Asymptotic properties of spurious regression and random walks with generalized drifts

John W. Dennis* and Kaiji Motegi†

IDA Kobe University

February 1, 2022

Abstract

This paper investigates the spurious regression where each of the regressand y and the regressor x follows a random walk with generalized drift. The drift specification includes a zero, nonzero local, and nonzero constant drift as special cases. This framework is a substantial extension of the existing literature, in which both y and x have zero or constant drifts. We derive the order of convergence or divergence of the estimated slope coefficient $\hat{\beta}$ and the squared t-statistic $\hat{t}^2_{\beta}$, as well as their asymptotic distributions. We find that $\hat{\beta}$ may converge, diverge, or neither depending on the drift specification. Further, the asymptotic distribution of the scaled $\hat{\beta}$ takes on various interesting shapes such as a bimodal and asymmetric distribution. We also reveal that $\hat{t}^2_{\beta}$ diverges at different rates across cases.

JEL codes: C22, C51.

MSC2020 codes: 62M10, 91B84.

Keywords: bimodal distribution, Brownian motion, functional central limit theorem, non-stationary time series, Type I Error.

Word count: 10000 words.

*Institute for Defense Analyses (IDA). Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of IDA. No funding was provided by IDA or an affiliated organization for this paper. E-mail: jay.dennis@alumni.unc.edu

†*Corresponding author*. Graduate School of Economics, Kobe University. 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. E-mail: motegi@econ.kobe-u.ac.jp

1 Introduction

Spurious regression has been well studied since its description by Granger and Newbold (1974). Phillips (1986) studies the case resulting when one of two independent random walks without drift is regressed onto the other, and proves that the estimated slope coefficient $\hat{\beta}$ from the regression does not converge in probability to zero; instead, he shows it to be $O_p(1)$.[1] The spurious regression for two independent random walks with drift is later studied by Entorf (1997), who finds that $\hat{\beta}$ converges in probability to the ratio of the nonzero drift terms at rate $n^{1/2}$. Those results have been a well-known caveat to the regression analysis of non-stationary time series.[2]

Note that *neither* the regressand y *nor* the regressor x has a drift term in Phillips (1986), while *both* of them have drift terms in Entorf (1997). This contrast naturally poses a question on intermediate cases between them: what are the consequence of spurious regression when *either* y *or* x has a drift term? Further, we seek an asymptotic theory that unifies these cases; we pursue this by introducing *generalized drifts* that include zero, nonzero local, and nonzero constant drifts as special cases. Most of these scenarios have been unexplored in the literature, and this paper fills that gap.

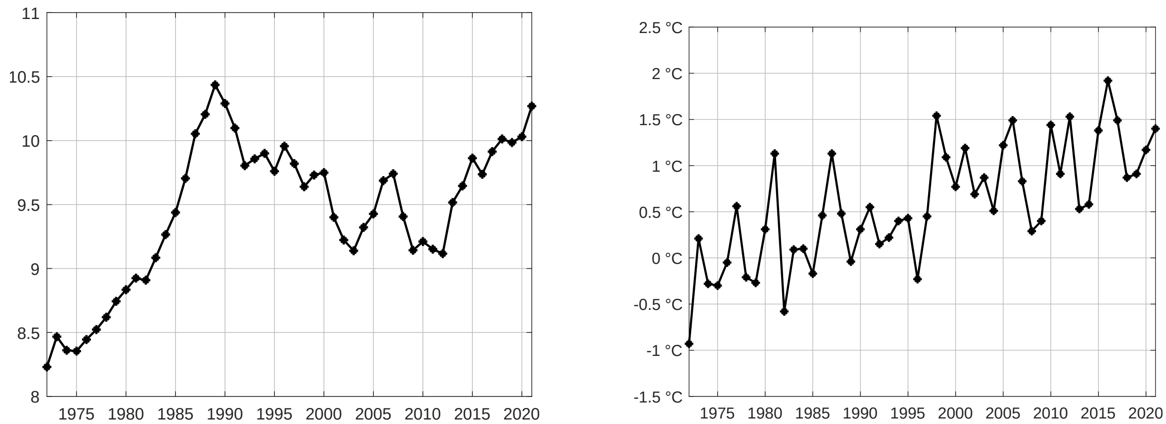
Characterizing the consequence of the generalized drift is motivated from not only theoretical but also empirical perspectives. In empirical studies, it is often hard to judge if variables of interest have zero or nonzero drifts. Here we present such a subtle example in order to illustrate the practical relevance of the generalized drift. Suppose that the log of the annual average of Nikkei 225 is regressed onto the annual land temperature anomalies of North America, where the sample period is 1972 through 2021 ($n = 50$ years); the two series are plotted in Figure 1. This is hypothetically a nonsense regression, since there is not a theoretical link between the Japanese stock price index and the North American temperature. The result of the regression, however, points to a significantly positive impact of the temperature on Nikkei. This is a common symptom of spurious regression caused by the non-stationarity of the two variables and the resulting residual; the result of the regression is summarized in Figure 2.

