r)dr\right)\left(\int_0^1 \widetilde{V}_x^2(r)dr\right)^{-1},\\
&T^{2d_x-2d_u}\widehat{\sigma}_\beta^2 \Rightarrow \left(\int_0^1 \widetilde{V}_x^2(r)dr\right)^{-2}\int_0^1\int_0^1 \widetilde{V}_x(r)\widetilde{V}_{u,x}(r)k(\frac{r-s}{b})\widetilde{V}_{u,x}(s)\widetilde{V}_x(s)drds,\\
&\widehat{t}_\beta \Rightarrow \left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_u(r)dr\right)\left(\int_0^1\int_0^1 \widetilde{V}_x(r)\widetilde{V}_{u,x}(r)k(\frac{r-s}{b})\widetilde{V}_{u,x}(s)\widetilde{V}_x(s)drds\right)^{-1/2}.
\end{aligned} \tag{18}$$

Theorem 1 shows that when $M = bT$, the variance estimate $\widehat{\sigma}_\beta^2$ captures the variability of $\hat{\beta}$ in the sense that both $\widehat{\beta}-\beta$ and $\widehat{\sigma}_\beta$ are of the same order $O_p(T^{-d_x+d_u})$. In contrast, the conventional variance estimator (also called the OLS variance estimator), which is $T^{-1}\sum_{t=1}^T \widehat{u}_t^2\left(\sum_{t=1}^T (x_t-\bar{x})^2\right)^{-1}$, is not only inconsistent but also underestimates var$(\widehat{\beta})$ by an order of magnitude. This is because both the regressor and the regression residuals are highly persistent in a spurious regression while the OLS variance estimator ignores this autocorrelation structure. When the OLS estimator is normalized by the conventional standard error estimate, the resulting t-statistic is bound to diverge. The rate of divergence is $\sqrt{T}$, as shown by Phillips (1986) and Marmol (1998). In contrast, the HAC estimator with $M = bT$ incorporates autocovariances of large lags and delivers a standard error estimate that is of the same stochastic order as $\widehat{\beta}-\beta$. Based on such a HAC estimate, the t-statistic is stochastically bounded and converges to a well-defined distribution.

We may use Theorem 1 to do hypothesis testing. Consider testing $H_0 : \beta = \beta_0$ against $H_1 : \beta = \beta_1$. Under the null hypothesis, the t-statistic $\widehat{t}_{\beta_0} = (\widehat{\beta}-\beta_0)/\widehat{\sigma}_\beta$ converges to the limiting distribution given by (18). Under the alternative hypothesis $\widehat{t}_{\beta_0} = O_p(T^{d_x-d_u})$, which is easily seen by writing $\widehat{t}_{\beta_0}$ as $(\widehat{\beta}-\beta_1)/\widehat{\sigma}_\beta + (\beta_1-\beta_0)/\widehat{\sigma}_\beta$. Therefore, when $d_x > d_u$, $\widehat{t}_{\beta_0}$ diverges under H_1 and the test is consistent. When

$d_x \leq d_u$, $\widehat{t}_{\beta_0}$ converges under both H_0 and H_1 and the test is inconsistent. In the latter case, the signal-noise ratio $\sum_{t=1}^{T} x_t^2 / \sum_{t=1}^{T} u_t^2$ is stochastically bounded in large samples. As a result, we can not consistently estimate the slope coefficient even in large samples.

Now we consider the spurious regression between a nonstationary $I(d)$ process and a linear trend. The data generating process for u_t is the same as before so that the invariance principle in (3) holds for $T^{-(2d_u-1)/2}u_{[Tr]}$. The data generating process for x_t is replaced by $x_t = t$ so that $T^{-(2d_x-1)/2}x_{[Tr]} \to r$ for $d_x = 3/2$. We regress y_t on a constant and x_t and construct the t-statistic as before. Using the arguments similar to the proof of Theorem 1, we can prove the following theorem immediately. The details are omitted.

Theorem 2 *Assume $x_t = t$, u_t satisfies the functional central limit theorem in (3), and $y_t = \alpha + \beta t + u_t$. Let $k(x)$ be a continuous function in $\mathcal{K}$ and $M = bT$ for $b \in (0, 1]$. Then*

$$T^{d_x-d_u}\left(\widehat{\beta} - \beta\right) \Rightarrow 12\left(\int_0^1 r\widetilde{V}_u(r)dr\right),$$

$$T^{2d_x-2d_u}\widehat{\sigma}_\beta^2 \Rightarrow 144\left(\int_0^1\int_0^1 (r-1/2)\widetilde{V}_{u,t}(r)k(\frac{r-s}{b})\widetilde{V}_{u,t}(s)(s-1/2)drds\right),$$

$$\widehat{t}_\beta \Rightarrow \left(\int_0^1 r\widetilde{V}_u(r)dr\right)\left(\int_0^1\int_0^1 (r-1/2)\widetilde{V}_{u,t}(r)k(\frac{r-s}{b})\widetilde{V}_{u,t}(s)(s-1/2)drds\right)^{-1/2}, \tag{19}$$

where

$$d_x = 3/2 \text{ and } \widetilde{V}_{u,t}(r) = \widetilde{V}_u(r) - \left(\int_0^1 \tau\widetilde{V}_u(\tau)d\tau\right)(12r-6). \tag{20}$$

Theorem 2 shows that when $M = bT$, the t-statistic is convergent, as in the case of a regression between two nonstationary fractional processes. In contrast, the usual t-statistic diverges at the rate of $\sqrt{T}$ (see Durlauf and Phillips (1988) and Marmol and Velasco (2002)). Our finding is consistent with a result by Phillips (1998), who considered regressing a unit root process on a complete orthonormal system in $L_2[0,1]$. He showed that the t-statistic based the usual HAC standard error with bandwidth M is of order $O_p\left((T/M)^{1/2}\right)$. From this, one may deduce heuristically that when $M = bT$, the t-statistic is stochastically bounded whereas when $M = o(T)$, the t-statistic diverges at the rate of $\sqrt{T/M}$.

As before, Theorem 2 permits us to make inference on the trend coefficient. When d_u is less than 3/2, the t-test is consistent. When $d_u \geq 3/2$, the t-test is inconsistent, reflecting the difficulty in distinguishing the signal from the noise. Similar results have also been obtained by Marmol and Velasco (2002) who analyzed the effects of spuriously detrending a nonstationary fractional process. Put in out context, they used the long run variance estimate of the form:

$$\widehat{\Omega} = \sum_{j=-T+1}^{T-1} \widehat{\Gamma}_{xx}(j)\widehat{\Gamma}_{\hat{u}\hat{u}}(j), \tag{21}$$

where $\widehat{\Gamma}_{xx}(j) = T^{-1}\sum_{t=1}^{T-|j|} x_{t+|j|}x_t$, $\widehat{\Gamma}_{\hat{u}\hat{u}}(j) = T^{-1}\sum_{t=1}^{T-|j|} \hat{u}_{t+|j|}\hat{u}_t$. Due to the difference in the construction of the variance estimate, their limiting distributions are different from ours. Nevertheless, the basic message is the same: it is necessary to let $M = bT$ to deliver a convergent t-statistic.

Together with Theorem 1, Theorem 2 shows that when $M = bT$, the t-statistic converges in distribution in the spurious regression with nonstationary fractional processes. This finding implies that the t-test based on the large M asymptotics will not point to a significant relationship between two independent processes.

3 Spurious Regressions with Stationary Fractional Processes

In this section, we consider the regression between two stationary $I(d)$ processes and that between a stationary $I(d)$ process and a linear trend.

