Title

A Convergent t-statistic in Spurious Regressions

Author

Sun, Yixiao

Publication Date

2003

A Convergent t-statistic in Spurious Regressions*

Yixiao Sun
Department of Economics
University of California, San Diego

December 23, 2002

*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 new t-statistic is based on the heteroscedasiticity and autocorrelation consistent (HAC) standard error estimate with the bandwidth equal to the sample size. Using autocovariances of all lags, the so-defined HAC estimator is capable of capturing the high persistence of the regressor and regression residuals. It is shown that the new 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 t-statistic and 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 a new estimator of the standard error and use it to construct a new t-statistic. We show that the new t-statistic converges in distribution to a non-degenerate random variable.

The new estimator of the standard error is based on the heteroscedasiticity and autocorrelation consistent (HAC) variance estimator that uses the full bandwidth (the bandwidth or the truncation lag is equal to the sample size). 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 be as large as the sample size to capture the high autocorrelation. In other words, we use auto-covariances of all lags and construct the HAC estimator without truncation.

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 t-statistic and critical values obtained via simulations will not lead to the finding of a spurious relationship.

The HAC estimator with the bandwidth equal to the sample size has been suggested by Kiefer and Vogelsang (2002a, 2002b) in other settings. Specifically, they considered this type of estimator in hypothesis testing in the presence of nonparametric autocorrelation. 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 (2002) 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 of the new t-statistics. 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 which renders the space complete and separable.

2 Spurious Regressions with Nonstationary Fractional Processes

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

Let x_t and y_t be two independent nonstationary $I(d)$ processes with $d > 1/2$. We assume that the following functional central limit theorem (FCLT) holds:

$$T^{-(2d_x-1)/2}x_{[Tr]} \Rightarrow \omega_x V_x(r),\ T^{-(2d_y-1)/2}y_{[Tr]} \Rightarrow \omega_y V_y(r),\ r \in (0,1], \tag{1}$$

where

$$V_x(r) = \frac{1}{\Gamma(d)}\int_0^r (r-s)^{d_x-1}dW^x(s), V_y(r) = \frac{1}{\Gamma(d)}\int_0^r (r-s)^{d_y-1}dW^y(s), \tag{2}$$

$W^x(s)$ and $W^y(s)$ are standard Brownian motions. The FCLT holds under a wide range of primitive conditions (e.g. Akonom and Gourieroux 1987; Marinucci and Robinson 2000). When x_t or y_t is a unit root process, the limiting process 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).

Consider regressing y_t on a constant and x_t,

$$y_t = \widehat{\alpha} + \widehat{\beta} x_t + \widehat{u}_t, t = 1, ..., T. \tag{3}$$

The ordinary least squares estimate of β is given by

$$\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}, \tag{4}$$

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 $\widehat{t}_\beta = \widehat{\beta}/\widehat{\sigma}_{\beta,M}$, where $\widehat{\sigma}_{\beta,M}$ is the HAC estimator defined as

$$\widehat{\sigma}^2_{\beta,M} = \left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)^{-1} T\widehat{\Omega}_M \left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)^{-1}, \tag{5}$$

where

$$\widehat{\Omega}_M = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\widehat{\Gamma}(j), \tag{6}$$

$$\widehat{\Gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} (x_{t+j} - \bar{x})\, \widehat{u}_{t+j}\widehat{u}_t\, (x_t - \bar{x}) & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} (x_{t+j} - \bar{x})\, \widehat{u}_{t+j}\widehat{u}_t\, (x_t - \bar{x}) & \text{for } j < 0 \end{cases} \tag{7}$$

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

The usual approach is to let $M \to \infty$ such that $M/T \to 0$ to get a consistent estimate of the long run variance of $(x_t - \bar{x})\,\widehat{u}_t$. 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. Therefore, we let $M = T$ throughout the paper and use the full bandwidth to estimate the long run variance.

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

$$\mathcal{K} = \{k(\cdot) : [-1, 1] \to [0, 1] \mid k(x) = k(-x),\ k(0) = 1,\ \text{and } K(\lambda) \geq 0, \forall \lambda \in \mathbb{R}\}, \tag{8}$$

where

$$K(\lambda) = \int_{-1}^{1} k(x) \exp(-i\lambda x)dx. \tag{9}$$

For a kernel function $k(x) \in \mathcal{K}$, we have $\int_{-1}^{1}\int_{-1}^{1} k(r - s)f(r)f(s)drds \geq 0$ 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 $\widehat{\beta}$, $\widehat{\sigma}^2_{\beta,T}$ and the resulting t-statistic $\widehat{t}_\beta$. The theorem uses the following notation:

$$\widetilde{V}_x(r) = V_x(r) - \int_0^1 V_x(\tau)d\tau, \tag{10}$$

$$\widetilde{V}_y(r) = V_y(r) - \int_0^1 V_y(\tau)d\tau, \tag{11}$$

and

$$\widetilde{V}_{y,x}(r) = \widetilde{V}_y(r) - \left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_x(r)dr\right)^{-1}\left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_y(r)dr\right)\widetilde{V}_x(r). \tag{12}$$

Theorem 1 *Assume that x_t and y_t satisfy the functional central limit theorem in (1). Let $k(x)$ be a continuous function in $\mathcal{K}$, then*

$$\begin{aligned}
T^{d_x-d_y}\widehat{\beta} &\Rightarrow \left(\frac{\omega_y}{\omega_x}\right)\left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_y(r)dr\right)\left(\int_0^1 \widetilde{V}_x^2(r)dr\right)^{-1}, \\
T^{2d_x-2d_y}\widehat{\sigma}^2_{\beta,T} &\Rightarrow \frac{\omega_y^2}{\omega_x^2}\left(\int_0^1 \widetilde{V}_x^2(r)dr\right)^{-2}\int_0^1\int_0^1 \widetilde{V}_x(r)\widetilde{V}_{y,x}(r)k(r-s)\widetilde{V}_{y,x}(s)\widetilde{V}_x(s)drds, \\
\widehat{t}_\beta &\Rightarrow \left(\int_0^1 \widetilde{V}_x(r)\widetilde{V}_y(r)dr\right)\left(\int_0^1\int_0^1 \widetilde{V}_x(r)\widetilde{V}_{y,x}(r)k(r-s)\widetilde{V}_{y,x}(s)\widetilde{V}_x(s)drds\right)^{-1/2}.
\end{aligned} \tag{13}$$

Theorem 1 shows that a t-statistic does not necessarily diverge, as long as a proper variance estimator is used. The conventional variance estimator (also called 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 conventional 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 new HAC estimator incorporates autocovariances of all lags and delivers a standard error estimate that is of the same stochastic order as $\widehat{\beta}$. Based on such a HAC estimate, the new t-statistic is stochastically bounded and converges to a well-defined distribution.