The exact consequence of the spurious regression depends on the drift types of y and x.

[1] Ernst, Shepp, and Wyner (2017) examine a related problem.

[2] Ventosa-Santaulària (2009) provides a survey of spurious regression. Vigen (2015) presents a number of empirical examples of spurious correlations.

Figure 1: The log of Nikkei 225 and the land temperature anomalies of North America



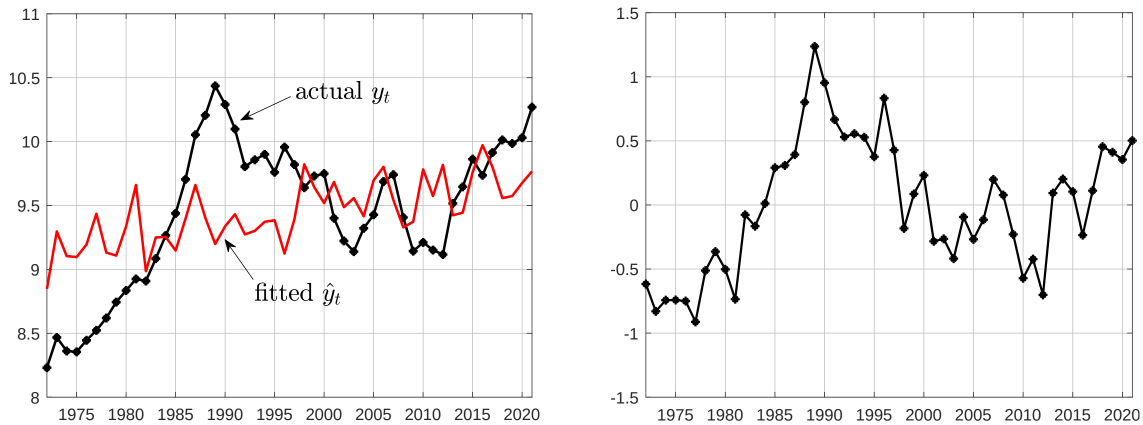
$y_t = \ln N225_t$ $\qquad$ $x_t = Temp_t$

$N225_t$ is the average value of Nikkei 225 at year t (measured in Japanese yen). $Temp_t$ is the land temperature anomalies of North America at year t (measured in Celsius); the anomalies are with respect to the 1910 to 2000 average. The sample period is 1972–2021 ($n = 50$ years). The data sources are described in Data Availability Statement of this paper.

In many empirical applications including the present one, it is hard to judge if y and x have zero or nonzero drifts (Figure 1). For Nikkei, there appears to be a strong positive drift until 1989, but the drift seems to have vanished afterward. For the temperature, there appears to be a weak positive drift throughout the sample period, but the magnitude of the drift seems rather small relative to the noise component. The generalized drift specification covers these fuzzy cases, which is the practical motivation of this study. Indeed, we reveal that the drift type has a crucial impact on the asymptotic properties of the estimated slope coefficient and the associated t-statistic.

