Econometric Theory, **20**, 2004, 943–962. Printed in the United States of America.
DOI: 10.1017/S0266466604205072

A CONVERGENT *t*-STATISTIC IN SPURIOUS REGRESSIONS

YIXIAO SUN
University of California, San Diego

This paper investigates the asymptotic properties of the *t*-statistic in spurious regressions when the bandwidth in the estimation of the heteroskedasticity and autocorrelation consistent (HAC) standard error is set 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 nondegenerate limiting distribution for all cases of spurious regressions considered in the literature. This finding sheds some new light on the nature of spurious regressions.

1. INTRODUCTION

Since the first Monte Carlo study by Granger and Newbold (1974), many efforts have been made to understand the nature of spurious regressions. Phillips (1986) develops 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 the approach of Phillips (1986), Durlauf and Phillips (1988) and Marmol (1995, 1998) find 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) have 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 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 ordinary least squares (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 heteroskedasticity and autocorrelation consistent (HAC) standard error estimator with bandwidth (M) proportional to the sample size. Specifically, we set $M = bT$ for some $b \in (0,1]$.

I am very grateful to Bruce Hansen and two anonymous referees for helpful comments and suggestions. All remaining errors and omissions are mine alone. Address correspondence to Yixiao Sun, Department of Economics, 0508, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA; e-mail: yisun@ucsd.edu.

© 2004 Cambridge University Press 0266-4666/04 $12.00

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

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. We show that when $M = bT$, the t-statistic converges to a well-defined distribution. This is true for spurious regressions between two independent fractional processes, stationary or nonstationary, and between a fractional process and a linear trend. Our finding suggests 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 the nature of 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.

For all the cases considered, the limiting distributions are nonstandard and depend on the memory of the underlying processes. If memory parameters are known, these limiting distributions can be simulated using Monte Carlo experiments. However, in practice memory parameters are generally unknown. Given the nonstandard distribution that is not nuisance parameter free, the new asymptotic theory is more of theoretical interest than of practical importance. In fact, when the underlying processes are $I(d)$, a narrow-band frequency-domain approach may dominate the time-domain approach employed in this paper. Robinson (1994) provides such an example. In view of these observations, this paper should be viewed as a theoretical investigation that helps deepen the understanding of spurious regressions.

The results of this paper are related to those of Kiefer and Vogelsang (2003). They consider 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 rise to 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 also. The difference is that in the usual regression the t-statistic converges regardless of whether $M = bT$ or $M = o(T)$, whereas in the spurious regression the t-statistic converges when $M = bT$ and diverges when $M = o(T)$. Other related papers that investigate the HAC estimator without truncation include the work of Robinson (1998), who provides conditions under which the HAC estimator without smoothing is consistent, Kiefer and Vogelsang (2002a, 2002b), who develop asymptotically valid tests in usual regressions under the specification $M = T$, Jansson (2004), who studies the accuracy of the $M = T$ approximation for Gaussian location models, and Phillips, Sun, and Jin (2003a, 2003b), who propose a new family of kernels for HAC estimation without a bandwidth parameter.

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.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

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

2. SPURIOUS REGRESSIONS WITH NONSTATIONARY PROCESSES

Consider the following data generating process:

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

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

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

where d_x, $d_u > \frac{1}{2}$ and ε_{xt} and ε_{ut} are two weakly dependent processes. For simplicity, we assume that x_t is a scalar process. The results of this paper easily can be extended to multivariate cases.

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

$$V_x(r) = \frac{1}{\Gamma(d_x)}\int_0^r (r-s)^{d_x-1}\,dW_x(s),$$

$$V_u(r) = \frac{1}{\Gamma(d_u)}\int_0^r (r-s)^{d_u-1}\,dW_u(s), \tag{4}$$

$W(s) = (W_x(s), W_u(s))'$ 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., 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, because the long-run equilibrium error u_t is nonstationary, Marmol (1998) refers to the second situation

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

as a partially spurious case. For convenience, we will refer to both cases as spurious cases.

Consider regressing y_t on a constant and x_t,

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t, \qquad t = 1, \ldots, T. \tag{6}$$

The OLS estimate of β is given by

$$\hat{\beta} = \frac{\sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, \tag{7}$$

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

$$\hat{\sigma}_\beta^2 = \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1} T \hat{\Omega} \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1}, \tag{8}$$

where

$$\hat{\Omega} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{M}\right) \hat{\Gamma}(j), \tag{9}$$

$$\hat{\Gamma}(j) = \begin{cases} \dfrac{1}{T} \displaystyle\sum_{t=1}^{T-j} (x_{t+j} - \bar{x}) \hat{u}_{t+j} \hat{u}_t (x_t - \bar{x}) & \text{for } j \geq 0 \\ \dfrac{1}{T} \displaystyle\sum_{t=-j+1}^{T} (x_{t+j} - \bar{x}) \hat{u}_{t+j} \hat{u}_t (x_t - \bar{x}) & \text{for } j < 0, \end{cases} \tag{10}$$