Consider the following data generating process

$$y_t = \alpha + x_t\beta + u_t \tag{22}$$

where x_t and u_t are linear Gaussian processes defined by

$$x_t = \sum_{j=0}^{\infty} a_j\varepsilon_{x,t-j} \text{ and } u_t = \sum_{j=0}^{\infty} b_j\varepsilon_{u,t-j}. \tag{23}$$

We assume the spectral density matrix of $(x_t, u_t)'$ is of the form

$$\begin{pmatrix} f_x(\lambda) & f_{xu}(\lambda) \\ f_{ux}(\lambda) & f_u(\lambda) \end{pmatrix} = \begin{pmatrix} \lambda^{-2d_x}\varphi_x(\lambda) & \lambda^{-(d_x+d_u)}\varphi_{xu}(\lambda) \\ \lambda^{-(d_x+d_u)}\varphi_{ux}(\lambda) & \lambda^{-2d_u}\varphi_u(\lambda) \end{pmatrix} \tag{24}$$

where $0 < d_x, d_u < 0.5$ and the $\varphi(\cdot)'s$ are continuous functions. Let $\varphi_x(0) = \omega_x^2$, $\varphi_u(0) = \omega_u^2$, $\varphi_{xu}(0) = \omega_{xu}$. We assume that the matrix ω as defined in (5) is positive definite.

Given the above spectral density matrix, x_t and u_t have spectral representations:

$$x_t = \int_{-\pi}^{\pi} \exp(it\lambda)\,|\lambda|^{-d_x}\,dW_x(\lambda) \text{ and } u_t = \int_{-\pi}^{\pi} \exp(it\lambda)\,|\lambda|^{-d_u}\,dW_u(\lambda), \tag{25}$$

$t = 1, 2, ..., T$, where $W_x(\cdot)$ and $W_u(\cdot)$ are complex-valued, orthogonal-increment Gaussian processes satisfying

$$W_z(d\lambda) = \overline{W_z(-d\lambda)}, \text{ for } z = x, u \tag{26}$$

$$EW_x(d\lambda)\overline{W_x(d\lambda)} = \varphi_x(\lambda)d\lambda, EW_u(d\lambda)\overline{W_u(d\lambda)} = \varphi_u(\lambda)d\lambda, \tag{27}$$

and

$$EW_x(d\lambda)\overline{W_u(d\mu)} = \varphi_{xu}(\lambda)1\{\lambda = \mu\}d\lambda. \tag{28}$$

The spectral representations help establish the following lemma, which will be used extensively in proving the asymptotic properties of the OLS estimator and the t-statistic. Before stating the lemma, we introduce some notation. Let

$$\sigma_{xu} = \sigma_{ux} = Ex_t u_t,\ \sigma_x^2 = Ex_t^2, \tag{29}$$

and define the random vector element

$$\begin{aligned} S_T(r) &= (S_T^x(r), S_T^u(r), S_T^{xu}(r)) \\ &= \left(T^{-(d_x+1/2)} \sum_{t=1}^{[Tr]} x_t, T^{-(d_u+1/2)} \sum_{t=1}^{[Tr]} u_t, T^{-d_x-d_u} \sum_{t=1}^{[Tr]} (x_t u_t - \sigma_{xu}) \right). \end{aligned} \tag{30}$$

Note that $S_T(r) \in D[0,1]^3$, the product space of all real valued functions on $[0,1]$ that are right continuous and possess finite left limits. We endow the product space with the product σ-algebra, which is generated by the open sets with respect to the metric that induces the Skorohod topology on the component space. The so-defined product σ-algebra makes $D[0,1]^3$ complete and separable.

Lemma 3 *Let x_t and u_t be the time series defined by (25). If $d_x, d_u \in (0, 1/2)$ and $d_x + d_u > 1/2$, then*

$$S_T(r) \Rightarrow (B_{d_x}(r), B_{d_u}(r), Z(r)), \tag{31}$$

where

$$B_{d_x}(r) = \int_{-\infty}^{\infty} \frac{\exp(i\xi r) - 1}{i\xi} |\xi|^{-d_x} dW_x(\xi), \tag{32}$$

$$B_{d_u}(r) = \int_{-\infty}^{\infty} \frac{\exp(i\eta r) - 1}{i\eta} |\eta|^{-d_u} dW_u(\eta), \tag{33}$$

and

$$Z(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(i(\xi + \eta)r) - 1}{i(\xi + \eta)} |\xi|^{-d_x} |\eta|^{-d_u} dW_x(\xi) dW_u(\eta). \tag{34}$$

Note that $B_{d_x}(r)$ and $B_{d_u}(r)$ are spectral representations of type I fractional Brownian motions (Samorodnitsky and Taqqu 1994; Marinucci and Robinson 1999). Lemma 3 shows that the partial sum of a fractional process converges to fractional Brownian motion. This result is not new and has been proved by several authors including Davydov (1970, Theorem 2), Avram and Taqqu (1987, Theorem 2 with $n = 1$), Chan and Terrin (1995, Theorem 3), and Davidson and de Jong (2000). Lemma 3 also shows that the partial sum of the product process $x_t u_t$ converges to the non-Gaussian process $Z(r)$. This result was obtained by Fox and Taqqu (1987) and Chung (2002) but under the stronger assumption that both d_x and d_u are greater than 0.25 and less than 0.5. The aforementioned papers considered either the partial sums of fractional processes or that of the product process, but not both (the only exception is Chung (2002)). Lemma 3 fills in this gap by considering them jointly and develops unified representations of the limiting processes.

Using Lemma 3 and following the same steps as the proof of Theorem 1, we can establish the asymptotic distributions of $\widehat{\beta}$ and $\widehat{\sigma}_{\beta}^2$ (defined in (9) and (10)) and the t-statistic in the following theorem.

Theorem 4 *Let x_t and u_t be the time series defined by (25) and $y_t = \alpha + x_t\beta + u_t$. Assume that $k(x)$ is a twice continuously differentiable function in $\mathcal{K}$ and $M = bT$ for $b \in (0, 1]$. If $d_x, d_u \in (0, 1/2)$ and $d_x + d_u > 1/2$, then*

$$T^{1-d_x-d_u}(\widehat{\beta} - \beta - \sigma_{xu}\sigma_x^{-2}) \Rightarrow \sigma_x^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(\xi, \eta, 1) \left|\xi\right|^{-d_x} \left|\eta\right|^{-d_u} dW_x(\xi)dW_u(\eta),$$

$$T^{2-2d_x-2d_u}\widehat{\sigma}_{\beta}^2 \Rightarrow -\sigma_x^{-4} \int_0^1\int_0^1 \frac{1}{b^2}k''(\frac{r-s}{b})U(r)U(s)drds,$$

$$\widehat{t}_{\beta} - \sigma_{xu}\sigma_x^{-2}\widehat{\sigma}_{\beta}^{-1} \Rightarrow \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(\xi, \eta, 1) \left|\xi\right|^{-d_x} \left|\eta\right|^{-d_u} dW_x(\xi)dW_u(\eta)\right) \times \left(\int_0^1\int_0^1 -\frac{1}{b^2}k''(\frac{r-s}{b})U(r)U(s)drds\right)^{-1/2}, \quad (35)$$

where

$$\psi(\xi, \eta, r) = \frac{\exp(i(\xi+\eta)r) - 1}{i(\xi+\eta)} - \frac{\exp(i\xi r) - 1}{i\xi}\frac{\exp(i\eta r) - 1}{i\eta}, \quad (36)$$

$$U(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\psi(\xi, \eta, r) - r\psi(\xi, \eta, 1)) \left|\xi\right|^{-d_x} \left|\eta\right|^{-d_u} dW_x(\xi)dW_u(\eta). \quad (37)$$

The theorem shows when x_t and u_t are correlated, the OLS estimate of β will be inconsistent. To facilitate the comparison with the existing literature on spurious regression between stationary long memory processes, we assume $\sigma_{ux} = 0$ in the discussion below. It should be noted that when $d_x > d_u$, β can be consistently estimated using the frequency domain approach even if $\sigma_{ux} \neq 0$ (c.f. Robinson 1994b).

When $\sigma_{ux} = 0$, Theorem 4 reveals the convergence of the t-statistics when $M = bT$. In contrast, Tsay and Chung (2000) showed that the t-statistic based on the OLS standard error diverges at the rate of $T^{d_x+d_u-0.5}$. As a consequence, the slope coefficient in the regression between two independent stationary long memory processes can be spuriously significant. The convergence of the t-statistic under the new asymptotics has profound implications. Note that the OLS estimator $\widehat{\beta}$ is consistent, the R^2 converges to zero, and the DW statistic does not approach zero (Tsay and Chung 2000). The behaviors of $\widehat{\beta}$, R^2 and DW are thus the same as in the case of no spurious effect. The only qualitative difference is the divergence of the usual t-statistic. Therefore, under the large M asymptotics, all of the statistics behave as in the case of usual regression. Hence, inferences based on the large M asymptotics will not result in the finding of a significant relationship that does not actually exist. We may conclude that there is no spurious effect between two independent stationary long memory processes, as long as M is allowed to grow at the same rate as T and correct critical values are employed.