Given the generalized drift, we derive the order of asymptotic convergence or divergence of the estimated slope coefficient $\hat{\beta}$ and the squared t-statistic $\hat{t}_{\beta}^2$, as well as their asymptotic distributions. We find that $\hat{\beta}$ may converge, diverge, or neither depending on the drift specification. Further, the asymptotic distribution of the properly scaled $\hat{\beta}$ takes on various interesting shapes such as the normal distribution, a unimodal and asymmetric distribution, a bimodal and symmetric distribution, and a bimodal and asymmetric distribution. We also reveal that $\hat{t}_{\beta}^2$ diverges for all cases considered, and the rate of divergence differs across cases. The asymptotic divergence of $\hat{t}_{\beta}^2$ implies that the probability of making Type I Error (i.e.,

Figure 2: Actual y_t, fitted $\hat{y}_t$, and residual $\hat{u}_t$ in the $N225$ vs. $Temp$ regression



Actual y_t and fitted $\hat{y}_t$ | Residual $\hat{u}_t$

The regression model is specified as $y_t = \alpha + \beta x_t + u_t$, where $y_t = \ln N225_t$ (i.e., the log of Nikkei 225) and $x_t = Temp_t$ (i.e., the land temperature anomalies of North America). The fitted value is given by $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$, where $\hat{\alpha} = 9.215$ and $\hat{\beta} = 0.395$ are the least squares estimates for α and β, respectively. The t-statistic for β is $\hat{t}_\beta = 3.352$; the standard error is 0.118; the p-value is 0.001. Hence, the zero hypothesis of β is rejected at any conventional significance level.

rejecting the correct null hypothesis $H_0 : \beta = 0$) approaches 1, a common symptom of the spurious regression.

The rest of this paper is organized as follows. The notation and analytical framework are set up in Section 2. The asymptotic properties of $\hat{\beta}$ are derived in Section 3. The asymptotic properties of $\hat{t}^2_\beta$ are derived in Section 4. Brief conclusions are provided in Section 5. Omitted proofs and technical discussions are presented in Appendices. Further details are provided separately in the supplemental material.

2 Notation and set-up

Suppose that the data generating process (DGP) is given by

$$y_t = d_{yn} + y_{t-1} + \epsilon_{yt}, \qquad x_t = d_{xn} + x_{t-1} + \epsilon_{xt}, \tag{1}$$

where the generalized drifts (d_{yn}, d_{xn}) are specified as follows:

$$
\begin{aligned}
d_{yn} &= d_y n^{-\delta_y}, \quad d_y \neq 0, \quad \delta_y \in [0, \infty), \\
d_{xn} &= d_x n^{-\delta_x}, \quad d_x \neq 0, \quad \delta_x \in [0, \infty).
\end{aligned}
\tag{2}
$$

The drifts (d_{yn}, d_{xn}) can take various forms depending on the power exponents (δ_y, δ_x). When $\delta_y \to \infty$ and $\delta_x \to \infty$, (2) reduces to the zero drift case of Phillips (1986): $d_{yn} \to 0$ and $d_{xn} \to 0$ for fixed sample size $n \in \mathbb{N}$. When $\delta_y = \delta_x = 0$, (2) reduces to the constant drift case of Entorf (1997): $d_{yn} = d_y \neq 0$ and $d_{xn} = d_x \neq 0$. Intermediate cases that have been unexplored in the literature include $\delta_y \to \infty$ and $\delta_x = 0$ (i.e., y has the zero drift and x has the constant drift $d_x \neq 0$) and vice versa. Furthermore, by setting $\delta_y \in (0, \infty)$, y is allowed to have a nonzero local drift which vanishes as the sample size grows (i.e., $d_{yn} \to 0$ as $n \to \infty$); the same goes for x too. The local drift scenario is new to the existing literature, which is another innovation of this paper.

Define:

$$
\boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix}, \quad \boldsymbol{s}_t = \begin{bmatrix} s_{xt} \\ s_{yt} \end{bmatrix} = \begin{bmatrix} \sum_{\tau=1}^{t} \epsilon_{x\tau} \\ \sum_{\tau=1}^{t} \epsilon_{y\tau} \end{bmatrix}, \quad \boldsymbol{s}_0 = \mathbf{0}, \quad y_0 = x_0 = 0. \tag{3}
$$

Impose the following assumptions.