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

To get a consistent estimate of the long-run variance of $(x_t - \bar{x})\hat{u}_t$, conventional asymptotic theory assumes that $M \to \infty$ such that $M = o(T)$. However, $(x_t - \bar{x})\hat{u}_t$ is nonstationary, and the variance of $\sum_{t=1}^{T} (x_t - \bar{x})\hat{u}_t / \sqrt{T}$ does not converge. In other words, the sum $\sum_{j=-\infty}^{\infty} |\hat{\Gamma}(j)|$ is infinite with the probability approaching one as $T \to \infty$. The infiniteness of this sum invalidates the usual truncation argument. As becomes clear later on, when M is taken to grow at a slower rate than T, the estimate $\hat{\Omega}$ underestimates the variability of $\sum_{t=1}^{T} (x_t - \bar{x})\hat{u}_t / \sqrt{T}$ 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.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

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

$$\mathcal{K} = \{k(\cdot) : \mathbb{R} \to [0,1] | k(x) = k(-x), k(0) = 1, \text{ and } K(\lambda) \geq 0, \forall \lambda \in \mathbb{R}\}, \quad \textbf{(11)}$$

where

$$K(\lambda) = \int_{-\infty}^{\infty} k(x) \exp(-i\lambda x)\, dx. \quad \textbf{(12)}$$

The class $\mathcal{K}$ is similar to class $\mathcal{K}_2$ in Andrews (1991).

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

$$\widetilde{V}_x(r) = V_x(r) - \int_0^1 V_x(\tau)\, d\tau, \quad \textbf{(13)}$$

$$\widetilde{V}_u(r) = V_u(r) - \int_0^1 V_u(\tau)\, d\tau, \quad \textbf{(14)}$$

and

$$\widetilde{V}_{u,x}(r) = \widetilde{V}_u(r) - \left(\int_0^1 \widetilde{V}_x(\tau) \widetilde{V}_x(\tau)\, d\tau \right)^{-1} \left(\int_0^1 \widetilde{V}_x(\tau) \widetilde{V}_u(\tau)\, d\tau \right) \widetilde{V}_x(r). \quad \textbf{(15)}$$

THEOREM 1. *Assume that x_t and u_t satisfy the FCLT 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 $T \to \infty$,*

$$\begin{aligned} T^{d_x - d_u}(\hat{\beta} - \beta) &\Rightarrow \left(\int_0^1 \widetilde{V}_x(r) \widetilde{V}_u(r)\, dr \right) \left(\int_0^1 \widetilde{V}_x^2(r)\, dr \right)^{-1}, \\ T^{2d_x - 2d_u} \hat{\sigma}_\beta^2 &\Rightarrow \left(\int_0^1 \widetilde{V}_x^2(r)\, dr \right)^{-2} \\ &\quad \times \int_0^1 \int_0^1 \widetilde{V}_x(r) \widetilde{V}_{u,x}(r) k\left(\frac{r-s}{b} \right) \widetilde{V}_{u,x}(s) \widetilde{V}_x(s)\, drds, \\ \hat{t}_\beta &\Rightarrow \left(\int_0^1 \widetilde{V}_x(r) \widetilde{V}_u(r)\, dr \right) \\ &\quad \times \left(\int_0^1 \int_0^1 \widetilde{V}_x(r) \widetilde{V}_{u,x}(r) k\left(\frac{r-s}{b} \right) \widetilde{V}_{u,x}(s) \widetilde{V}_x(s)\, drds \right)^{-1/2}. \end{aligned} \quad \textbf{(16)}$$

Theorem 1 shows that when $M = bT$, the HAC estimator delivers a standard error estimate that is of the same stochastic order as $\hat{\beta} - \beta$. Based on such a HAC estimate, the *t*-statistic is stochastically bounded and converges to a well-

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

defined distribution. In contrast, the conventional variance estimator (also called the OLS variance estimator), which is $(T^{-1}\sum_{t=1}^{T}\hat{u}_t^2)/\sum_{t=1}^{T}(x_t-\bar{x})^2$, ignores the autocorrelation in the underlying processes and underestimates $\text{var}(\hat{\beta})$ by an order of magnitude. 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).

Our finding is consistent with a result by Phillips (1998), who considers regressing a unit root process on a complete orthonormal system in $L_2[0,1]$. He shows that the t-statistic based on the usual HAC standard error with bandwidth M is of order $O_p((T/M)^{1/2})$. 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 $\sqrt{T/M}$.

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 $T^{-(2d_u-1)/2}u_{[\text{Tr}]}$. The data generating process for x_t is replaced by $x_t=t$ so that $T^{-(2d_x-1)/2}x_{[\text{Tr}]}\to r$ for $d_x=\frac{3}{2}$. Using arguments similar to the proof of Theorem 1, we can prove the following theorem immediately. The details are omitted.