Now we consider the spurious regression between a nonstationary $I(d)$ process and a linear trend. The data generating process for y_t is the same as before so that the invariance principle in (1) holds for $T^{-(2d_y-1)/2}y_{[Tr]}$. 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 new 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 $x_t = t$ and y_t satisfy the functional central limit theorem in (1). Let $k(x)$ be a continuous function in $\mathcal{K}$, then*

$$\begin{aligned}
&T^{d_x-d_y}\widehat{\beta} \Rightarrow 12\omega_y \left(\int_0^1 r\widetilde{V}_y(r)dr\right),\\
&T^{2d_x-2d_y}\widehat{\sigma}^2_{\beta,T} \Rightarrow 144\omega_y^2 \left(\int_0^1\int_0^1 (r-1/2)\widetilde{V}_{y,t}(r)k(r-s)\widetilde{V}_{y,t}(s)(s-1/2)drds\right),\\
&\widehat{t}_\beta \Rightarrow \left(\int_0^1 r\widetilde{V}_y(r)dr\right)\left(\int_0^1\int_0^1 (r-1/2)\widetilde{V}_{y,t}(r)k(r-s)\widetilde{V}_{y,t}(s)(s-1/2)drds\right)^{-1/2},
\end{aligned} \tag{14}$$

where

$$d_x = 3/2 \text{ and } \widetilde{V}_{y,t}(r) = \widetilde{V}_y(r) - \left(\int_0^1 r\widetilde{V}_y(r)dr\right)(12r-6). \tag{15}$$

Theorem 2 shows the new 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 $\sqrt{T}$ (for the unit root case, see Phillips and Durlauf 1988). This finding is consistent with a result by Phillips (1998), who considered regressing a unit root process on a complete orthnomal system in $L_2[0,1]$. He showed that the t-statistic based the usual HAC standard error with bandwidth M is of order $O_p\left((T/M)^{1/2}\right)$. For the new and usual t-statistics, the bandwidths are $M = T$ and $M = 1$, respectively. The former is thus stochastically bounded while the latter diverges at the rate of $\sqrt{T}$.

Together with Theorem 1, Theorem 2 shows that the new t-statistic converges in distribution in the spurious regression with nonstationary fractional processes. This finding implies that the new t-statistic 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 independent stationary $I(d)$ processes and that between a stationary $I(d)$ process and a linear trend.

Consider two Gaussian processes x_t and y_t with the following spectral densities $f_x(\lambda)$ and $f_y(\lambda)$:

$$f_x(\lambda) = \lambda^{-2d_x}\varphi_x(\lambda) \text{ and } f_y(\lambda) = \lambda^{-2d_y}\varphi_y(\lambda), \tag{16}$$

where $0 < d_x, d_y < 0.5$, $\varphi_x(\lambda)$ and $\varphi_y(\lambda)$ are continuous functions with $\varphi_x(0) = \omega_x^2 \in (0, \infty)$ and $\varphi_y(0) = \omega_y^2 \in (0, \infty)$. Given the above spectral densities, x_t and y_t have spectral representations:

$$x_t = \int_{-\pi}^{\pi} \exp(it\lambda) f_x^{1/2}(\lambda) dW_x(\lambda) \text{ and } y_t = \int_{-\pi}^{\pi} \exp(it\lambda) f_y^{1/2}(\lambda) dW_y(\lambda), \tag{17}$$

$t = 1, 2, ..., T$, where $W_x(\cdot)$ and $W_y(\cdot)$ are complex-valued, Gaussian random measures satisfying $EW_x(d\lambda)\overline{W_y(d\mu)} = 0$,

$$W_z(d\lambda) = \overline{W_z(-d\lambda)},\ EW_z(d\lambda) = 0, \text{for } z = x, y, \tag{18}$$

and

$$EW_z(d\lambda)\overline{W_z(d\mu)} = 1\{\lambda = \mu\} d\lambda, \text{ for } z = x, y, \tag{19}$$

where $1\{\cdot\}$ is the indicator function.

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

$$\begin{aligned} S_T(r) &= \left(S_T^x(r), S_T^y(r), S_T^{xy}(r)\right) \\ &= \left(T^{-(d_x+1/2)} \sum_{t=1}^{[Tr]} x_t, T^{-(d_y+1/2)} \sum_{t=1}^{[Tr]} y_t, T^{-d_x-d_y} \sum_{t=1}^{[Tr]} x_t y_t\right). \end{aligned} \tag{20}$$

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 y_t be the time series defined by (17). If $d_x, d_y \in (0, 1/2)$ and $d_x + d_y > 1/2$, then*

$$S_T(r) \Rightarrow \left(\omega_x B_{d_x}(r), \omega_y B_{d_y}(r), \omega_x \omega_y Z(r)\right), \tag{21}$$

where

$$B_{d_x}(r) = \int_{-\infty}^{\infty} \frac{\exp(i\xi r) - 1}{i\xi} |\xi|^{-d_x} dW_x(\xi), \tag{22}$$

$$B_{d_y}(r) = \int_{-\infty}^{\infty} \frac{\exp(i\eta r) - 1}{i\eta} |\eta|^{-d_x} dW_y(\eta), \tag{23}$$

and

$$Z(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(i(\xi + \eta)r) - 1}{i(\xi + \eta)} |\xi|^{-d_x} |\eta|^{-d_y} dW_x(\xi) dW_y(\eta). \tag{24}$$

Note that $B_{d_x}(r)$ and $B_{d_y}(r)$ are spectral representations of type I fractional Brownian motions (Samorodnisky 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 $x_t y_t$ 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_y 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 $\widehat{\sigma}^2_{\beta,T}$ (defined in (4) and (5)) and the t-statistic in the following theorem.