Assumption 1. *(i)* $\mathrm{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}$ *for all* $t \in \mathbb{N}$; *(ii)* $\sup_{t\in\mathbb{N}} \max\{\mathrm{E}|\epsilon_{xt}|^{2+\delta}, \mathrm{E}|\epsilon_{yt}|^{2+\delta}\} < \infty$ *for some* $\delta > 0$; *(iii)* $\lim_{n\to\infty} n^{-1}\mathrm{E}(\boldsymbol{s}_n \boldsymbol{s}_n^\top)$ *exists and is positive definite; (iv)* $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{N}}$ *is* α*-mixing with some mixing coefficient* $\{\alpha_m\}_{m\in\mathbb{N}}$ *satisfying* $\sum_{m=1}^{\infty} \alpha_m^{\varphi} < \infty$ *for some* $\varphi > 0$; *(v)* $\{\epsilon_{xt}\}$ *and* $\{\epsilon_{yt}\}$ *are mutually independent sequences.*

Assumption 1 is essentially identical to the assumptions of Phillips (1986). Conditions (*i*)-(*iv*) are mild assumptions which admit a functional central limit theorem (FCLT). The mixing coefficient $\{\alpha_m\}_{m\in\mathbb{N}}$ appearing in condition (*iv*) measures the dependence between $\{\boldsymbol{\epsilon}_\tau\}_{\tau=1}^{t}$ and $\{\boldsymbol{\epsilon}_\tau\}_{\tau=t+m}^{\infty}$ (see, e.g., Davidson, 1994, Bradley, 2005, for details on the α-mixing condition). Condition (*v*), which implies the mutual independence between x and y, is imposed to investigate the properties of spurious regression.

The regression model is specified as

$$
y_t = \alpha + \beta x_t + u_t, \qquad t \in \{1, \dots, n\}, \tag{4}
$$

where (α, β) are parameters to estimate and u_t is the error term.[3] Equation (4) is a spurious regression under Assumption 1(v). Let $(\hat{\alpha}, \hat{\beta})$ be the least squares estimators of (α, β); let $\hat{t}_\beta$ be the conventional t-statistic with respect to $H_0 : \beta = 0$. As is well known,

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \tag{5}$$

$$\hat{t}_\beta = \frac{n^{1/2}\hat{\beta}}{\left(\hat{\sigma}_u^2/\hat{V}_x\right)^{1/2}}, \tag{6}$$

where

$$\bar{y} = n^{-1}\sum_{t=1}^{n} y_t, \quad \bar{x} = n^{-1}\sum_{t=1}^{n} x_t, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{y}_t = \hat{\alpha} + \hat{\beta}x_t,$$
$$\hat{\sigma}_u^2 = n^{-1}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2, \quad \hat{V}_x = n^{-1}\sum_{t=1}^{n}(x_t - \bar{x})^2. \tag{7}$$

We derive the asymptotic properties of $\hat{\beta}$ and $\hat{t}_\beta^2$ in Sections 3 and 4, respectively.

3 Asymptotic properties of the estimated slope

In this section, we derive the asymptotic properties of the estimated slope parameter $\hat{\beta}$ under the generalized drifts (2). Equation (5) can be rewritten as

$$\hat{\beta} = \frac{\lambda_n^{xy}}{\lambda_n^{xx}}, \tag{8}$$

where

$$\lambda_n^{xx} = n^{-3}\sum_{t=1}^{n}(x_t - \bar{x})^2, \qquad \lambda_n^{xy} = n^{-3}\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y}). \tag{9}$$

To express the rate of convergence or divergence of $\hat{\beta}$, we introduce the notion of boundedness in probability. For a stochastic sequence $\{X_n\}_{n\in\mathbb{N}}$ and $k \in \mathbb{R}$, we write $X_n = O_p(n^k)$ if for all $\epsilon > 0$, there exists $B_\epsilon < \infty$ such that $\Pr(n^{-k}|X_n| > B_\epsilon) < \epsilon$. Further, we express

[3] In the supplemental material, we perform a simulation-based study of extended regression models which contain time trends as extra regressors. We find that the symptom of spurious regression is present even when the time trends are included in the model.

asymptotically negligible terms as follows: $X_n = o_p(n^k)$ if $\Pr(n^{-k}|X_n| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$. In what follows, we use the notations of O_p and o_p in order to simplify equations.

The asymptotic approximations of λ_n^{xx} and λ_n^{xy} appearing in (9) are as follows.