THEOREM 2. *Assume that $x_t=t$, u_t satisfies the FCLT 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*

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

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

$$\hat{t}_\beta\Rightarrow\left(\int_0^1 r\widetilde{V}_u(r)\,dr\right)\times\left(\int_0^1\int_0^1\left(r-\frac{1}{2}\right)\widetilde{V}_{u,t}(r)k\left(\frac{r-s}{b}\right)\times\widetilde{V}_{u,t}(s)\left(s-\frac{1}{2}\right)drds\right)^{-1/2}, \tag{17}$$

where

$$d_x=\frac{3}{2}\quad\text{and}\quad\widetilde{V}_{u,t}(r)=\widetilde{V}_u(r)-\left(\int_0^1\tau\widetilde{V}_u(\tau)\,d\tau\right)(12r-6). \tag{18}$$

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 con-

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

trast, the usual *t*-statistic diverges at the rate of $\sqrt{T}$ (see Durlauf and Phillips, 1988; Marmol and Velasco, 2002).

Results similar to Theorem 2 have also been obtained by Marmol and Velasco (2002), who analyze the effects of spuriously detrending a nonstationary fractional process. Put in our context, they used the long-run variance estimate of the form:

$$\hat{\Omega} = \sum_{j=-T+1}^{T-1} \hat{\Gamma}_{xx}(j)\hat{\Gamma}_{\hat{u}\hat{u}}(j), \tag{19}$$

where $\hat{\Gamma}_{xx}(j) = T^{-1}\sum_{t=1}^{T-|j|} x_{t+|j|}x_t$, $\hat{\Gamma}_{\hat{u}\hat{u}}(j) = T^{-1}\sum_{t=1}^{T-|j|} \hat{u}_{t+|j|}\hat{u}_t$. As a result of 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 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.

The first data generating process we consider is the same as that in Chung (2002). For completeness, we briefly describe the data generating process. We assume that y_t is generated according to

$$y_t = \alpha + x_t\beta + u_t, \tag{20}$$

where $(x_t, u_t)'$ is a vector linear process defined by

$$(x_t, u_t)' = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \tag{21}$$

$\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(\ldots, \varepsilon_{t-1}, \varepsilon_t)$, $E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma$ for some matrix Σ, and $\{\Psi_j\}$ is a sequence of 2×2 matrices. As in Chung (2002), we maintain the following assumptions.

A1. For some $\delta > 0$, $\sup_t E(|\varepsilon_{it}|^{2+\delta}|\mathcal{F}_{t-1}) < \infty$, a.s. for $i = 1$ and 2.

A2. For some nonsingular constant matrix Π,

$$\Psi_j \sim \begin{pmatrix} \frac{1}{\Gamma(d_x)} j^{d_x-1} & 0 \\ 0 & \frac{1}{\Gamma(d_u)} j^{d_u-1} \end{pmatrix} \Pi \quad \text{as } j \to \infty,$$

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

where $\sim$ means the ratio of the left- and right-hand-side elements tends to one and $\Gamma(\cdot)$ is the gamma function.

A3. $0 < d_x, d_u < 0.5$.

Define the random vector element

$$S_T(r) = (S_T^x(r), S_T^u(r), S_T^{xu}(r))$$
$$= \left(T^{-(d_x+1/2)} \sum_{t=1}^{[\mathrm{Tr}]} x_t, T^{-(d_u+1/2)} \sum_{t=1}^{[\mathrm{Tr}]} u_t, T^{-d_x-d_u} \sum_{t=1}^{[\mathrm{Tr}]} (x_t u_t - E x_t u_t) \right).$$

Under Assumptions A1–A3, the following lemma holds.

LEMMA 3. *Let Assumptions A1–A3 hold. If* $d_x + d_u > \frac{1}{2}$, *then*

$$S_T(r) \Rightarrow (B_{d_x}(r), B_{d_u}(r), Z(r)), \tag{22}$$

where

$$B_{d_x}(r) = \frac{1}{\Gamma(d_x+1)} \left\{ \int_0^r (r-t)^{d_x} \, dW_x(t) + \int_{-\infty}^0 [(r-t)^{d_x} - (-t)^{d_x}] \, dW_x(t) \right\},$$

$$B_{d_u}(r) = \frac{1}{\Gamma(d_u+1)} \left\{ \int_0^r (r-t)^{d_u} \, dW_u(t) + \int_{-\infty}^0 [(r-t)^{d_u} - (-t)^{d_u}] \, dW_u(t) \right\},$$

and

$$Z(r) = \frac{1}{\Gamma(d_x)\Gamma(d_u)} \times \int_{-\infty}^r \int_{-\infty}^r \left[\int_0^r (\tau - t_1)_+^{d_x-1} (\tau - t_2)_+^{d_u-1} \, d\tau \right] dW_x(t_1) \, dW_u(t_2).$$

Here $\xi_+^d = [\max(0, \xi)]^d$ *and* $(W_x(t), W_u(t))'$ *is a vector Brownian motion with variance matrix* $\Pi\Sigma\Pi'$.