Theorem 4 *Let x_t and y_t be the time series defined by (17). Assume that $k(x)$ is a twice continuously differentiable function in $\mathcal{K}$. If $d_x, d_y \in (0, 1/2)$ and $d_x + d_y > 1/2$, then*

$$\begin{aligned}
T^{1-d_x-d_y}\widehat{\beta} &\Rightarrow \omega_x\omega_y \left(\int_{-\pi}^{\pi} f_x(\lambda)d\lambda\right)^{-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(\xi,\eta,1)\,|\xi|^{-d_x}\,|\eta|^{-d_y}\,dW_x(\xi)dW_y(\eta),\\
T^{2-2d_x-2d_y}\widehat{\sigma}^2_{\beta,T} &\Rightarrow \omega_x^2\omega_y^2 \left(\int_{0}^{1} f_x(\lambda)d\lambda\right)^{-2} \int_0^1\int_0^1 -k''(r-s)U(r)U(s)drds,\\
\widehat{t}_\beta = \widehat{\beta}/\widehat{\sigma}_{\beta,T} &\Rightarrow \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(\xi,\eta,1)\,|\xi|^{-d_x}\,|\eta|^{-d_y}\,dW_x(\xi)dW_y(\eta)\right)\\
&\quad\times\left(\int_0^1\int_0^1 -k''(r-s)U(r)U(s)drds\right)^{-1/2}, \qquad (25)
\end{aligned}$$

where

$$\psi(\xi,\eta,r) = \frac{\exp(i(\xi+\eta)r)-1}{i(\xi+\eta)} - \frac{\exp(i\xi r)-1}{i\xi}\frac{\exp(i\eta r)-1}{i\eta}, \qquad (26)$$

$$U(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\psi(\xi,\eta,r) - r\psi(\xi,\eta,1))\,|\xi|^{-d_x}\,|\eta|^{-d_y}\,dW_x(\xi)dW_y(\eta). \qquad (27)$$

The most important finding in the above theorem is the convergence of the new t-statistic. In contrast, Tsay and Chung (2000) showed that the t-statistic based on the OLS standard error diverges at the rate of $T^{d_x+d_y-0.5}$. As a consequence, the slope coefficient in the regression between two stationary long memory processes can be spuriously significant. The convergence of the new t-statistic 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, when the new t-statistic is used in place of the conventional one, all of the statistics behave as in the case of usual regression. Hence, inferences based on the new t-statistic 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 stationary long memory processes, as long as a proper t-statistic 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 (2002). 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 stationary 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 (2002) 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 $\widehat{\beta}$, $\widehat{\sigma}^2_{\beta,T}$ and the t-statistic when the sharp kernels are employed.

Theorem 5 *Let x_t and y_t be the time series defined by (17). If $k(x) = (1-|x|)^{\rho}1\{|x| \leq 1\}$, $d_x, d_y \in (0, 1/2)$ and $d_x + d_y > 1/2$, then the results of Theorem 4 hold with $\int_0^1\int_0^1 -k''(r-s)U(r)U(s)drds$ replaced by*

$$2\rho\int_0^1 U^2(r)dr - \rho(\rho-1)\iint''_{[0,1]^2} U(r)(1-|r-s|)^{\rho-2}U(s)drds, \tag{28}$$

where the second term is defined to be zero when $\rho = 1$ and $\iint''_{[0,1]^2}$ indicates that the integration on the diagonal $r = s$ is excluded.

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

$$T^{-(d_y+1/2)}\sum_{t=1}^{[Tr]} y_t \Rightarrow B_{d_y}(r). \tag{29}$$

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

$$T^{-(d_y+3/2)} \sum_{t=1}^{[Tr]} t y_t \Rightarrow r B_{d_y}(r) - \int_0^r B_{d_y}(s)ds. \tag{30}$$

Let

$$\begin{aligned} G(r) &= (r - \frac{1}{2}) B_{d_y}(r) - \int_0^r B_{d_y}(s)ds - B_{d_y}(1) \left(\int_0^r (s - 1/2)ds \right) \\ &\quad - \left(6 B_{d_y}(1) - 12 \int_0^1 B_{d_y}(s)ds \right) \int_0^r \left(s - \frac{1}{2} \right)^2 ds. \end{aligned} \tag{31}$$

Then we can prove the following theorem using (29) and (30) and the arguments similar to the proof of Theorem 4. Details are omitted.

Theorem 6 *Let y_t be the time series defined by (17) with $d_y \in (0, 1/2)$ and x_t be the linear trend: $x_t = t$. If $k(x)$ is a twice continuously differentiable function in $\mathcal{K}$, then*

$$T^{3/2-d_y} \widehat{\beta} \Rightarrow \omega_y \left(6 B_{d_y}(1) - 12 \int_0^1 B_{d_y}(s)ds \right), \tag{32}$$

$$T^{3-2d_y} \widehat{\sigma}^2_{\beta,T} \Rightarrow 144 \omega_y^2 \int_0^1 \int_0^1 -k''(r-s) G(r) G(s) dr ds, \tag{33}$$

$$\widehat{t}_\beta \Rightarrow \left(\frac{1}{2} B_{d_y}(1) - \int_0^1 B_{d_y}(s) \right) \left(\int_0^1 \int_0^1 -k''(r-s) G(r) G(s) dr ds \right)^{-1/2}. \tag{34}$$

If $k(x) = (1 - |x|)^\rho 1\{|x| \le 1\}$ for some integer $\rho \ge 1$, then (32), (33), and (34) hold provided that $\int_0^1 \int_0^1 -k''(r-s) G(r) G(s) dr ds$ is replaced by

$$2\rho \int_0^1 G^2(r) dr - \rho(\rho - 1) \iint''_{[0,1]^2} G(r)(1 - |r - s|)^{\rho-2} G(s) dr ds, \tag{35}$$

where the second term is defined to be zero when $\rho = 1$.

Theorem 6 shows that the OLS estimator is consistent and the new 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 a proper t-statistic is employed and critical values from the correct limiting distribution are used.

4 Kernel Estimates of Asymptotic Distributions

The limiting distributions of $\widehat{t}_\beta$ in (13), (14), (25) and (34) are nonstandard. In this section, we use Monte Carlo simulations to approximate their probability densities.