Lemma 1. *Let the DGP* (1) *and Assumption 1 hold, then*

$$\lambda_n^{xx} = \frac{1}{12} n^{-2\delta_x} d_x^2 + 2n^{-(1/2+\delta_x)} d_x \hat{\xi}_x + n^{-1} \hat{B}_{xx} + o_p(n^{-1}), \tag{10}$$

$$\lambda_n^{xy} = \frac{1}{12} n^{-(\delta_x+\delta_y)} d_x d_y + n^{-(1/2+\delta_y)} d_y \hat{\xi}_x + n^{-(1/2+\delta_x)} d_x \hat{\xi}_y + n^{-1} \hat{B}_{xy} + o_p(n^{-1}), \tag{11}$$

where

$$\hat{\xi}_x = n^{-5/2} \sum_{t=1}^{n} t s_{xt} - \frac{1}{2} n^{-3/2} \sum_{t=1}^{n} s_{xt}, \qquad \hat{\xi}_y = n^{-5/2} \sum_{t=1}^{n} t s_{yt} - \frac{1}{2} n^{-3/2} \sum_{t=1}^{n} s_{yt}, \tag{12}$$

$$\hat{B}_{xx} = n^{-2} \sum_{t=1}^{n} s_{xt}^2 - \left(n^{-3/2} \sum_{t=1}^{n} s_{xt} \right)^2, \tag{13}$$

$$\hat{B}_{xy} = n^{-2} \sum_{t=1}^{n} s_{xt} s_{yt} - \left(n^{-3/2} \sum_{t=1}^{n} s_{xt} \right) \left(n^{-3/2} \sum_{t=1}^{n} s_{yt} \right),$$

and s_{xt} and s_{yt} are defined in (3).

See Appendix A.1 for the proof of Lemma 1.

It is clear from (8) and Lemma 1 that the asymptotic properties of $\hat{\beta}$ depend on those of $(\hat{\xi}_x, \hat{\xi}_y, \hat{B}_{xx}, \hat{B}_{xy})$, and the latter are well known in the time series theory. Let $\{W_x(t)\}_{t\geq 0}$ and $\{W_y(t)\}_{t\geq 0}$ be standard Brownian motions that are mutually independent. Define

$$\xi_x = \int_0^1 t\sigma_{\epsilon x} W_x(t) dt - \frac{1}{2} \int_0^1 \sigma_{\epsilon x} W_x(t) dt, \qquad \xi_y = \int_0^1 t\sigma_{\epsilon y} W_y(t) dt - \frac{1}{2} \int_0^1 \sigma_{\epsilon y} W_y(t) dt,$$

$$B_{xx} = \int_0^1 \sigma_{\epsilon x}^2 W_x^2(t) dt - \left\{ \int_0^1 \sigma_{\epsilon x} W_x(t) dt \right\}^2, \tag{14}$$

$$B_{xy} = \int_0^1 \sigma_{\epsilon x} \sigma_{\epsilon y} W_x(t) W_y(t) dt - \int_0^1 \sigma_{\epsilon x} W_x(t) dt \int_0^1 \sigma_{\epsilon y} W_y(t) dt,$$

where $\sigma_{\epsilon x}^2 = \lim_{n\to\infty} n^{-1} \mathrm{E}(s_{xn}^2)$ and $\sigma_{\epsilon y}^2 = \lim_{n\to\infty} n^{-1} \mathrm{E}(s_{yn}^2)$ are the long-run variances of ϵ_x and ϵ_y, respectively; see Phillips (1986) for a discussion.

Assumptions 1(i)-(iv) are sufficient for the regularity assumptions of Entorf (1997), and they can similarly be found in Park and Phillips (1988, 1989) and Sims, Stock, and Watson

(1990). In particular, these regularity assumptions allow weak convergence of the partial sum process of the model errors to a Brownian motion via FCLT, as first described by Donsker (1951) and later elaborated by Phillips (1986) and many others (see Billingsley, 1968, for a classic textbook treatment). Lemma 2 restates some of these results, where the symbol $\Rightarrow$ denotes weak convergence.

Lemma 2. *Let Assumptions 1(i)-(iv) hold, then* $\hat{\xi}_x \Rightarrow \xi_x$, $\hat{\xi}_y \Rightarrow \xi_y$, $\hat{B}_{xx} \Rightarrow B_{xx}$, *and* $\hat{B}_{xy} \Rightarrow B_{xy}$.

Refer to Billingsley (1968) and Phillips (1986) for proofs of Lemma 2.