The above theorem assumes that the kernel function is twice continuously differentiable. This excludes the widely used Bartlett kernel and the sharp kernels studied by Phillips, Sun and Jin (2003). The sharp kernels are defined by $k(x) = (1 - |x|)^{\rho}1\{|x| \leq 1\}$, where ρ is the sharpness index. These kernels, as so defined, exhibit a sharp peak at the origin and include the Bartlett kernel as a special case. It can be shown that the sharp kernels are positive semi-definite. In the I(0) framework, Kiefer and Vogelsang (2002a,b) showed that the Bartlett kernel delivers a class of test with the highest powers within a group of popular kernels. Subsequently, Phillips, Sun and Jin (2003) showed that the sharp kernels can deliver more powerful tests than the Bartlett kernel. Thus, it is of interest to consider the sharp kernels in the present context.

The following theorem establishes the asymptotic distributions of $\widehat{\beta}$, $\widehat{\sigma}_\beta$ and the t-statistic when the sharp kernels are employed.

Theorem 5 *Let x_t and u_t be the time series defined by (25) and $y_t = \alpha + x_t\beta + u_t$. If $k(x) = (1-|x|)^\rho 1\{|x| \le 1\}$, $M = bT$ for $b \in (0,1]$, $d_x, d_u \in (0, 1/2)$ and $d_x + d_u > 1/2$, then the results of Theorem 4 hold with $-1/b^2 \int_0^1 \int_0^1 k''((r-s)/b)U(r)U(s)drds$ replaced by*

$$\frac{2\rho}{b}\int_0^1 U^2(r)dr - \frac{2\rho}{b}\int_0^{1-b} U(r)U(r+b)dr$$
$$-\frac{\rho(\rho-1)}{b^2}\int_0^1\int_0^1 U(r)\left(1-\frac{|r-s|}{b}\right)^{\rho-2} 1\{0<|r-s|<b\}\, U(s)drds \quad (38)$$

where (38) is defined to be zero when $\rho = 1$.

Tsay and Chung (2000) showed that when a stationary $I(d_u)$ process is regressed on a linear trend, the usual t-statistic diverges at the rate of T^{d_u}. We proceed to investigate whether this is the case when $M = bT$. To this end, we assume that u_t satisfies the functional central limit theorem as before:

$$T^{-(d_u+1/2)}\sum_{t=1}^{[Tr]} u_t \Rightarrow B_{d_u}(r). \quad (39)$$

Using sum by parts and the continuous mapping theorem, we have

$$T^{-(d_u+3/2)}\sum_{t=1}^{[Tr]} tu_t \Rightarrow rB_{d_u}(r) - \int_0^r B_{d_u}(s)ds. \quad (40)$$

Let

$$\begin{aligned} G(r) &= (r-\frac{1}{2})B_{d_u}(r) - \int_0^r B_{d_u}(s)ds - B_{d_u}(1)\left(\int_0^r (s-1/2)ds\right) \\ &\quad -\left(6B_{d_u}(1) - 12\int_0^1 B_{d_u}(s)ds\right)\int_0^r \left(s-\frac{1}{2}\right)^2 ds. \end{aligned} \quad (41)$$

Then we can prove the following theorem using (39) and (40) and the arguments similar to the proof of Theorem 4. Details are omitted.

Theorem 6 *Let u_t be the time series defined by (25) with $d_u \in (0, 1/2)$, x_t be the linear trend $x_t = t$, and $y_t = \alpha + \beta x_t + u_t$. If $k(x)$ is a twice continuously differentiable function in $\mathcal{K}$ and $M = bT$ for $b \in (0, 1]$, then*

$$T^{3/2-d_u}(\widehat{\beta} - \beta) \Rightarrow 6B_{d_u}(1) - 12\int_0^1 B_{d_u}(s)ds, \quad (42)$$

$$T^{3-2d_u}\widehat{\sigma}_\beta^2 \Rightarrow -\frac{144}{b^2}\int_0^1\int_0^1 k''(\frac{r-s}{b})G(r)G(s)drds, \quad (43)$$

$$\widehat{t}_\beta \Rightarrow \left(\frac{1}{2}B_{d_u}(1) - \int_0^1 B_{d_u}(s)\right)\left(\int_0^1\int_0^1 -\frac{1}{b^2}k''(\frac{r-s}{b})G(r)G(s)drds\right)^{-1/2} \quad (44)$$

If $k(x) = (1 - |x|)^\rho 1\{|x| \leq 1\}$ for some integer $\rho \geq 1$, then (42), (43), and (44) hold with $\int_0^1\int_0^1 -\frac{1}{b^2}k''(\frac{r-s}{b})G(r)G(s)drds$ replaced by

$$\frac{2\rho}{b}\int_0^1 G^2(r)dr - \frac{2\rho}{b}\int_0^{1-b} G(r)G(r+b)dr \quad (45)$$

$$-\frac{\rho(\rho-1)}{b^2}\int_0^1\int_0^1 G(r)\left(1 - \frac{|r-s|}{b}\right)^{\rho-2} 1\{0 < |r-s| \leq b\} G(s)drds \quad (46)$$

where (46) is defined to be zero when $\rho = 1$.

Theorem 6 shows that the OLS estimator is consistent and the t-statistic converges as in other cases. Therefore, detrending a stationary fractionally integrated process will not lead to the spurious effect of finding a significant trend, as long as critical values from the large M asymptotics are used.

4 Kernel Estimates of Asymptotic Distributions

The limiting distributions of $\widehat{t}_\beta$ are nonstandard. In this section, we use Monte Carlo simulations to approximate their probability densities.

For simplicity, we assume that there is no correlation between $\{x_t\}$ and $\{u_t\}$. In this case, it suffices to simulate simple fractionally integrated processes. Specifically, we generate the fractional processes x_t and u_t according to $(1 - L)^{d_x}x_t = \varepsilon_{xt}$ and $(1 - L)^{d_u}u_t = \varepsilon_{ut}$, where $\varepsilon_{xt} \sim iidN(0,1)$, $\varepsilon_{ut} \sim iidN(0,1)$ for $t > 0$, $\varepsilon_{xt} = \varepsilon_{ut} = 0$ for $t \leq 0$, and $\{\varepsilon_{xt}\}$ is independent of $\{\varepsilon_{ut}\}$. Without loss of generality, we let $\beta = 0$. For spurious regressions between nonstationary $I(d)$ processes, we consider $(d_x, d_u) = (0.6, 0.6), (0.6, 1), (1, 0.6)$, or $(1, 1)$; and for those between stationary ones, we let $(d_x, d_u) = (0.3, 0.3), (0.4, 0.2)$ or $(0.2, 0.4)$. We let $k(\cdot)$ be the Bartlett kernel and consider the following values of $b = 0, 0.025, 0.05, ..., 1$. When $b = 0$, $\widehat{\Omega}$ becomes $T^{-1}\sum \widehat{u}_t^2 (x_t - \bar{x})^2$. We use 5000 replications and a sample size of 1000. For the results reported below, 10000 replications are employed.