Note that the limiting distributions are invariant to ω_x and ω_y. It suffices to simulate simple fractionally integrated processes. Specifically, we generate the fractional processes x_t and y_t according to $(1-L)^{d_x} x_t = \varepsilon_{xt}$ and $(1-L)^{d_y} y_t = \varepsilon_{yt}$, where $\varepsilon_{xt} \sim iid(0,1)$, $\varepsilon_{yt} \sim iid(0,1)$ for $t > 0$, $\varepsilon_{xt} = \varepsilon_{yt} = 0$ for $t \leq 0$, and $\{\varepsilon_{xt}\}$ is independent of $\{\varepsilon_{yt}\}$. We let $k(\cdot)$ be the sharp kernels with the sharpness index $\rho = 1, 4, 8$. The simulated estimates use 2000 replications and a sample size of 1000. For spurious regressions between nonstationary $I(d)$ processes, we consider $(d_x, d_y) = (0.6, 0.6), (0.6, 1)$, $(1, 0.6)$, or $(1, 1)$; and for those between stationary ones, we let $(d_x, d_y) = (0.3, 0.3), (0.4, 0.2)$ or $(0.2, 0.4)$.

We first consider spurious regressions with nonstationary fractional processes. Figure 1 reports the kernel estimates of the probability densities for the case $x_t \sim I(d_x)$, $y_t \sim I(d_y)$ with $d_x = d_y = 0.6$. The qualitative results for other (d_x, d_y) combinations are similar. The probability densities appear to be symmetric and are apparently more dispersed than the standard normal density. For example, when the Bartlett kernel is used, the 95% quantile of the limiting distribution is 4.153, which is larger than 1.645, the 95% quantile of the standard normal distribution. Interestingly, the larger the sharpness index, the less dispersed the limiting distribution. For example, when $\rho = 8$, the 95% quantile becomes 2.463, which is quite close to 1.645. In this case, the probability of $\left|\widehat{t}_\beta\right| > 1.96$ is 30.30%. Therefore, when $\rho = 8$ and the new t-statistic is used to test the significance of the slope coefficient, we will erroneously reject the null 30.30% of the times when the wrong critical value is used. In contrast, when the usual t-statistic is employed, the rejection probability goes to one as the sample size increases. When $T = 1000$, the rejection probability is 75.9%, as shown by simulations. Hence the use of the new t-statistic 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 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 with $(d_x, d_y) = (0.4, 0.2)$. The density estimates for the other two cases turn out to be close to the case $(d_x, d_y) = (0.4, 0.2)$. The figure shows that the limiting distributions are more concentrated around the origin than in the nonstationary cases. For example, the 90% quantiles when the sharpness index $\rho = 1, 4, 8$ are $2.677, 1.736$, and 1.556, respectively. The corresponding 95% quantiles are $3.647, 2.339$, and 2.064. Simulation results show that the limiting distribution becomes closer to the standard normal when ρ is larger. For example, when $\rho = 8$, the probability of $\left|\widehat{t}_\beta\right| > 1.96$ is 12.90%, which is very close to 10%, the size of the test when $\widehat{t}_\beta$ is standard normal. In other words, when the new t-statistic is used to test the null of $\beta = 0$, the probability of wrong rejection is only 12.90% 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 more concentrated as ρ increases. 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 has proposed a new t-statistic that is convergent in all the cases of spurious regressions considered in the literature. This new t-statistic is based on the HAC estimator using a truncation lag or bandwidth equal to the sample size. The paper argues 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. The paper reinforces the warnings that hypothesis testing using the OLS standard error can lead to misleading inference. This is true even if the underlying processes are stationary (Granger, Hyung and Jeon 2001). To avoid or alleviate this problem, a pre-whitening HAC standard error or a HAC standard error with the truncation lag growing with the persistence of the underlying processes may be used (Andrews 1991; Andrews and Monahan 1992). It turns out that in a spurious regression, the truncation lag needs to be as large as the sample size or at least proportional to the sample size.

In view of the papers by Kiefer and Vogelsang (2002a, 2002b), Phillips, Sun and Jin (2002) and Sun (2002), the new t-statistic converges to a well-defined distribution in the usual regressions with stationary covariates and regression errors, cointegrating regressions and spurious regressions. Therefore, it has the potential to deliver a unified inferential framework. The advantage of the new t-statistic is that it converges in distribution without any normalization. In contrast, to make an asymptotically valid inference, the usual t-statistic has to be normalized by T^κ, where κ depends on unknown memory parameters that characterize the persistence of the underlying processes.

The paper is a first step towards the asymptotic properties of the new t-statistic with 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 bandwidth does not have to be the sample size to deliver a convergent t-statistic. It suffices that the bandwidth is proportional to the sample size. Finally, the limiting distributions depend on the kernel used. Therefore, it is desirable to investigate whether there exists an optimal kernel according to a certain criterion, such as the power of the t-test that it delivers.