Durlauf and Phillips (1988) show that

$$\begin{bmatrix} \xi_x \\ \xi_y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon x}^2/120 & 0 \\ 0 & \sigma_{\epsilon y}^2/120 \end{bmatrix} \right). \tag{15}$$

Hence, by Lemma 2, it follows that

$$\begin{bmatrix} \hat{\xi}_x \\ \hat{\xi}_y \end{bmatrix} \Rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon x}^2/120 & 0 \\ 0 & \sigma_{\epsilon y}^2/120 \end{bmatrix} \right). \tag{16}$$

Equation (16) is useful for simplifying and interpreting the asymptotic distributions of $\hat{\beta}$.

In view of (10)-(11), the dominant terms of $(\lambda_n^{xx}, \lambda_n^{xy})$ alter depending on (δ_x, δ_y). The key threshold is $\delta_x = 1/2$, at which $2\delta_x = 1/2 + \delta_x = 1$; the same insight holds for δ_y as well. Hence, we prepare the following cases and terminologies:

Y1 (weak drift in y): $\delta_y \in (1/2, \infty)$.

Y2 (semi-strong drift in y): $\delta_y = 1/2$.

Y3 (strong drift in y): $\delta_y \in [0, 1/2)$.

X1 (weak drift in x): $\delta_x \in (1/2, \infty)$.

X2 (semi-strong drift in x): $\delta_x = 1/2$.

X3 (strong drift in x): $\delta_x \in [0, 1/2)$.

The *weak* drift vanishes at a faster rate than $n^{1/2}$; the *semi-strong* drift vanishes at rate $n^{1/2}$; the *strong* drift vanishes at a slower rate than $n^{1/2}$. The zero drift is the limit case of the weak drift, whereas the constant drift is the boundary case of the strong drift. We write Case Y1X1, for example, to express the combination of Cases Y1 and X1. The framework of Phillips (1986), in which both y and x have zero drifts, is the limit case of Y1X1 with $\delta_y \to \infty$ and $\delta_x \to \infty$. The framework of Entorf (1997), in which both y and x have nonzero constant drifts, is the boundary case of Y3X3 with $\delta_y = \delta_x = 0$. These terminologies help us interpret the impact of the drift specification on the asymptotic properties of $\hat{\beta}$.

For each of Cases X1-X3, the relevant terms of (10) are given as follows:

$$\lambda_n^{xx} = \begin{cases} n^{-1}\hat{B}_{xx} + o_p(n^{-1}) & \text{(Case X1)}, \\ n^{-1}\hat{Z}_{xx} + o_p(n^{-1}) & \text{(Case X2)}, \\ \frac{1}{12}n^{-2\delta_x}d_x^2 + 2n^{-(1/2+\delta_x)}d_x\hat{\xi}_x + n^{-1}\hat{B}_{xx} + o_p(n^{-1}) & \text{(Case X3)}, \end{cases} \tag{17}$$

where

$$\hat{Z}_{xx} = \frac{1}{12}d_x^2 + 2d_x\hat{\xi}_x + \hat{B}_{xx}.$$

Similarly, the relevant terms of (11) are given as follows:

$$\lambda_n^{xy} = \begin{cases} n^{-1}\hat{B}_{xy} + o_p(n^{-1}) & \text{(Case Y1X1)}, \\ n^{-1}(d_x\hat{\xi}_y + \hat{B}_{xy}) + o_p(n^{-1}) & \text{(Case Y1X2)}, \\ n^{-(1/2+\delta_x)}d_x\hat{\xi}_y + o_p\left\{n^{-(1/2+\delta_x)}\right\} & \text{(Case Y1X3)}, \\ n^{-1}(d_y\hat{\xi}_x + \hat{B}_{xy}) + o_p(n^{-1}) & \text{(Case Y2X1)}, \\ n^{-1}\hat{Z}_{xy} + o_p(n^{-1}) & \text{(Case Y2X2)}, \\ n^{-(1/2+\delta_x)}d_x(\frac{1}{12}d_y + \hat{\xi}_y) + o_p\left\{n^{-(1/2+\delta_x)}\right\} & \text{(Case Y2X3)}, \\ n^{-(1/2+\delta_y)}d_y\hat{\xi}_x + o_p\left\{n^{-(1/2+\delta_y)}\right\} & \text{(Case Y3X1)}, \\ n^{-(1/2+\delta_y)}d_y(\frac{1}{12}d_x + \hat{\xi}_x) + o_p\left\{n^{-(1/2+\delta_y)}\right\} & \text{(Case Y3X2)}, \\ \frac{1}{12}n^{-(\delta_x+\delta_y)}d_xd_y + n^{-(1/2+\delta_y)}d_y\hat{\xi}_x + n^{-(1/2+\delta_x)}d_x\hat{\xi}_y & \\ \qquad +n^{-1}\hat{B}_{xy} + o_p(n^{-1}) & \text{(Case Y3X3)}, \end{cases} \tag{18}$$

where

$$\hat{Z}_{xy} = \frac{1}{12}d_xd_y + d_y\hat{\xi}_x + d_x\hat{\xi}_y + \hat{B}_{xy}.$$