We first exam spurious regressions with nonstationary fractional processes. Figure 1 reports the kernel estimates of the probability densities for the case $x_t \sim I(d_x)$, $u_t \sim I(d_u)$ with $d_x = d_u = 0.6$. The qualitative results for other (d_x, d_u) combinations are similar. The probability densities appear to be symmetric and are apparently more dispersed than the standard normal density. For example, when $b = 0.1$, the 95% quantile of the limiting distribution is 3.1012, which is larger than 1.645, the 95% quantile of the standard normal distribution. The quantile is not a monotonic function of b. When b is small, the quantile is decreasing in b whereas when b is large, the quantile is increasing in b. In other words, when b moves closer to zero or one, the limiting distribution becomes more dispersed. This result is well expected when b is very close to zero. In this case, the behavior of the t-statistic may be better described by the conventional limit theory. However, when b is very close to one, one may think that the limiting distribution should becomes less dispersed as the variance estimate with a large bandwidth captures the strong autocorrelation in $x_t u_t$. This is not what is happening. In fact, as b becomes close to one, the variance estimate $\widehat{\Omega}$ does not capture the strong autocorrelation very well. This is because u_t is not observed and has to be replaced by the estimated residual $\widehat{u}_t$. As a consequence, when $k(x) = 1$ and $b = 1$, $\widehat{\Omega}_1 = \sum_{j=-T+1}^{T-1} \widehat{\Gamma}(j)$ is zero by construction. Of course, this is an extreme case. But for other kernels, when b is close to one, the effect of increasing b is similar. The estimate $\widehat{\Omega}$ tends to bias toward zero when $\widehat{\Omega}$ incorporates autocovariances of almost all lags.

For all the cases considered, the simulation results indicate that the limiting distribution is least dispersed or close to be least dispersed when $b = 0.1$. When $d_x = d_u = 0.6$ and $b = 0.1$, the probability of $\left|\widehat{t}_\beta\right| > 1.96$ is 28.39%. Therefore, when $M = 0.1T$, we will erroneously reject the null 28.39% of the times when the wrong critical value is used. In contrast, when $M = o(T)$, the rejection probability goes to one as the sample size increases. When $T = 1000$ and the OLS standard error is used, the rejection probability is 75.9%, as shown by simulations. Hence, the use of $M = 0.1T$ reduces the spurious effect substantially.

Figure 2 presents the same graph when y_t is an $I(0.6)$ process and x_t is a linear deterministic trend. The graph is representative of other cases. The qualitative observations made for Figure 1 apply. However, the limiting distributions become more dispersed than those in Figure 1.

We next consider spurious regressions with stationary fractional processes. Figure 3 graphs the density estimates when $(d_x, d_u) = (0.4, 0.2)$. The density estimates for the other two cases turn out to be close to the case $(d_x, d_u) = (0.4, 0.2)$. The figure shows that the limiting distributions are more concentrated around the origin than in the nonstationary cases. For example, the 90% quantiles when $b = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ are $1.9265, 1.5602, 1.6963, 1.8440, 2.0120$, and 2.1654, respectively. The corresponding 95% quantiles are 2.4695, 2.0525, 2.2118, 2.4741, 2.6988, and 2.9385, respectively. Simulation results show that the limiting distribution becomes close to the standard normal when b is close to 0.1. For example, when $b = 0.1$, the probability of $\left|\widehat{t}_\beta\right| > 1.96$ is 11.70%, which is very close to 10%, the size of the test when $\widehat{t}_\beta$ is standard normal. In other words, when $M = 0.1T$ is used to test the null of $\beta = 0$, the probability of wrong rejection is only 11.70% even if the critical value does not come from the true limiting distribution. To a great extent, the new t-test eliminates the spurious effect.

Figure 4 graphs the density estimates when $y_t \sim I(0.3)$ and $x_t = t$. Again, we find that the densities are more concentrated than in the nonstationary cases and become most concentrated when b is close to 0.1. Another feature of Figures 3 and 4 is that the densities appear to be slightly negatively skewed (skewed to the left).

5 Conclusion

This paper proposes a t-statistic that is convergent in all the cases of spurious regressions considered in the literature. This t-statistic is based on the HAC estimator using a truncation lag or bandwidth proportional to the sample size. It was shown that the usual t-statistic diverges because the OLS standard error does not take into account the high persistence of the regressor and regression residuals. This finding helps to explain why the significant regression coefficients occur in nonsense regressions. Conventional wisdom is that the variables in a nonsense regression share the common feature of a trending mechanism. But there is no trending mechanism in a stationary long memory process, yet the usual t-statistic diverges, and, therefore, ultimately exceeds any finite critical values and indicates a statistically significant relationship between two independent long memory processes. Our study suggests that it is the use of a small bandwidth ($M = o(T)$) in estimating the standard error that gives arise the spurious effects. If a large bandwidth ($M = bT$) is used, the convergence of the t-statistic can be restored and valid inferences can be made.

When the bandwidth is proportional to the sample size, the t-statistic also converges to a well-defined distribution in the usual regressions with weakly dependent processes (Kiefer and Vogelsang (2003)). Therefore, the specification of $M = bT$ has the potential to deliver a unified inferential framework. The advantage of setting $M = bT$ is that the t-statistic converges in distribution without any normalization. In contrast, when M is chosen to grow at a slower rat than T, the t-statistic may have to be normalized to be convergent. The normalization factor is typically of the form T^{κ} for some parameter κ that depends on unknown memory parameters. In practice, researchers routinely use the t-statistic without any normalization. This practice thus runs the risk of over rejection when M is small and the underlying processes exhibit some persistence. The use of $M = bT$ has the potential to reduce this risk.

The paper is a first step towards the large M asymptotics for highly persistent, possibly nonstationary time series. It can be extended in several directions. First, the results of the paper are readily extended to the multiple regression with two or more regressors. Second, the limiting distribution depends on the memory parameters and the long run covariance matrix. In empirical applications, we may consistently estimate these quantities and simulate the limiting distribution. Another possible solution is to investigate the validity of subsampling in this context (Politis, Romano and Wolf (1999)). Finally, Phillips (1998) argued that there is a valid mathematical representation underlying the spurious regression between I(1) processes. His argument relied on the statistical significance of the coefficients manifested in the usual t-statistic or the t-statistic based on a small bandwidth ($1/M + M/T \to 0$). However, when M is allowed to grow at the same rate as T, the t-statistic is expected to converge and statistical significance does not seem to be inevitable. It is an open question how to interpret the mathematical representation under the specification of $M = bT$.

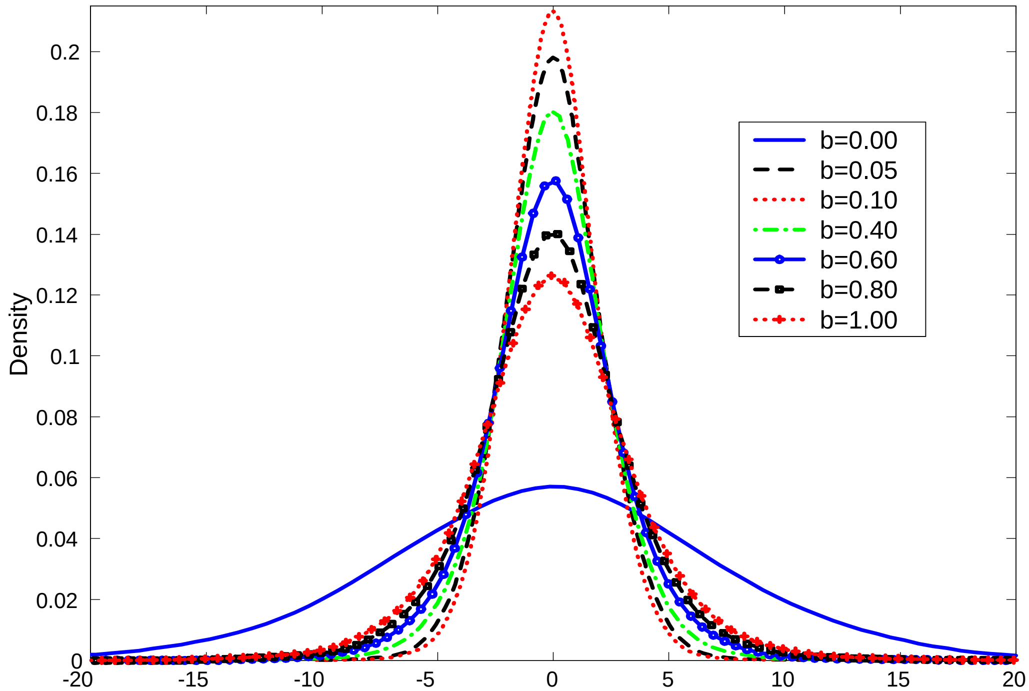

Figure 1. kernel estimates of densities of $\widehat{t}_{\beta}$ when $x_t \sim I(0.6)$ and $y_t \sim I(0.6)$