Note that $B_{d_x}(r)$ and $B_{d_u}(r)$ are type I fractional Brownian motions (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 numerous authors, including Davidson and de Jong (2000) and Chung (2002). 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 is obtained by Chung (2002, Theorem 2) but under the stronger assumption that $d_x, d_u \in (0.25, 0.5)$. A careful inspection of Chung's proof reveals that his Theorem 2 holds as long as $d_x + d_u > \frac{1}{2}$.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

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

THEOREM 4. *Let Assumptions A1–A3 hold. Assume that* $k(x)$ *is a twice continuously differentiable function in* $\mathcal{K}$ *and* $M = bT$ *for* $b \in (0,1]$. *If* $d_x + d_u > \frac{1}{2}$, *then*

$$T^{1-d_x-d_u}(\hat{\beta} - \beta - \sigma_{xu}\sigma_x^{-2}) \Rightarrow \sigma_x^{-2}[Z(1) - B_{d_x}(1)B_{d_u}(1)],$$

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

$$\hat{t}_{\beta} - \sigma_{xu}\sigma_x^{-2}\hat{\sigma}_{\beta}^{-1} \Rightarrow (Z(1) - B_{d_x}(1)B_{d_u}(1)) \times \left(\int_0^1\int_0^1 -\frac{1}{b^2}\,k''\left(\frac{r-s}{b}\right) \times U(r)U(s)\,drds\right)^{-1/2}, \tag{23}$$

where

$$\sigma_x^2 = Ex_t^2, \qquad \sigma_{xu} = Ex_t u_t,$$

and

$$U(r) = Z(r) - rZ(1) - [B_{d_x}(r) - rB_{d_x}(1)]B_{d_u}(1) - B_{d_x}(1)[B_{d_u}(r) - rB_{d_u}(1)].$$

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 that follows. It should be noted that when $d_x > d_u$, β can be consistently estimated using a frequency domain approach even if $\sigma_{ux} \neq 0$ (cf. Robinson, 1994).

When $\sigma_{ux} = 0$, Theorem 4 reveals the convergence of the *t*-statistic when $M = bT$. In contrast, Tsay and Chung (2000) show 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 $\hat{\beta}$ is consistent, the R^2 converges to zero, and the Durbin–Watson (DW)-statistic does not approach zero (Tsay and Chung, 2000). The behaviors of $\hat{\beta}$, 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

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

finding 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 preceding theorem assumes that the kernel function is twice continuously differentiable. This excludes the widely used Bartlett kernel. The following theorem establishes the asymptotic distributions of $\hat{\beta}$, $\hat{\sigma}_\beta$, and the t-statistic when the Bartlett kernel is employed.

THEOREM 5. *Let Assumptions A1–A3 hold. If* $k(x) = (1 - |x|)1\{|x| \leq 1\}$, $M = bT$ *for* $b \in (0,1]$, *and* $d_x + d_u > \frac{1}{2}$, *then Theorem 4 holds with*

$$-1/b^2 \int_0^1 \int_0^1 k''((r-s)/b)U(r)U(s)\,drds \tag{24}$$

replaced by

$$\frac{2}{b}\int_0^1 U^2(r)\,dr - \frac{2}{b}\int_0^{1-b} U(r)U(r+b)\,dr. \tag{25}$$

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

$$T^{-(d_u+1/2)} \sum_{t=1}^{[\mathrm{Tr}]} u_t \Rightarrow B_{d_u}(r). \tag{26}$$

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

$$T^{-(d_u+3/2)} \sum_{t=1}^{[\mathrm{Tr}]} tu_t \Rightarrow rB_{d_u}(r) - \int_0^r B_{d_u}(s)\,ds. \tag{27}$$

Let

$$G(r) = \left(r - \frac{1}{2}\right)B_{d_u}(r) - \int_0^r B_{d_u}(s)\,ds - B_{d_u}(1)\left(\int_0^r \left(s - \frac{1}{2}\right)ds\right) - \left(6B_{d_u}(1) - 12\int_0^1 B_{d_u}(s)\,ds\right)\int_0^r \left(s - \frac{1}{2}\right)^2 ds. \tag{28}$$

Then we can prove the following theorem using (26) and (27) and arguments similar to the proof of Theorem 4. Details are omitted.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

THEOREM 6. *Let $x_t = t$ and u_t satisfy (26) with $d_u \in (0,\frac{1}{2})$. If $k(x)$ is a twice continuously differentiable function in $\mathcal{K}$ and $M = bT$ for $b \in (0,1]$, then*

$$T^{3/2-d_u}(\hat{\beta} - \beta) \Rightarrow 6B_{d_u}(1) - 12\int_0^1 B_{d_u}(s)\,ds, \tag{29}$$

$$T^{3-2d_u}\hat{\sigma}_\beta^2 \Rightarrow -\frac{144}{b^2}\int_0^1\int_0^1 k''\left(\frac{r-s}{b}\right)G(r)G(s)\,drds, \tag{30}$$

$$\hat{t}_\beta \Rightarrow \left(\frac{B_{d_u}(1)}{2} - \int_0^1 B_{d_u}(s)\right) \times \left(\int_0^1\int_0^1 -\frac{1}{b^2}\,k''\left(\frac{r-s}{b}\right)G(r)G(s)\,drds\right)^{-1/2}. \tag{31}$$

If $k(x) = (1 - |x|)1\{|x| \le 1\}$, then (29), (30), and (31) hold with

$$\int_0^1\int_0^1 -\frac{1}{b^2}\,k''\left(\frac{r-s}{b}\right)G(r)G(s)\,drds \tag{32}$$

replaced by

$$\frac{2}{b}\int_0^1 G^2(r)\,dr - \frac{2}{b}\int_0^{1-b} G(r)G(r+b)\,dr. \tag{33}$$

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 $\hat{t}_\beta$ are nonstandard. In this section, we use Monte Carlo simulations to approximate their probability densities. Our goal is to see how the new limiting distribution differs from the standard normal distribution and to what extent the use of $M = bT$ helps reduce the spurious effects.