By Lemma 2, we have that $\hat{Z}_{xx} \Rightarrow Z_{xx}$ and $\hat{Z}_{xy} \Rightarrow Z_{xy}$, where

$$Z_{xx} = \frac{1}{12}d_x^2 + 2d_x\xi_x + B_{xx}, \qquad Z_{xy} = \frac{1}{12}d_xd_y + d_y\xi_x + d_x\xi_y + B_{xy}.$$

Case Y3X3 requires extra caution when computing the ratio $\hat{\beta} = \lambda_n^{xy}/\lambda_n^{xx}$. As shown in

(17)-(18), the dominant term of λ_n^{xx} is of order $n^{-2\delta_x}$, and the dominant term of λ_n^{xy} is of order $n^{-(\delta_x+\delta_y)}$. The relative magnitude of these terms depends on the values of $\delta_x \in [0, 1/2)$ and $\delta_y \in [0, 1/2)$. Hence, Case Y3X3 should be divided into three subcases:

$$\textbf{Y3X3}(i)\textbf{:}\ \ \delta_y < \delta_x; \qquad \textbf{Y3X3}(ii)\textbf{:}\ \ \delta_y > \delta_x; \qquad \textbf{Y3X3}(iii)\textbf{:}\ \ \delta_y = \delta_x. \tag{19}$$

In Case (i), y has the stronger drift than x; in Case (ii), x has the stronger drift than y; in Case (iii), the drift in y is as strong as the drift in x. The constant drift scenario of Entorf (1997) is the boundary case of (iii) with $\delta_y = \delta_x = 0$.

Substitute (17) and (18) into (8) and then use Lemma 2 to characterize the asymptotic properties of $\hat{\beta}$.

Theorem 3. *Let the DGP* (1) *and Assumption 1 hold, then for each drift specification, the order of stochastic convergence or divergence of $\hat{\beta}$ and the asymptotic distribution of scaled $\hat{\beta}$ are characterized in Table 1.*

See Appendix A.2 for the proof of Theorem 3.

The asymptotic properties of $\hat{\beta}$ are dramatically different across cases (Table 1). First, $\hat{\beta} = O_p(1)$ (i.e., $\hat{\beta}$ neither diverges nor converges) *if and only if* neither y nor x has a strong drift; see Cases Y1X1, Y1X2, Y2X1, and Y2X2. This result is a substantial extension of the classical result of Phillips (1986), which states that $\hat{\beta} = O_p(1)$ if both y and x have zero drifts. Second, $\hat{\beta}$ diverges in Cases Y3X1, Y3X2, and Y3X3(i). In these cases, y has a strong drift (i.e., $\delta_y < 1/2$) and the drift in y is stronger than the drift in x (i.e., $\delta_y < \delta_x$). Third, $\hat{\beta}$ converges to 0 in Cases Y1X3, Y2X3, and Y3X3(ii). In these cases, x has a strong drift and the drift in x is stronger than the drift in y. Fourth, $\hat{\beta}$ converges to $d_y/d_x \neq 0$ at rate $n^{1/2-\delta_x}$ in Case Y3X3(iii), where $\delta_y = \delta_x < 1/2$. If we further impose $\delta_x = 0$, then our result reduces to Entorf's (1997) well-known result that $\hat{\beta} - d_y/d_x = O_p(n^{-1/2})$.