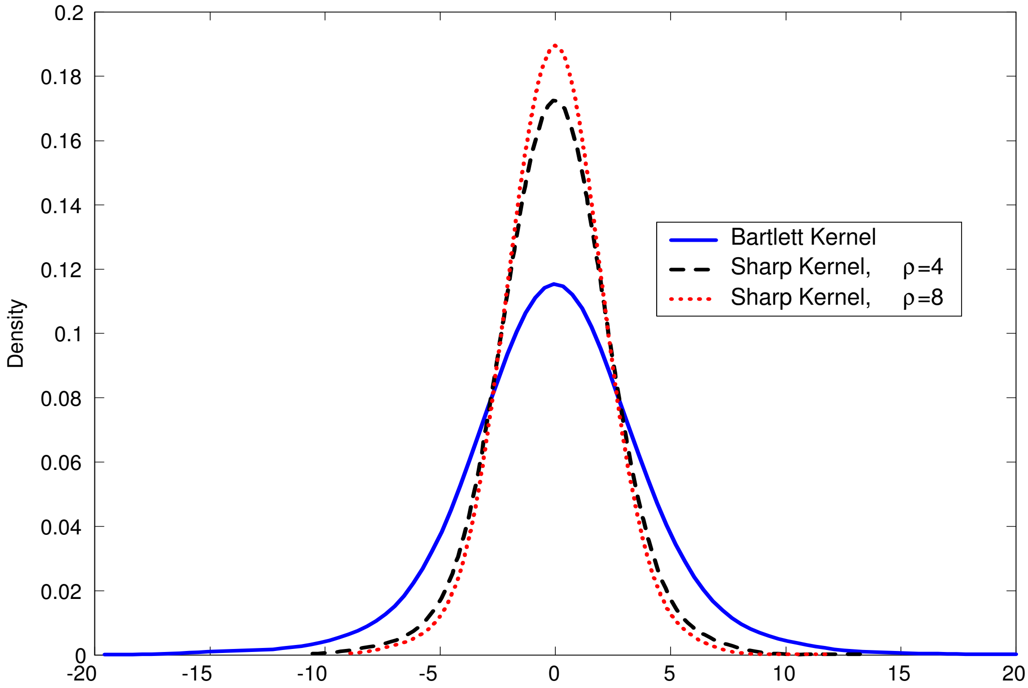

Figure 1. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t \sim I(0.6)$ and $y_t \sim I(0.6)$

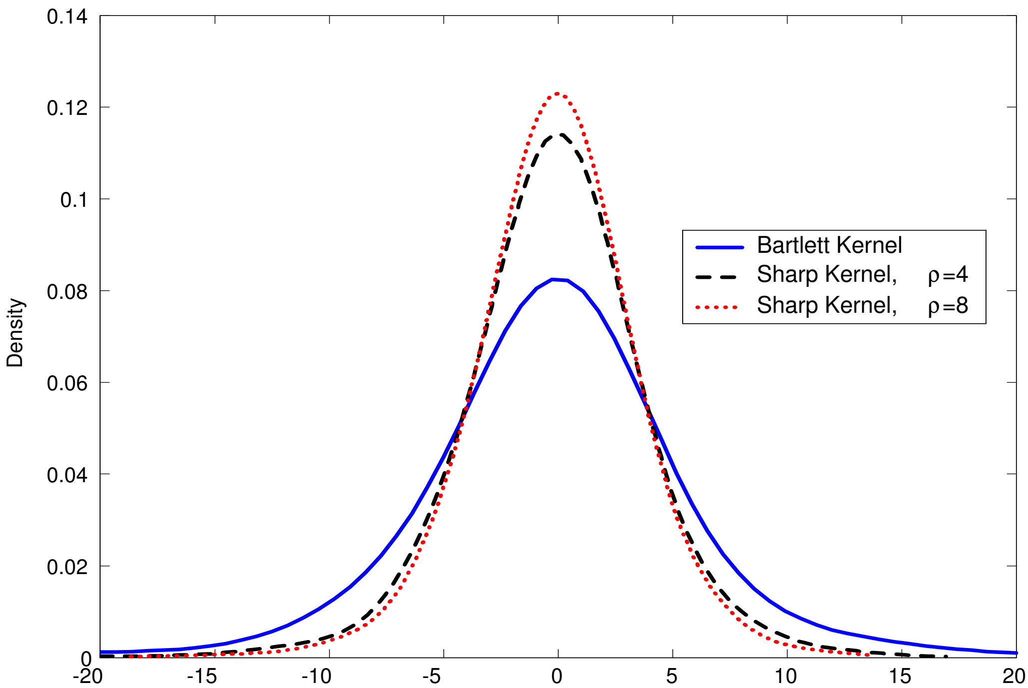

Figure 2. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t = t$ and $y_t \sim I(0.6)$

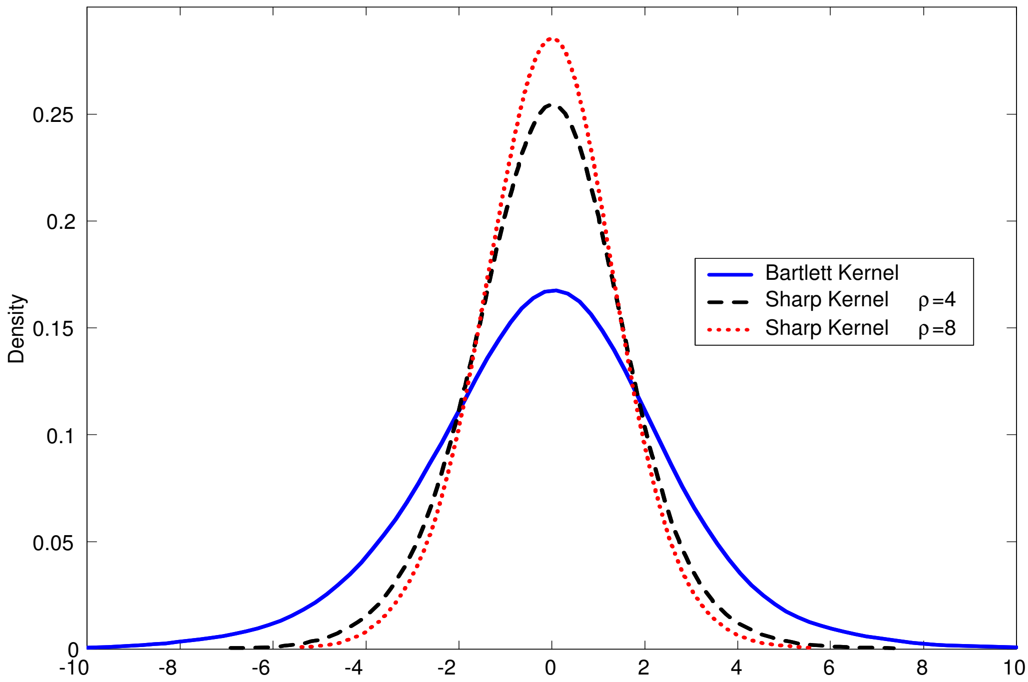

Figure 3. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t \sim I(0.4)$ and $y_t \sim I(0.2)$