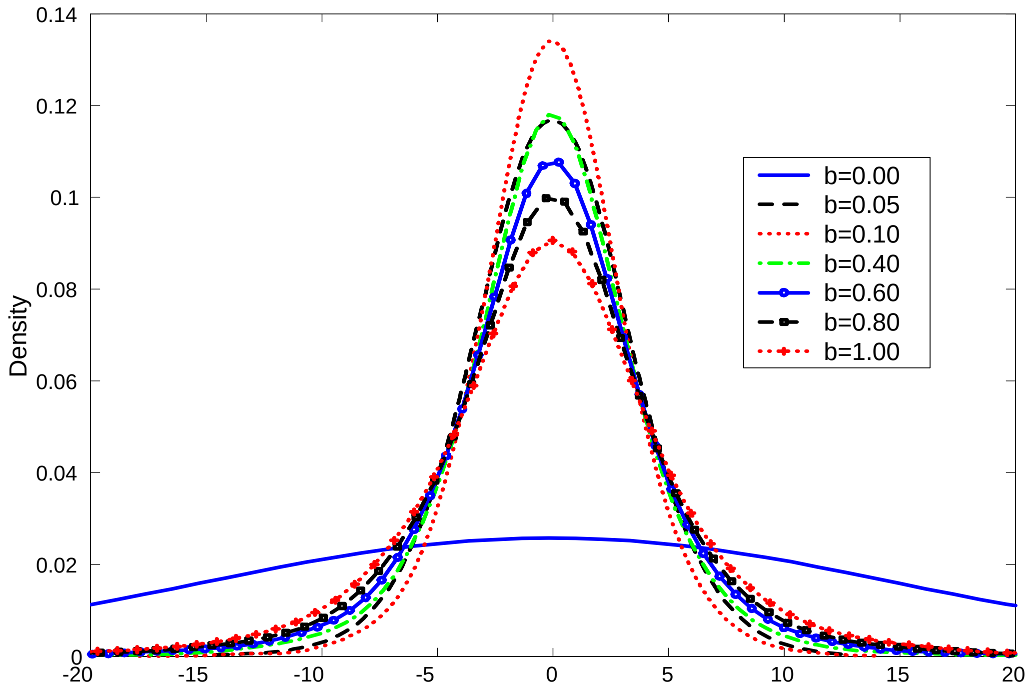

Figure 2. kernel estimates of densities of $\widehat{t}_{\beta}$ when $x_t = t$ and $y_t \sim I(0.6)$

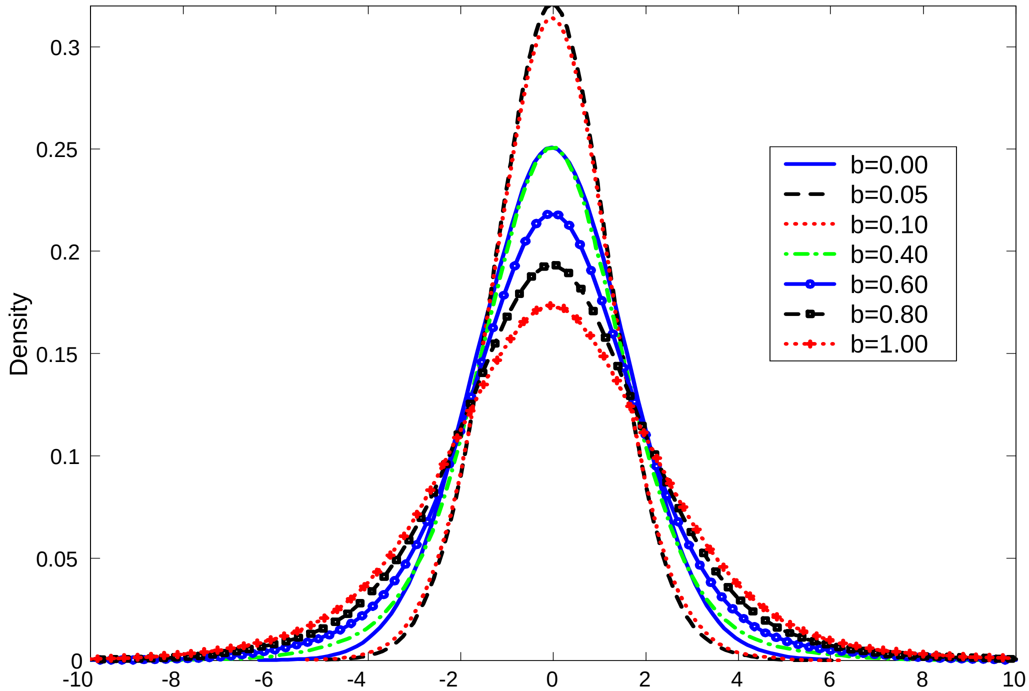

Figure 3. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t \sim I(0.4)$ and $y_t \sim I(0.2)$

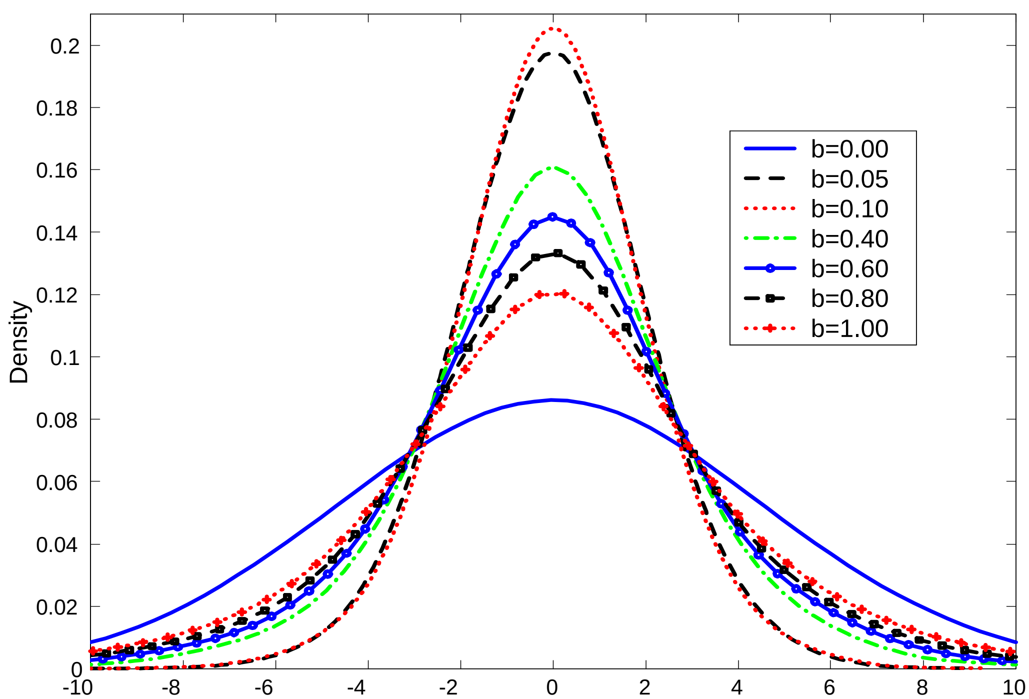

Figure 4. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t = t$ and $y_t \sim I(0.3)$

6 Appendix of Proofs

Proof of Theorem 1. Combine the functional central limit theorem with the continuous mapping theorem, we have

$$T^{-d_x-d_u}\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u})\Rightarrow\int_0^1\widetilde{V}_x(r)\widetilde{V}_u(r)dr \tag{47}$$

and

$$T^{-2d_x}\sum_{t=1}^{T}(x_t-\bar{x})^2\Rightarrow\int_0^1\widetilde{V}_x^2(r)dr. \tag{48}$$

Hence

$$\begin{aligned}T^{d_x-d_u}(\widehat{\beta}-\beta) &= \frac{T^{-d_x-d_u}\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u})}{T^{-2d_x}\sum_{t=1}^{T}(x_t-\bar{x})^2}\\ &\Rightarrow \left(\int_0^1\widetilde{V}_x(r)\widetilde{V}_u(r)dr\right)\left(\int_0^1\widetilde{V}_x^2(r)dr\right)^{-1}.\end{aligned} \tag{49}$$