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 $t \le 0$, 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, $0.05,\ldots,1$. When $b = 0$, $\hat{\Omega}$ becomes $T^{-1}\sum\hat{u}_t^2(x_t - \bar{x})^2$. We

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

use 5,000 replications and a sample size of 1,000. For the results reported here, 10,000 replications are employed.

We first examine 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 become 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 $\hat{\Omega}$ 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 $\hat{u}_t$. As a consequence, when $k(x) = 1$ and $b = 1$, $\hat{\Omega} = \sum_{j=-T+1}^{T-1} \hat{\Gamma}(j)$ is zero by construction.

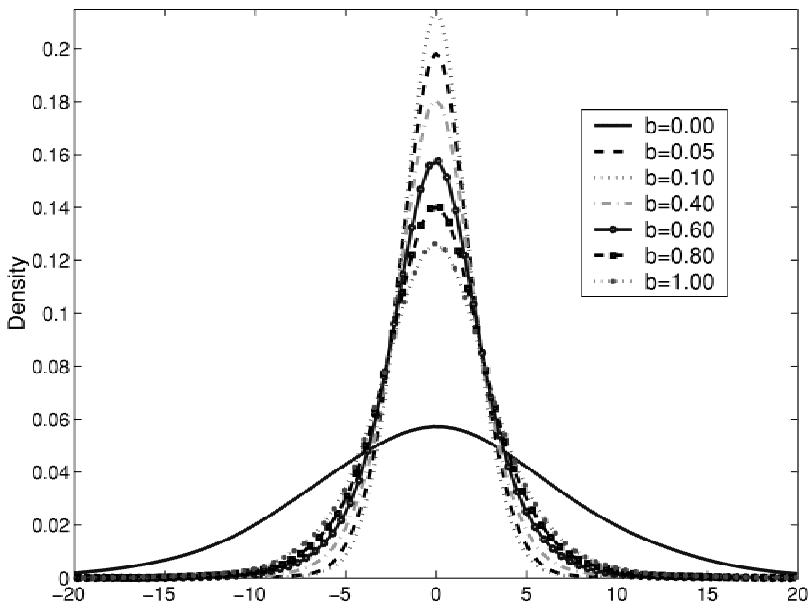


FIGURE 1. Kernel estimates of densities of $\hat{t}_\beta$ when $x_t \sim I(0.6)$ and $y_t \sim I(0.6)$.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

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 $\hat{\Omega}$ tends to bias toward zero when $\hat{\Omega}$ 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 being least dispersed when $b = 0.1$. When $d_x = d_u = 0.6$ and $b = 0.1$, the probability of $|\hat{t}_\beta| > 1.96$ 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 $T = 1{,}000$ 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%

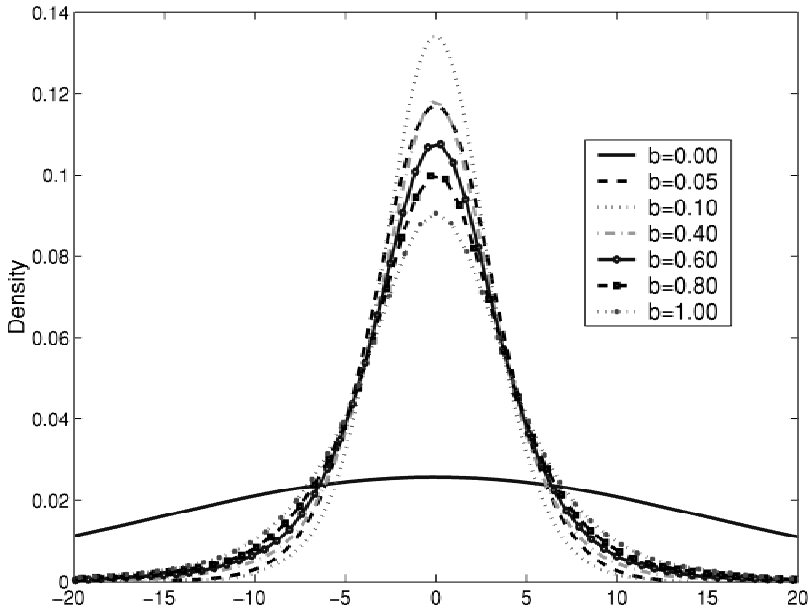


FIGURE 2. Kernel estimates of densities of $\hat{t}_\beta$ when $x_t = t$ and $y_t \sim I(0.6)$.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