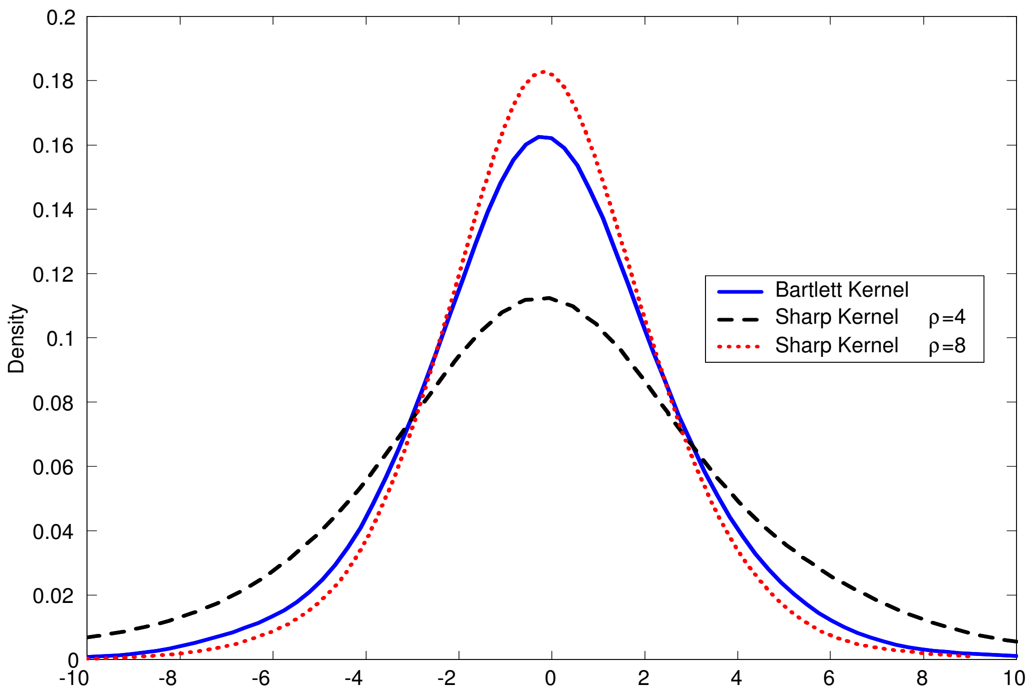

Figure 4. kernel estimates of densities of $\widehat{t}_\beta$ when $x_t = t$ and $y_t \sim I(0.3)$

6 Appendix of Proofs

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

$$T^{-d_x-d_y}\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})\Rightarrow\omega_x\omega_y\int_0^1\widetilde{V}_x(r)\widetilde{V}_y(r)dr \tag{36}$$

and

$$T^{-2d_x}\sum_{t=1}^{T}(x_t-\bar{x})^2\Rightarrow\omega_x^2\int_0^1\widetilde{V}_x^2(r)dr. \tag{37}$$

Hence

$$\begin{aligned} T^{d_x-d_y}\widehat{\beta} &= \frac{T^{-d_x-d_y}\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})}{T^{-2d_x}\sum_{t=1}^{T}(x_t-\bar{x})^2} \\ &\Rightarrow (\omega_y/\omega_x)\left(\int_0^1\widetilde{V}_x(r)\widetilde{V}_y(r)dr\right)\left(\int_0^1\widetilde{V}_x^2(r)dr\right)^{-1}. \end{aligned} \tag{38}$$

As a consequence,

$$\begin{aligned} T^{-(2d_y-1)/2}\widehat{u}_{[Tr]} &= T^{-(2d_y-1)/2}\left(y_{[Tr]}-\bar{y}\right)-T^{d_x-d_y}\widehat{\beta}T^{-(2d_x-1)/2}(x_{[Tr]}-\bar{x}) \\ &\Rightarrow \omega_y\widetilde{V}_{y,x}(r). \end{aligned} \tag{39}$$

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

$$\begin{aligned} &\left(\frac{1}{T}\sum_{t=1}^{T}(\frac{x_t-\bar{x}}{T^{d_x-1/2}})^2\right)^{-2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{x_t-\bar{x}}{T^{d_x-1/2}}\frac{\widehat{u}_t}{T^{d_y-1/2}}k(\frac{r-s}{T})\frac{\widehat{u}_s}{T^{d_y-1/2}}\frac{x_s-\bar{x}}{T^{d_x-1/2}} \\ \Rightarrow\ &\omega_y^2\omega_x^{-2}\left(\int_0^1\widetilde{V}_x^2(r)dr\right)^{-2}\int_0^1\int_0^1\widetilde{V}_x(r)\widetilde{V}_{y,x}(r)k(r-s)\widetilde{V}_{y,x}(s)\widetilde{V}_x(s)drds, \end{aligned} \tag{40}$$

where the last line follows from the continuous mapping theorem. In view of (38) and (40), we have

$$\begin{aligned} &\widehat{t}_\beta=\left(\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})\right)\left(\sum_{t=1}^{T}\sum_{s=1}^{T}(x_t-\bar{x})\widehat{u}_t k(\frac{r-s}{T})\widehat{u}_s(x_s-\bar{x})\right)^{-1/2} \\ \Rightarrow &\left(\int_0^1\widetilde{V}_x(r)\widetilde{V}_y(r)dr\right)\left(\int_0^1\int_0^1\widetilde{V}_x(r)\widetilde{V}_{y,x}(r)k(r-s)\widetilde{V}_{y,x}(s)\widetilde{V}_x(s)drds\right)^{-1/2}. \end{aligned} \tag{41}$$

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 $S_T^y(r)$ 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 $\{S_T^y(r)\}$ are tight. It remains to show the tightness of $\{S_T^{xy}(r)\}$. In view of Theorem 13.5 of Billingsley (1999), it suffices to show that, for almost all sample paths, some constant $C > 0$ and $0 \leq r_1 \leq r \leq r_2 \leq 1$,

$$P\left(\left|S_T^{xy}(r) - S_T^{xy}(r_1)\right| \geq \lambda, \left|S_T^{xy}(r_2) - S_T^{xy}(r)\right| \geq \lambda\right) \leq C\lambda^{-4}(r_2 - r_1)^{2\nu}, \tag{42}$$

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

$$\begin{aligned} & P\left(\left|S_T^{xy}(r) - S_T^{xy}(r_1)\right| \geq \lambda, \left|S_T^{xy}(r_2) - S_T^{xy}(r)\right| \geq \lambda\right) & (43) \\ \leq & \frac{E\left(S_T^{xy}(r) - S_T^{xy}(r_1)\right)^2}{\lambda^2} \frac{E\left(S_T^{xy}(r_2) - S_T^{xy}(r)\right)^2}{\lambda^2}. & (44) \end{aligned}$$