As a consequence,

$$\begin{aligned}&T^{-(2d_u-1)/2}\widehat{u}_{[Tr]}\\ =\;&T^{-(2d_u-1)/2}\left(u_{[Tr]}-\bar{u}\right)-\left(T^{d_x-d_u}(\widehat{\beta}-\beta)\right)\left(T^{-(2d_x-1)/2}(x_{[Tr]}-\bar{x})\right)\Rightarrow\widetilde{V}_{u,x}(r).\end{aligned}$$

Now write $T^{2d_x-2d_u}\widehat{\sigma}^2_{\beta}$ as

$$\begin{aligned}&\left(\frac{1}{T}\sum_{t=1}^{T}(\frac{x_t-\bar{x}}{T^{d_x-1/2}})^2\right)^{-2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\frac{x_t-\bar{x}}{T^{d_x-1/2}}\frac{\widehat{u}_t}{T^{d_u-1/2}}k(\frac{t-\tau}{bT})\frac{\widehat{u}_\tau}{T^{d_u-1/2}}\frac{x_\tau-\bar{x}}{T^{d_x-1/2}}\\ \Rightarrow\;&\left(\int_0^1\widetilde{V}_x^2(r)dr\right)^{-2}\int_0^1\int_0^1\widetilde{V}_x(r)\widetilde{V}_{u,x}(r)k(\frac{r-s}{b})\widetilde{V}_{u,x}(s)\widetilde{V}_x(s)drds,\end{aligned} \tag{50}$$

where the last line follows from the continuous mapping theorem. In view of (49) and (50), we have

$$\begin{aligned}\widehat{t}_\beta &= \left(\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u})\right)\left(\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\widehat{u}_tk(\frac{t-\tau}{bT})\widehat{u}_\tau(x_\tau-\bar{x})\right)^{-1/2}\\ &\Rightarrow \left(\int_0^1\widetilde{V}_x(r)\widetilde{V}_u(r)dr\right)\left(\int_0^1\int_0^1\widetilde{V}_x(r)\widetilde{V}_{u,x}(r)k(\frac{r-s}{b})\widetilde{V}_{u,x}(s)\widetilde{V}_x(s)drds\right)^{-1/2}.\end{aligned} \tag{51}$$

This completes the proof of the theorem. ■

Proof of Lemma 3. We first prove the tightness of $S_T(r)$. From Lemma A.3 of Phillips and Durlauf (1986), we know that the necessary and sufficient condition for the tightness of $S_T(r)$ is that each element of $S_T(r)$ is tight in the respective component space. But several authors (Davydov (1970, Theorem 2), Avram and Taqqu (1987, Theorem 2), Davidson and de Jong (1998)) have proved that the partial sum processes $S_T^x(r)$ and $S_T^u(r)$ converge weakly to fractional Brownian motions. It follows from a theorem of Prohorov (Billingsley (1999), Theorem 5.1, p. 59) that, since $D[0,1]^3$ is complete and separable, both $\{S_T^x(r)\}$ and $\{S_T^u(r)\}$ are tight. It remains to show the tightness of $\{S_T^{xu}(r)\}$. In view of Theorem 13.5 of Billingsley (1999), it suffices to show that, for almost all sample paths, some constant $C > 0$ and $0 \le r_1 \le r \le r_2 \le 1$,

$$P\left(|S_T^{xu}(r) - S_T^{xu}(r_1)| \ge \lambda, |S_T^{xu}(r_2) - S_T^{xu}(r)| \ge \lambda\right) \le C\lambda^{-2}(r_2 - r_1)^{2\nu}, \tag{52}$$

where $\lambda > 0$ and $\nu > 1/2$. By the Cauchy and Markov inequalities, we have

$$\begin{aligned}
& P\left(|S_T^{xu}(r) - S_T^{xu}(r_1)| \ge \lambda, |S_T^{xu}(r_2) - S_T^{xu}(r)| \ge \lambda\right) \\
\le\; & \left[P\left(|S_T^{xu}(r) - S_T^{xu}(r_1)| \ge \lambda\right)\right]^{1/2} \left[P\left(|S_T^{xu}(r_2) - S_T^{xu}(r)| \ge \lambda\right)\right]^{1/2} \\
\le\; & \frac{\left(E\left(S_T^{xu}(r) - S_T^{xu}(r_1)\right)^2\right)^{1/2}}{\lambda} \frac{\left(E\left(S_T^{xu}(r) - S_T^{xu}(r_1)\right)^2\right)^{1/2}}{\lambda}.
\end{aligned} \tag{53}$$

Note that, for a generic constant C that may be different across lines,

$$\begin{aligned}
& E\left(S_T^{xu}(r) - S_T^{xu}(r_1)\right)^2 \\
=\; & T^{-2d_x-2d_u} E\left(\sum_{t=[Tr_1]+1}^{[Tr]} (x_t u_t - \sigma_{xu})\right)^2 \\
=\; & T^{-2d_x-2d_u} E\left(\sum_{t=[Tr_1]+1}^{[Tr]} x_t u_t\right)^2 - T^{-2d_x-2d_u}\sigma_{xu}^2\left([Tr] - [Tr_1]\right)^2 \\
=\; & T^{-2d_x-2d_u} \sum_{t=[Tr_1]+1}^{[Tr]} \sum_{\tau=[Tr_1]+1}^{[Tr]} E x_t u_t x_\tau u_\tau - T^{-2d_x-2d_u}\sigma_{xu}^2\left([Tr] - [Tr_1]\right)^2.
\end{aligned}$$

Since x_t and u_t are Gaussian processes, we have

$$E x_t u_t x_\tau u_\tau = E x_t u_t E x_\tau u_\tau + E x_t x_\tau E u_t u_\tau + E x_t u_\tau E u_t x_\tau. \tag{54}$$

Using the above formula, we can rewrite $E\left(S_T^{xu}(r) - S_T^{xu}(r_1)\right)^2$ as $I_1 + I_2$ where

$$\begin{aligned}
I_1 &= T^{-2d_x-2d_u} \sum_{t=[Tr_1]+1}^{[Tr]} \sum_{\tau=[Tr_1]+1}^{[Tr]} E x_t x_\tau E u_t u_\tau, \\
I_2 &= T^{-2d_x-2d_u} \sum_{t=[Tr_1]+1}^{[Tr]} \sum_{\tau=[Tr_1]+1}^{[Tr]} E x_t u_\tau E u_t x_\tau.
\end{aligned} \tag{55}$$

We now consider I_1 and I_2. First,

$$\begin{aligned}
I_1 &= T^{-2d_x-2d_u}2\sum_{t=[Tr_1]+1}^{[Tr]}\sum_{\tau=t+1}^{[Tr]}(Ex_tx_\tau)(Eu_tu_\tau)+T^{-2d_x-2d_u}\sum_{t=[Tr_1]+1}^{[Tr]}\left(Ex_t^2\right)\left(Eu_t^2\right) \\
&= CT^{-2d_x-2d_u}\sum_{t=[Tr_1]+1}^{[Tr]}\sum_{\tau=t+1}^{[Tr]}(\tau-t)^{2d_x+2d_u-2}+CT^{1-2d_x-2d_u}(\frac{[Tr]-[Tr_1]}{T}) \\
&\leq C\int_{r_1}^{r}\left(\int_t^r(\tau-t)^{2d_x+2d_u-2}d\tau\right)dt+CT^{1-2d_x-2d_u}(\frac{[Tr]-[Tr_1]}{T}) \\
&= \frac{2C}{(2d_x+2d_u-1)(2d_x+2d_u)}(r-r_1)^{2d_x+2d_u}+o(1) \\
&\leq C(r-r_1)^{2d_x+2d_u}, \qquad (56)
\end{aligned}$$

where we use the fact that $Ex_tx_\tau \leq C|\tau-t|^{2d_x-1}$ and $Eu_tu_\tau \leq C|\tau-t|^{2d_u-1}$. Similarly, using $Ex_tu_\tau \leq C|\tau-t|^{d_x+d_u-1}$, we can show that $I_2 \leq C(r-r_1)^{2d_x+2d_u}$. Therefore,

$$E\left(S_T^{xu}(r)-S_T^{xu}(r_1)\right)^2 \leq C(r-r_1)^{2d_x+2d_u}. \qquad (57)$$

Combining (53) with (57) yields

$$\begin{aligned}
&P\left(|S_T^{xu}(r)-S_T^{xu}(r_1)| \geq \lambda, |S_T^{xu}(r_2)-S_T^{xu}(r)| \geq \lambda\right) \\
\leq\ & C\lambda^{-2}(r-r_1)^{d_x+d_u}(r_2-r)^{d_x+d_u} \leq C\lambda^{-2}(r_2-r_1)^{2d_x+2d_u}. \qquad (58)
\end{aligned}$$

Therefore (52) holds and $\{S_T^{xu}(r)\}$ is tight.