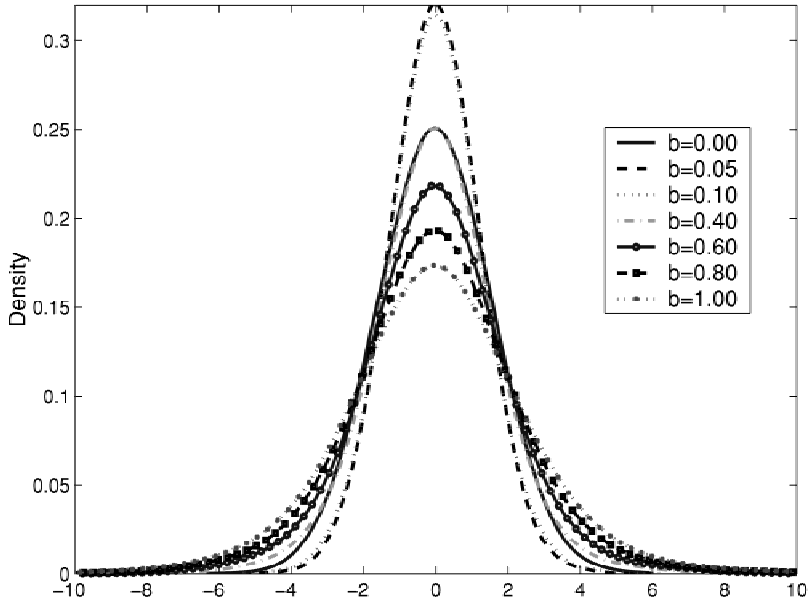


FIGURE 3. Kernel estimates of densities of $\hat{t}_\beta$ when $x_t \sim I(0.4)$ and $y_t \sim I(0.2)$.

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 $|\hat{t}_\beta| > 1.96$ is 11.70%, which is very close to 10%, the size of the test when $\hat{t}_\beta$ 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 investigates the asymptotic properties of the t-statistic in spurious regressions. It was shown that the usual t-statistic diverges because the OLS

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

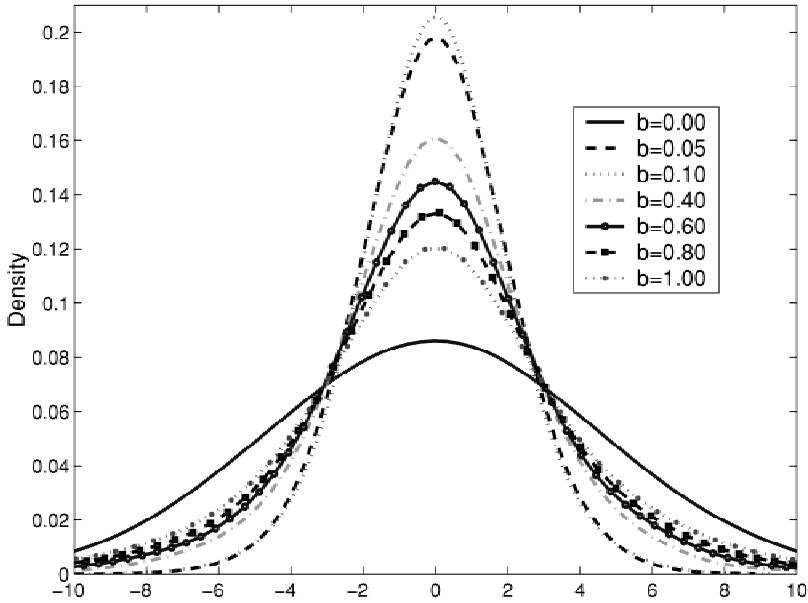


FIGURE 4. Kernel estimates of densities of $\hat{t}_\beta$ when $x_t = t$ and $y_t \sim I(0.3)$.

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 rise to spurious effects. If a large bandwidth ($M = bT$) is used, the convergence of the *t*-statistic can be restored.

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 rate 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 overrejection when

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

M is small and the underlying processes exhibit some persistence. The use of $M = bT$ has the potential to reduce this risk.

REFERENCES

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

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

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

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

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

Jansson, M. (2004) Autocorrelation robust tests with good size and power. *Econometrica*, forthcoming.

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

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

Kiefer, N.M. & T.J. Vogelsang (2003) A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. Working paper, Center for Analytic Economics, Cornell University.

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

Marinucci, D. & 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.

Marmol, F. & C. Velasco (2002) Trend stationarity versus long-range dependence in time series analysis. *Journal of Econometrics* 108, 25–42.

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., Y. Sun, & S. Jin (2003a) Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation. Cowles Foundation Discussion paper 1407, http://cowles.econ.yale.edu/P/au/d_phillips.htm.

Phillips, P.C.B., Y. Sun, & S. Jin (2003b) Long Run Variance Estimation with Steep Origin Kernels without Truncation. Cowles Foundation Discussion paper 1437, http://cowles.econ.yale.edu/P/au/d_phillips.htm.

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

Robinson, P.M. (1998) Inference-without-smoothing in the presence of nonparametric autocorrelation. *Econometrica* 66, 1163–1182.