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

$$\begin{aligned} & E\left(S_T^{xy}(r) - S_T^{xy}(r_1)\right)^2 \\ = & T^{-2d_x-2d_y} E\left(\sum_{t=[Tr_1]+1}^{[Tr]} x_t y_t\right)^2 \\ = & T^{-2d_x-2d_y} 2 \sum_{t=[Tr_1]+1}^{[Tr]} \sum_{\tau=t+1}^{[Tr]} (E x_t x_\tau)(E y_t y_\tau) + T^{-2d_x-2d_y} \sum_{t=[Tr_1]+1}^{[Tr]} \left(E x_t^2\right)\left(E y_t^2\right) \\ = & C T^{-2d_x-2d_y} \sum_{t=[Tr_1]+1}^{[Tr]} \sum_{\tau=t+1}^{[Tr]} (\tau - t)^{2d_x+2d_y-2} + C T^{1-2d_x-2d_y} \left(\frac{[Tr] - [Tr_1]}{T}\right) \\ \leq & C \int_{r_1}^{r} \left(\int_t^r (\tau - t)^{2d_x+2d_y-2} d\tau\right) dt + C T^{1-2d_x-2d_y} \left(\frac{[Tr] - [Tr_1]}{T}\right) \\ = & \frac{2C}{(2d_x + 2d_y - 1)(2d_x + 2d_y)} (r - r_1)^{2d_x+2d_y} + o(1) \\ \leq & C(r - r_1)^{2d_x+2d_y}, \end{aligned} \tag{45}$$

where we use the fact that $E x_t x_\tau \leq C(\tau - t)^{2d_x-1}$ and $E y_t y_\tau \leq C(\tau - t)^{2d_y-1}$. Combining (43) with (45) yields

$$\begin{aligned} & P\left(\left|S_T^{xy}(r) - S_T^{xy}(r_1)\right| \geq \lambda, \left|S_T^{xy}(r_2) - S_T^{xy}(r)\right| \geq \lambda\right) \\ \leq & C\lambda^{-4}(r - r_1)^{2d_x+2d_y}(r_2 - r)^{2d_x+2d_y} \leq C\lambda^{-4}(r_2 - r_1)^{4d_x+4d_y}. \end{aligned} \tag{46}$$

Therefore (42) holds and $\{S_T^{xy}(r)\}$ 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 $(S_T^x(r), S_T^y(r))$ converges to that of $(\omega_x B_{d_x}(r), \omega_y B_{d_y}(r))$. The fidi convergence of $S_T^{xy}(r)$ follows from Theorem 7.4 of Giraitis and Taqqu (1999). Put in our context, this theorem says that if x_t and y_t follow linear processes with the same iid innovation sequences, then

$$T^{-d_x-d_y}\sum_{t=1}^{[Tr]} x_t y_t \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(i(\xi+\eta)r)-1}{i(\xi+\eta)} |\xi|^{-d_x} |\eta|^{-d_y} \, dW_x(\xi)dW_y(\eta) \quad (47)$$

for any $r \in (0,1]$. To see this, we use the notation in Giraitis and Taqqu (1999), and set $i=1, m_i=n_i=1, l=2, \alpha^{(i,1)}=2d_x, \alpha^{(i,2)}=2d_y, N=T$, $b(\tau)=1\{\tau=0\}$. For these special parameter and function specifications,

$$Q_{[Tr]} = \sum_{t=1}^{[Tr]}\sum_{s=1}^{[Tr]} b(t-s)P_{m_i,n_i}(x_t, y_s), \quad (48)$$

the partial sum process considered by Giraitis and Taqqu (1999), becomes

$$Q_{[Tr]} = \sum_{t=1}^{[Tr]} x_t y_t, \quad (49)$$

and the limiting process can be shown to be $\omega_x\omega_y Z(r)$ (Note that they use $Z(\cdot)$ for the orthogonal Gaussian measure where we use $W(\cdot)$). Our case differs from the above special case of Giraitis and Taqqu (1999) only in that we assume that x_t and y_t are independent processes where Giraitis and Taqqu assume that x_t and y_t share the same innovation sequences. Nevertheless, their proof goes through for the independent case with obvious and minor modifications. Finally, the joint fidi convergence of $(S_T^x(r), S_T^y(r))$ with $S_T^{xy}(r)$ 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 $W_y(\cdot)$. ■

Proof of Theorem 4. By ergodicity, we have

$$\text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} x_t^2 = \int_{-\pi}^{\pi} f_x(\lambda)d\lambda, \text{ and } \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} x_t = 0. \quad (50)$$

Therefore

$$\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T} x_t^2 - (\bar{x})^2 = \int_{-\pi}^{\pi} f_x(\lambda)d\lambda. \quad (51)$$

Combining (51) with Lemma 3 yields

$$
\begin{aligned}
& T^{1-d_x-d_y}\widehat{\beta} = T^{-d_x-d_y}\left(\sum_{t=1}^{T}(x_t y_t - T\bar{x}\bar{y})\right)\left(T^{-1}\sum_{t=1}^{T}x_t^2 - (\bar{x})^2\right)^{-1} \\
& \Rightarrow \omega_x\omega_y\left(\int_{-\pi}^{\pi} f_x(\lambda)d\lambda\right)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(\xi,\eta,1)\,|\xi|^{-d_x}\,|\eta|^{-d_y}\,dW_x(\xi)dW_y(\eta). \qquad (52)
\end{aligned}
$$

We next consider the limiting distribution of

$$
\sum_{t=1}^{T}\sum_{s=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-s}{T})\widehat{u}_s\,(x_s-\bar{x})\,. \qquad (53)
$$