It remains to prove the finite dimensional (fidi) convergence of $S_T(r)$. From Theorem 3.3 of Chan and Terrin (1995), we know that the fidi distribution of $(S_T^x(r), S_T^u(r))$ converges to that of $(B_{d_x}(r), B_{d_u}(r))$. The fidi convergence of $S_T^{xu}(r)$ follows from Theorem 7.4 of Giraitis and Taqqu (1999, page 30). We show below that this theorem implies that, if x_t and u_t follow linear processes with the same iid innovation sequences, then

$$T^{-d_x-d_u}\sum_{t=1}^{[Tr]}(x_tu_t-\sigma_{xu}) \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(i(\xi+\eta)r)-1}{i(\xi+\eta)}|\xi|^{-d_x}|\eta|^{-d_u}dW_x(\xi)dW_u(\eta) \qquad (59)$$

for any $r \in (0,1]$. Our case differs from that of Giraitis and Taqqu (1999) only in that we assume that x_t and u_t have correlated innovation sequences where they assume that x_t and u_t share the same innovation sequences. Nevertheless, their proof goes through for our case with obvious and minor modifications.

To show (59), we use the notation in Giraitis and Taqqu (1999), and set $i = 1, m_i = n_i = 1, l = 2, \alpha^{(i,1)} = 2d_x, \alpha^{(i,2)} = 2d_u, N = T$, $b(\tau) = 1\{\tau = 0\}$. For these special parameter and function specifications, the partial sum process considered by Giraitis and Taqqu (1999), viz,

$$Q_{[Tr]} = \sum_{t=1}^{[Tr]}\sum_{s=1}^{[Tr]}b(t-s)P_{m_i,n_i}(x_t,u_s), \qquad (60)$$

becomes

$$Q_{[Tr]} = \sum_{t=1}^{[Tr]} (x_t u_t - \sigma_{xu}) . \tag{61}$$

Therefore, (59) is a special case of Theorem 7.4 of Giraitis and Taqqu (1999).

Finally, the joint fidi convergence of $(S_T^x(r), S_T^u(r))$ with $S_T^{xu}(r)$ follows from the fact that they are defined as stochastic integrals of deterministic functions with respect to the same Gaussian measures $W_x(\cdot)$ and $W_u(\cdot)$. ■

Proof of Theorem 4. From the definition of $\{x_t\}$, we have

$$x_t^2 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp(it(\lambda - \mu)) \left|\lambda\right|^{-d_x} \left|\mu\right|^{-d_x} dW_x(\lambda)\overline{dW}_x(\mu). \tag{62}$$

So

$$T^{-1} \sum_{t=1}^{T} x_t^2 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} T^{-1} \sum_{t=1}^{T} \exp(it(\lambda - \mu)) \left|\lambda\right|^{-d_x} \left|\mu\right|^{-d_x} dW_x(\lambda)\overline{dW}_x(\mu). \tag{63}$$

Combining (63) with dominated convergence and the fact that

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \exp(it(\lambda - \mu)) = 1_{\{\lambda=\mu\}}, \tag{64}$$

we have

$$\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} x_t^2 = \int_{-\pi}^{\pi} f_x(\lambda) d\lambda = \sigma_x^2. \tag{65}$$

Similarly, $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} x_t = 0$. Therefore

$$\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} x_t^2 - (\bar{x})^2 = \int_{-\pi}^{\pi} f_x(\lambda) d\lambda = \sigma_x^2. \tag{66}$$

Combining (66) with Lemma 3 yields

$$\begin{aligned} & T^{1-d_x-d_u}(\widehat{\beta} - \beta - \sigma_{xu}/\sigma_x^2) \\ = {} & T^{-d_x-d_u} \left(\sum_{t=1}^{T} (x_t u_t - \sigma_{xu} - T\bar{x}\bar{u})\right) \left(T^{-1} \sum_{t=1}^{T} x_t^2 - (\bar{x})^2\right)^{-1} + o_p(1) \\ \Rightarrow {} & \sigma_x^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(\xi, \eta, 1) \left|\xi\right|^{-d_x} \left|\eta\right|^{-d_u} dW_x(\xi) dW_u(\eta). \end{aligned} \tag{67}$$

We next consider the limiting distribution of

$$\sum_{t=1}^{T} \sum_{\tau=1}^{T} (x_t - \bar{x}) \widehat{u}_t k(\frac{t-\tau}{bT}) \widehat{u}_\tau (x_\tau - \bar{x}) . \tag{68}$$

Let $v_t = (x_t - \bar{x})\,\widehat{u}_t$ and $S_T^v(r) = \sum_{t=1}^{[Tr]} v_t$, for $r \geq 1/T$ and $S_T^v(r) = 0$, for $0 \leq r < 1/T$. Then

$$\begin{aligned}
T^{-d_x-d_u} S_T^v(r) &= T^{-d_x-d_u} \sum_{t=1}^{[Tr]} (x_t - \bar{x})\,(y_t - \bar{y}) - T^{-d_x-d_u}\widehat{\beta} \sum_{t=1}^{[Tr]} (x_t - \bar{x})^2 \\
&= T^{-d_x-d_u} \sum_{t=1}^{[Tr]} (x_t - \bar{x})\,(u_t - \bar{u}) - T^{-d_x-d_u}(\widehat{\beta} - \beta) \sum_{t=1}^{[Tr]} (x_t - \bar{x})^2 \\
&= T^{-d_x-d_u} \sum_{t=1}^{[Tr]} (x_t - \bar{x})\,(u_t - \bar{u}) - T^{-d_x-d_u} r \sum_{t=1}^{T} (x_t u_t - T\bar{x}\bar{u})(1 + o_p(1)) \\
&\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\psi(\xi,\eta,r) - r\psi(\xi,\eta,1))\,|\xi|^{-d_x}\,|\eta|^{-d_u}\,dW_x(\xi)dW_u(\eta) \qquad (69) \\
&: = U(r),
\end{aligned}$$

where we have used

$$\text{plim}_{T\to\infty}\left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)^{-1}\sum_{t=1}^{[Tr]}(x_t - \bar{x})^2 = r. \qquad (70)$$

Using summation by parts twice, we have

$$\begin{aligned}
& T^{-2d_x-2d_u} \sum_{t=1}^{T}\sum_{\tau=1}^{T} (x_t - \bar{x})\,\widehat{u}_t k(\frac{t-\tau}{bT})\widehat{u}_\tau\,(x_\tau - \bar{x}) \\
= \;& T^{-2d_x-2d_u} \sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} \left\{ S_T^v(t/T)\left(2k(\frac{t-\tau}{bT}) - k(\frac{t-\tau-1}{bT}) - k(\frac{t-\tau+1}{bT})\right)\right. \\
& \left.\times S_T^v(\tau/T)\right\} + T^{-2d_x-2d_u} S_T^v(1) \sum_{\tau=1}^{T-1}\left(k(\frac{T-\tau}{bT}) - k(\frac{T-\tau-1}{bT})\right) S_T^v(\tau/T) \\
& + T^{-2d_x-2d_u}\left\{\sum_{t=1}^{T-1} S_T^v(t/T)\left(k(\frac{t-T}{bT}) - k(\frac{t-T+1}{bT})\right) S_T^v(1) + S_T^v(1)S_T^v(1)\right\} \\
= \;& \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} T^{-d_x-d_u} S_T^v(t/T) T^2 D_T(\frac{t-\tau}{bT}) T^{-d_x-d_u} S_T^v(\tau/T), \qquad (71)
\end{aligned}$$

where

$$D_T(\frac{t-\tau}{bT}) = 2k(\frac{t-\tau}{bT}) - k(\frac{t-\tau-1}{bT}) - k(\frac{t-\tau+1}{bT}) \qquad (72)$$

and the last line follows from the identity that $S_T^v(1) = 0$. Note that when $T \to \infty$ such that $(t-\tau)/T \to r - s$, we have

$$T^2 D_T(\frac{t-\tau}{bT}) \to -\frac{1}{b^2}k''(\frac{r-s}{b}). \qquad (73)$$

Combining (69), (71), and (73), and invoking the continuous mapping theorem, we get

$$\begin{aligned} & T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{bT})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\ \Rightarrow & -\frac{1}{b^2}\int_0^1\int_0^1 U(r)k''(\frac{r-s}{b})U(s)drds. \end{aligned} \tag{74}$$