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

Weinstock, R. (1957) On continuous differentiability. *American Mathematical Monthly* 64, 492.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

APPENDIX OF PROOFS

Proof of Theorem 1. Combining the FCLT with the continuous mapping theorem, we have

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

and

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

Hence

$$\begin{aligned} T^{d_x-d_u}(\hat{\beta}-\beta) &= \frac{T^{-d_x-d_u}\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u})}{T^{-2d_x}\sum_{t=1}^{T}(x_t-\bar{x})^2} \\ &\Rightarrow \left(\int_0^1 \tilde{V}_x(r)\tilde{V}_u(r)\,dr\right)\left(\int_0^1 \tilde{V}_x^2(r)\,dr\right)^{-1}. \end{aligned} \tag{A.3}$$

As a consequence,

$$\begin{aligned} &T^{-(2d_u-1)/2}\hat{u}_{[\mathrm{Tr}]} \\ &\quad = T^{-(2d_u-1)/2}(u_{[\mathrm{Tr}]}-\bar{u}) - (T^{d_x-d_u}(\hat{\beta}-\beta))(T^{-(2d_x-1)/2}(x_{[\mathrm{Tr}]}-\bar{x})) \Rightarrow \tilde{V}_{u,x}(r). \end{aligned}$$

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

$$\begin{aligned} &\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t-\bar{x}}{T^{d_x-1/2}}\right)^2\right)^{-2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\frac{x_t-\bar{x}}{T^{d_x-1/2}}\,\frac{\hat{u}_t}{T^{d_u-1/2}}\,k\left(\frac{t-\tau}{bT}\right)\frac{\hat{u}_\tau}{T^{d_u-1/2}}\,\frac{x_\tau-\bar{x}}{T^{d_x-1/2}} \\ &\quad \Rightarrow \left(\int_0^1 \tilde{V}_x^2(r)\,dr\right)^{-2}\int_0^1\int_0^1 \tilde{V}_x(r)\tilde{V}_{u,x}(r)k\left(\frac{r-s}{b}\right)\tilde{V}_{u,x}(s)\tilde{V}_x(s)\,drds, \end{aligned} \tag{A.4}$$

where the last line follows from the continuous mapping theorem. In view of (A.3) and (A.4), we have

$$\begin{aligned} \hat{t}_\beta &= \left(\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u})\right)\left(\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau-\bar{x})\right)^{-1/2} \\ &\Rightarrow \left(\int_0^1 \tilde{V}_x(r)\tilde{V}_u(r)\,dr\right)\left(\int_0^1\int_0^1 \tilde{V}_x(r)\tilde{V}_{u,x}(r)k\left(\frac{r-s}{b}\right)\tilde{V}_{u,x}(s)\tilde{V}_x(s)\,drds\right)^{-1/2}. \end{aligned} \tag{A.5}$$

This completes the proof of the theorem. ■

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

Proof of Lemma 3. The results of the lemma have been established in Chung (2002) under the stronger assumption that both d_x and d_u are greater than 0.25 and less than 0.5. A careful inspection reveals that his proof goes through without modification as long as $d_x + d_u > 0.5$. ■

Proof of Theorem 4. It follows from Corollary 1 in Chung (2002) that

$$\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} x_t^2 - (\bar{x})^2 = \sigma_x^2. \tag{A.6}$$

Combining (A.6) with Lemma 3 yields

$$\begin{aligned} &T^{1-d_x-d_u}(\hat{\beta} - \beta - \sigma_{xu}/\sigma_x^2) \\ &\quad = T^{-d_x-d_u}\left(\sum_{t=1}^{T}(x_t u_t - \sigma_{xu} - T\bar{x}\bar{u})\right)\left(T^{-1}\sum_{t=1}^{T} x_t^2 - (\bar{x})^2\right)^{-1} + o_p(1) \\ &\quad \Rightarrow \sigma_x^{-2}[Z(1) - B_{d_x}(1)B_{d_u}(1)]. \end{aligned} \tag{A.7}$$

We next consider the limiting distribution of

$$\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t - \bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau - \bar{x}). \tag{A.8}$$

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

$$\begin{aligned} T^{-d_x-d_u}S_T^v(r) &= T^{-d_x-d_u}\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})(y_t - \bar{y}) - T^{-d_x-d_u}\hat{\beta}\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})^2 \\ &= T^{-d_x-d_u}\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})(u_t - \bar{u}) - T^{-d_x-d_u}(\hat{\beta} - \beta)\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})^2 \\ &= T^{-d_x-d_u}\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})(u_t - \bar{u}) - T^{-d_x-d_u}r\sum_{t=1}^{T}(x_t u_t - T\bar{x}\bar{u})(1 + o_p(1)) \\ &\Rightarrow Z(r) - rZ(1) - [B_{d_x}(r) - rB_{d_x}(1)]B_{d_u}(1) - B_{d_x}(1)[B_{d_u}(r) - rB_{d_u}(1)] \\ &:= U(r), \end{aligned} \tag{A.9}$$

where we have used

$$\text{plim}_{T\to\infty}\left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)^{-1}\sum_{t=1}^{[\text{Tr}]}(x_t - \bar{x})^2 = r \tag{A.10}$$

uniformly in r.