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

$$
\begin{aligned}
& T^{-d_x-d_y}S_T^v(r) = T^{-d_x-d_y}\sum_{t=1}^{[Tr]}(x_t-\bar{x})\,(y_t-\bar{y}) - T^{-d_x-d_y}\widehat{\beta}\sum_{t=1}^{[Tr]}(x_t-\bar{x})^2 \\
\Rightarrow \quad & \omega_x\omega_y\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\psi(\xi,\eta,r) - r\psi(\xi,\eta,1))\,|\xi|^{-d_x}\,|\eta|^{-d_y}\,dW_x(\xi)dW_y(\eta) \qquad (54) \\
: \quad & = \omega_x\omega_y U(r),
\end{aligned}
$$

where we have used

$$
\text{plim}_{T\to\infty}\left(\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^{-1}\sum_{t=1}^{[Tr]}(x_t-\bar{x})^2 = r. \qquad (55)
$$

Using a well-known formula, we have

$$
\begin{aligned}
& T^{-2d_x-2d_y}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{T})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\
= \; & T^{-2d_x-2d_y}\sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1}\left\{S_T^v(t/T)\left(2k(\frac{t-\tau}{T}) - k(\frac{t-\tau-1}{T}) - k(\frac{t-\tau+1}{T})\right)\right. \\
& \left.\times S_T^v(\tau/T)\right\} + T^{-2d_x-2d_y}S_T^v(1)\sum_{\tau=1}^{T-1}\left(k(\frac{T-\tau}{T}) - k(\frac{T-\tau-1}{T})\right)S_T^v(\tau/T) \\
& + T^{-2d_x-2d_y}\left\{\sum_{t=1}^{T-1}S_T^v(t/T)\left(k(\frac{t-T}{T}) - k(\frac{t-T+1}{T})\right)S_T^v(1) + S_T^v(1)S_T^v(1)\right\} \\
= \; & \frac{1}{T^2}\sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1}T^{-d_x-d_y}S_T^v(t/T)T^2 D_T(\frac{t-\tau}{T})T^{-d_x-d_y}S_T^v(\tau/T), \qquad (56)
\end{aligned}
$$

where

$$D_T(\frac{t-\tau}{T}) = 2k(\frac{t-\tau}{T}) - k(\frac{t-\tau-1}{T}) - k(\frac{t-\tau+1}{T}) \tag{57}$$

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

$$T^2 D_T(\frac{t-\tau}{T}) \to -k''(r-s). \tag{58}$$

Combining (54), (56), and (58), and invoking the continuous mapping theorem, we get

$$\begin{aligned} & T^{-2d_x-2d_y} \sum_{t=1}^{T} \sum_{\tau=1}^{T} (x_t - \bar{x})\, \widehat{u}_t k(\frac{t-\tau}{T}) \widehat{u}_\tau\, (x_\tau - \bar{x}) \\ \Rightarrow \quad & -\omega_x^2 \omega_y^2 \int_0^1 \int_0^1 U(r) k''(r-s) U(s) dr ds. \end{aligned} \tag{59}$$

Therefore

$$\begin{aligned} & T^{2-2d_x-2d_y} \widehat{\sigma}_{\beta,T}^2 \\ = \quad & \left(\frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-2} T^{-2d_x-2d_y} \sum_{t=1}^{T} \sum_{\tau=1}^{T} (x_t - \bar{x})\, \widehat{u}_t k(\frac{t-\tau}{T}) \widehat{u}_\tau\, (x_\tau - \bar{x}) \\ \Rightarrow \quad & -\omega_x^2 \omega_y^2 \left(\int_0^1 f_x(\lambda) d\lambda \right)^{-2} \int_0^1 \int_0^1 U(r) k''(r-s) U(s) dr ds. \end{aligned} \tag{60}$$

Combining (52) with (60) 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,

$$D_T(\frac{t-\tau}{T}) = \frac{2}{T} 1(t=\tau).$$

Hence

$$\begin{aligned} & T^{-2d_x-2d_y} \sum_{t=1}^{T} \sum_{\tau=1}^{T} (x_t - \bar{x})\, \widehat{u}_t k(\frac{t-\tau}{T}) \widehat{u}_\tau\, (x_\tau - \bar{x}) \\ = \quad & \frac{2}{T} \sum_{\tau=1}^{T-1} T^{-2d_x-2d_y} S_T^v(t/T) S_T^v(t/T) \Rightarrow 2\omega_x^2 \omega_y^2 \int_0^1 U^2(r) dr. \end{aligned} \tag{61}$$

For other sharp kernels $k(x) = (1-|x|)^\rho \{|x| \le 1\}$ for $\rho \ge 2$, we have,

$$\lim_{T\to\infty} T^2 D_T(\frac{t-\tau}{T}) = -\rho(\rho-1)(1-|r-s|)^{\rho-2}$$

provided that $\lim_{T\to\infty}(t-\tau)/T = r-s \neq 0$. When $\lim_{T\to\infty}(t-\tau)/T = 0$, it is easy to see that

$$\lim_{T\to\infty} TD_T(\frac{t-\tau}{T}) = 0, \text{ if } t-\tau \neq 0, \tag{62}$$

and

$$\lim_{T\to\infty} TD_T(\frac{t-\tau}{T}) = 2\rho, \text{ if } t-\tau = 0. \tag{63}$$

Therefore

$$\begin{aligned}
& T^{-2d_x-2d_y}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\,\widehat{u}_t k(\frac{t-\tau}{T})\widehat{u}_\tau\,(x_\tau-\bar{x}) \\
= \quad & \frac{1}{T^2}\sum_{\tau=1}^{T-1}\sum_{t=1,t\neq\tau}^{T-1} T^{-d_x-d_y}S_T^v(t/T)T^2D_T(\frac{t-\tau}{T})T^{-d_x-d_y}S_T^v(\tau/T) \\
& +\frac{2\rho}{T}\sum_{\tau=1}^{T-1}T^{-2d_x-2d_y}S_T^v(\tau/T)S_T^v(\tau/T) \\
\Rightarrow \quad & \omega_x^2\omega_y^2\left(2\rho\int_0^1 U^2(r)dr - \rho(\rho-1)\iint_{[0,1]^2}'' U(r)(1-|r-s|)^{\rho-2}U(s)drds\right).
\end{aligned} \tag{64}$$

The theorem now follows from (61), (64) 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.

Andrews, D. W. K. and J. C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953-966.

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 autocovariance, 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.

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.

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. 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 (2002) Simple robust hypothesis testing using sharp and steep kernels without truncation, Manuscript in preparation. Department of Economics, Yale University.

Samorodnisky, 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.