Therefore

$$\begin{aligned} & T^{2-2d_x-2d_u}\widehat{\sigma}_\beta^2 \\ = & \left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^{-2} T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{bT})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\ \Rightarrow & -\sigma_x^{-4}\frac{1}{b^2}\int_0^1\int_0^1 U(r)k''(\frac{r-s}{b})U(s)drds. \end{aligned} \tag{75}$$

Combining (67) with (75) yields the limiting distribution of the t-statistic. This completes the proof of the theorem. ■

Proof of Theorem 5. For the Bartlett kernel, we have, after simple calculations,

$$D_T(\frac{t-\tau}{bT}) = \begin{cases} \frac{2}{bT}, & \text{if } |t-\tau| = 0 \\ -\frac{1}{bT}, & \text{if } |t-\tau| = [bT] \\ 0, & \text{otherwise} \end{cases} \tag{76}$$

Hence

$$\begin{aligned} & T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{bT})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\ = & T^{-2d_x-2d_u}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}S_T^v(t/T)D_T(\frac{t-\tau}{bT})S_T^v(\tau/T), \\ = & \frac{2}{bT}\sum_{\tau=1}^{T-1}T^{-2d_x-2d_u}S_T^v(t/T)S_T^v(t/T) \\ & -\frac{2}{bT}\sum_{\tau=1}^{T-[bT]-1}T^{-2d_x-2d_u}S_T^v(\tau/T)S_T^v((\tau+[bT])/T) \\ \Rightarrow & \frac{2}{b}\int_0^1 U^2(r)dr - \frac{2}{b}\int_0^{1-b}U(r)U(r+b)dr. \end{aligned} \tag{77}$$

Next, consider other sharp kernels $k(x) = (1-|x|)^\rho\{|x| \leq 1\}$ for $\rho \geq 2$. Let $\lim_{T\to\infty}(t-\tau)/T = r-s$. When $|t-\tau| \in (0,[bT])$, we have

$$\lim_{T\to\infty} b^2T^2D_T(\frac{t-\tau}{bT}) = -\rho(\rho-1)\left(1-\frac{|r-s|}{b}\right)^{\rho-2}. \tag{78}$$

When $|t-\tau| = 0$, we have

$$\lim_{T\to\infty} (bT)\, D_T\left(\frac{t-\tau}{bT}\right) = 2\rho. \tag{79}$$

When $|t-\tau| = [bT]$, we have

$$\lim_{T\to\infty} (bT)\, D_T\left(\frac{t-\tau}{bT}\right) = -\rho. \tag{80}$$

Therefore,

$$\begin{aligned}
& T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{bT})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\
= \quad & \frac{1}{b^2T^2}\sum_{t=1}^{T-1}\sum_{\tau=1,\tau\neq t,|\tau-t|<[bT]}^{T-1} T^{-2d_x-2d_u} S_T^v(t/T) b^2T^2 D_T(\frac{t-\tau}{bT}) S_T^v(\tau/T) \\
& +\frac{1}{bT}\sum_{t=1}^{T-1} T^{-2d_x-2d_u} S_T^v(t/T)\,(bT)\,D_T\,(0)\,S_T^v(t/T) \\
& -\frac{2}{bT}\sum_{t=1}^{T-[bT]-1} T^{-2d_x-2d_u} S_T^v(t/T)\,(bT)\,D_T\left(\frac{[bT]}{bT}\right) S_T^v((t+[bT])\,/T) \\
\Rightarrow \quad & \frac{2\rho}{b}\int_0^1 U^2(r)dr - \frac{2\rho}{b}\int_0^{1-b} U(r)U(r+b)dr \\
& -\frac{\rho(\rho-1)}{b^2}\iint_{[0,1]^2} U(r)\left(1-\frac{|r-s|}{b}\right)^{\rho-2} 1\left\{0<|r-s|<b\right\} U(s)drds.
\end{aligned} \tag{81}$$

The theorem now follows from (77), (81) and the steps in the proof of Theorem 4. ■

References

Akonom, J. and C. Gourieroux (1987) A functional central limit theorem for fractional processes. Discussion paper #8801, CEPREMAP, Paris.

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-854.

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

Billingsley, P. (1999) *Convergence of Probability Measures.* 2nd Edition, New York: John Wiley & Sons.

Chan, N. H. and N. Terrin (1995) Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Annals of Statistics* 23(5), 1662-1683.

Chung, C.-F. (2002) Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Theory* 18, 51-78.

Davidson, J. and R. de Jong (2000) The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes. *Econometric Theory* 16, 643-666.

Davydov, Y. A. (1970) The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15, 487– 489.

Durlauf, S. N. and P. C. B. Phillips (1988) Trends versus random walks in time series analysis. *Econometrica* 56, 1333–1354.

Fox, R. and M. S. Taqqu (1987) Multiple stochastic integrals with dependent integrators. *Journal of Multivariate Analysis* 21, 105-127.

Giraitis, L. and M. S. Taqqu (1999) Convergence of normalized quadratic forms. *Journal of Statistical Planning and Inference* 80, 15–35.

Granger, C. W. J. and P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.

Hurvich C. M. and B. K. Ray (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16: 17–42.

Jansson, M. (2002) Autocorrelation robust tests with good size and power. Department of Economics, University of California, Berkeley.

Kiefer, N. M. and T. J. Vogelsang (2002a) Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.

Kiefer, N. M. and T. J. Vogelsang (2002b) Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70(5), 2093-2095.

Kiefer, N. M. and T. J. Vogelsang (2003) A new asymptotic theory for heteroskedasiticy-autocorrelation robust tests, Working Paper, Center for Analytic Economics, Cornell University.

Marinucci D. and P. M. Robinson (1999) Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* 80, 111-122

Marinucci D. and P. M. Robinson (2000) Weak convergence to fractional Brownian motion. *Stochastic Processes and their Applications* 86, 103-120.

Marmol, F. (1995) Spurious regressions between I(d) processes. *Journal of Time Series Analysis* 16, 313–321.

Marmol, F. (1998) Spurious regression theory with nonstationary fractionally integrated processes. *Journal of Econometrics* 84, 233–250.

Phillips, P. C. B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

Phillips, P. C. B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299-1325.

Phillips, P. C. B. (1999) Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion Paper #1243, Yale University.

Phillips, P. C. B. and S. N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495

Phillips, P. C. B., Y. Sun and S. Jin (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with No truncation. Cowles Foundation Discussion Paper No. 1407
(http://cowles.econ.yale.edu/P/au/d_phillips.htm).

Politis, D.N., J. P. Romano and M. Wolf, *Subsampling,* Springer-Verlag, New York, 1999.

Robinson P. M. (1994a) Time series with strong dependence. In *Advances in Econometrics,* Sixth World Congress, Sims CA (eds), 1: 47-96. Cambridge: Cambridge University Press.

Robinson, P. M. (1994b) Semiparametric analysis of long memory time series, *Annals of Statistics*, 22, 515-539

Samorodnitsky, G. and M. S. Taqqu (1994) *Stable Non-Gaussian Random Processes.* New York: Chapman and Hall.

Sun, Y. (2002) Estimation of long run average relationships in nonstationary panel time series. Department of Economics, University of California, San Diego.

Tsay, W.-J. and C.-F. Chung (2000) The spurious regression of fractionally integrated processes. *Journal of Econometrics* 96(1), 155–182.

Velasco C. (1999). Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325-371.