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

Using summation by parts twice, we have

$$
\begin{aligned}
&T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau-\bar{x})\\
&\quad= T^{-2d_x-2d_u}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\left\{S_T^v(t/T)\left(2k\left(\frac{t-\tau}{bT}\right)-k\left(\frac{t-\tau-1}{bT}\right)-k\left(\frac{t-\tau+1}{bT}\right)\right)\right.\\
&\qquad\qquad\left.\times S_T^v(\tau/T)\right\}\\
&\qquad+T^{-2d_x-2d_u}S_T^v(1)\sum_{\tau=1}^{T-1}\left(k\left(\frac{T-\tau}{bT}\right)-k\left(\frac{T-\tau-1}{bT}\right)\right)S_T^v(\tau/T)\\
&\qquad+T^{-2d_x-2d_u}\left\{\sum_{t=1}^{T-1}S_T^v(t/T)\left(k\left(\frac{t-T}{bT}\right)-k\left(\frac{t-T+1}{bT}\right)\right)S_T^v(1)+S_T^v(1)S_T^v(1)\right\}\\
&\quad=\frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}T^{-d_x-d_u}S_T^v(t/T)T^2D_T\left(\frac{t-\tau}{bT}\right)T^{-d_x-d_u}S_T^v(\tau/T),
\end{aligned}
\tag{A.11}
$$

where

$$
D_T\left(\frac{t-\tau}{bT}\right)=2k\left(\frac{t-\tau}{bT}\right)-k\left(\frac{t-\tau-1}{bT}\right)-k\left(\frac{t-\tau+1}{bT}\right)
\tag{A.12}
$$

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

$$
\lim_{T\to\infty}T^2D_T\left(\frac{t-\tau}{bT}\right)=-\frac{1}{b^2}k''\left(\frac{r-s}{b}\right).
\tag{A.13}
$$

Because $k(\cdot)$ is twice continuously differentiable, the preceding convergence is uniform in r and s. In other words, for any given $\varepsilon>0$, there exists a positive Δ that is independent of r and s such that

$$
\left|T^2D_T\left(\frac{t-\tau}{bT}\right)+\frac{1}{b^2}k''\left(\frac{r-s}{b}\right)\right|<\varepsilon
$$

whenever $\|(t/T,\tau/T)-(r,s)\|<\Delta$ for all (r,s) in $[0,1]\times[0,1]$. For a proof of the uniformity, see Weinstock (1957).

Combining (A.9), (A.11), and (A.13) and invoking the continuous mapping theorem, we get

$$
\begin{aligned}
&T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau-\bar{x})\\
&\quad\Rightarrow-\frac{1}{b^2}\int_0^1\int_0^1U(r)k''\left(\frac{r-s}{b}\right)U(s)\,drds.
\end{aligned}
\tag{A.14}
$$

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press

Therefore

$$
\begin{aligned}
&T^{2-2d_x-2d_u}\hat{\sigma}_\beta^2 \\
&\quad = \left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^{-2} T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau-\bar{x}) \\
&\quad \Rightarrow -\sigma_x^{-4}\frac{1}{b^2}\int_0^1\int_0^1 U(r)k''\left(\frac{r-s}{b}\right)U(s)\,drds. \qquad \textbf{(A.15)}
\end{aligned}
$$

Combining (A.7) with (A.15) 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,

$$
D_T\left(\frac{t-\tau}{bT}\right) = \begin{cases} \dfrac{2}{bT}, & \text{if } |t-\tau| = 0 \\ -\dfrac{1}{bT}, & \text{if } |t-\tau| = [bT] \\ 0, & \text{otherwise.} \end{cases} \qquad \textbf{(A.16)}
$$

Hence

$$
\begin{aligned}
&T^{-2d_x-2d_u}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(x_t-\bar{x})\hat{u}_t k\left(\frac{t-\tau}{bT}\right)\hat{u}_\tau(x_\tau-\bar{x}) \\
&\quad = T^{-2d_x-2d_u}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} S_T^v(t/T)D_T\left(\frac{t-\tau}{bT}\right)S_T^v(\tau/T), \\
&\quad = \frac{2}{bT}\sum_{\tau=1}^{T-1} T^{-2d_x-2d_u}S_T^v(t/T)S_T^v(t/T) \\
&\qquad - \frac{2}{bT}\sum_{\tau=1}^{T-[bT]-1} T^{-2d_x-2d_u}S_T^v(\tau/T)S_T^v((\tau+[bT])/T) \\
&\quad \Rightarrow \frac{2}{b}\int_0^1 U^2(r)\,dr - \frac{2}{b}\int_0^{1-b} U(r)U(r+b)\,dr. \qquad \textbf{(A.17)}
\end{aligned}
$$

The theorem now follows from (A.17) and the steps in the proof of Theorem 4. ■

https://doi.org/10.1017/S0266466604205072 Published online by Cambridge University Press