To better interpret the divergence versus convergence of $\hat{\beta}$, take Case Y2X2 as a benchmark, where both y and x have the semi-strong drifts and consequently $\hat{\beta} = O_p(1)$. An interesting contrast is that $\hat{\beta}$ *converges* to 0 when the drift in x becomes stronger (Case Y2X3), while $\hat{\beta}$ *diverges* when the drift in y becomes stronger (Case Y3X2). The former result suggests that the spurious regression is *mitigated* when x has the stronger drift than y, since the error term u has a relatively weak drift under $\beta = 0$. The latter result suggests that the spurious regression is *exacerbated* when y has the stronger drift than x, since u has

Table 1: Asymptotic properties of the least squares estimator $\hat{\beta}$

Case	δ_y	δ_x	Subcase	Order of $\hat{\beta}$	Asy. dist. of scaled $\hat{\beta}$
Y1X1	$(1/2, \infty)$	$(1/2, \infty)$	-	$\hat{\beta} = O_p(1)$	$\frac{B_{xy}}{B_{xx}}$
Y1X2	$(1/2, \infty)$	$1/2$	-	$\hat{\beta} = O_p(1)$	$\frac{d_x \xi_y + B_{xy}}{Z_{xx}}$
Y1X3	$(1/2, \infty)$	$[0, 1/2)$	-	$\hat{\beta} = O_p\left\{n^{-(1/2-\delta_x)}\right\}$	$\mathcal{N}\left(0,\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right)$
Y2X1	$1/2$	$(1/2, \infty)$	-	$\hat{\beta} = O_p(1)$	$\frac{d_y \xi_x + B_{xy}}{B_{xx}}$
Y2X2	$1/2$	$1/2$	-	$\hat{\beta} = O_p(1)$	$\frac{Z_{xy}}{Z_{xx}}$
Y2X3	$1/2$	$[0, 1/2)$	-	$\hat{\beta} = O_p\left\{n^{-(1/2-\delta_x)}\right\}$	$\mathcal{N}\left(\frac{d_y}{d_x},\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right)$
Y3X1	$[0, 1/2)$	$(1/2, \infty)$	-	$\hat{\beta} = O_p\left(n^{1/2-\delta_y}\right)$	$\frac{d_y \xi_x}{B_{xx}}$
Y3X2	$[0, 1/2)$	$1/2$	-	$\hat{\beta} = O_p\left(n^{1/2-\delta_y}\right)$	$\frac{d_y(\frac{1}{12}d_x + \xi_x)}{Z_{xx}}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(i) $\delta_y < \delta_x$	$\hat{\beta} = O_p\left(n^{\delta_x-\delta_y}\right)$	$\frac{d_y}{d_x}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(ii) $\delta_y > \delta_x$	$\hat{\beta} = O_p\left\{n^{-(\delta_y-\delta_x)}\right\}$	$\frac{d_y}{d_x}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(iii) $\delta_y = \delta_x$	$\hat{\beta} - \frac{d_y}{d_x} = O_p\left\{n^{-(1/2-\delta_x)}\right\}$	$\mathcal{N}\left\{0,\ \frac{6d_y^2}{5d_x^2}\left(\frac{\sigma_{\epsilon x}^2}{d_x^2} + \frac{\sigma_{\epsilon y}^2}{d_y^2}\right)\right\}$

DGP: $y_t = d_y n^{-\delta_y} + y_{t-1} + \epsilon_{yt}$ and $x_t = d_x n^{-\delta_x} + x_{t-1} + \epsilon_{xt}$, where $d_y \neq 0$ and $d_x \neq 0$. Model: $y_t = \alpha + \beta x_t + u_t$. $\hat{\beta}$ is the least squares estimator for β. This table summarizes the order of stochastic convergence or divergence of $\hat{\beta}$ and the asymptotic distribution of properly scaled $\hat{\beta}$ for each drift specification. Taking Case Y3X1 as an example, the last column should be interpreted as $n^{-(1/2-\delta_y)}\hat{\beta} \Rightarrow d_y \xi_x / B_{xx}$.

a relatively strong drift under $\beta = 0$.

As shown in Table 1, the properly scaled $\hat{\beta}$ is asymptotically normal in Cases Y1X3, Y2X3, and Y3X3(iii). Comparing these cases helps us understand the consequence of the stronger drift in y on the spurious regression. Switching from a weak drift in y to the semi-strong drift makes the scaled $\hat{\beta}$ have mean d_y/d_x instead of 0. Further, switching from the semi-strong drift to the strong drift makes $\hat{\beta}$ itself converge to d_y/d_x instead of 0. These results indicate that the stronger drift in y exacerbates the spurious regression by pushing $\hat{\beta}$ further away from 0.

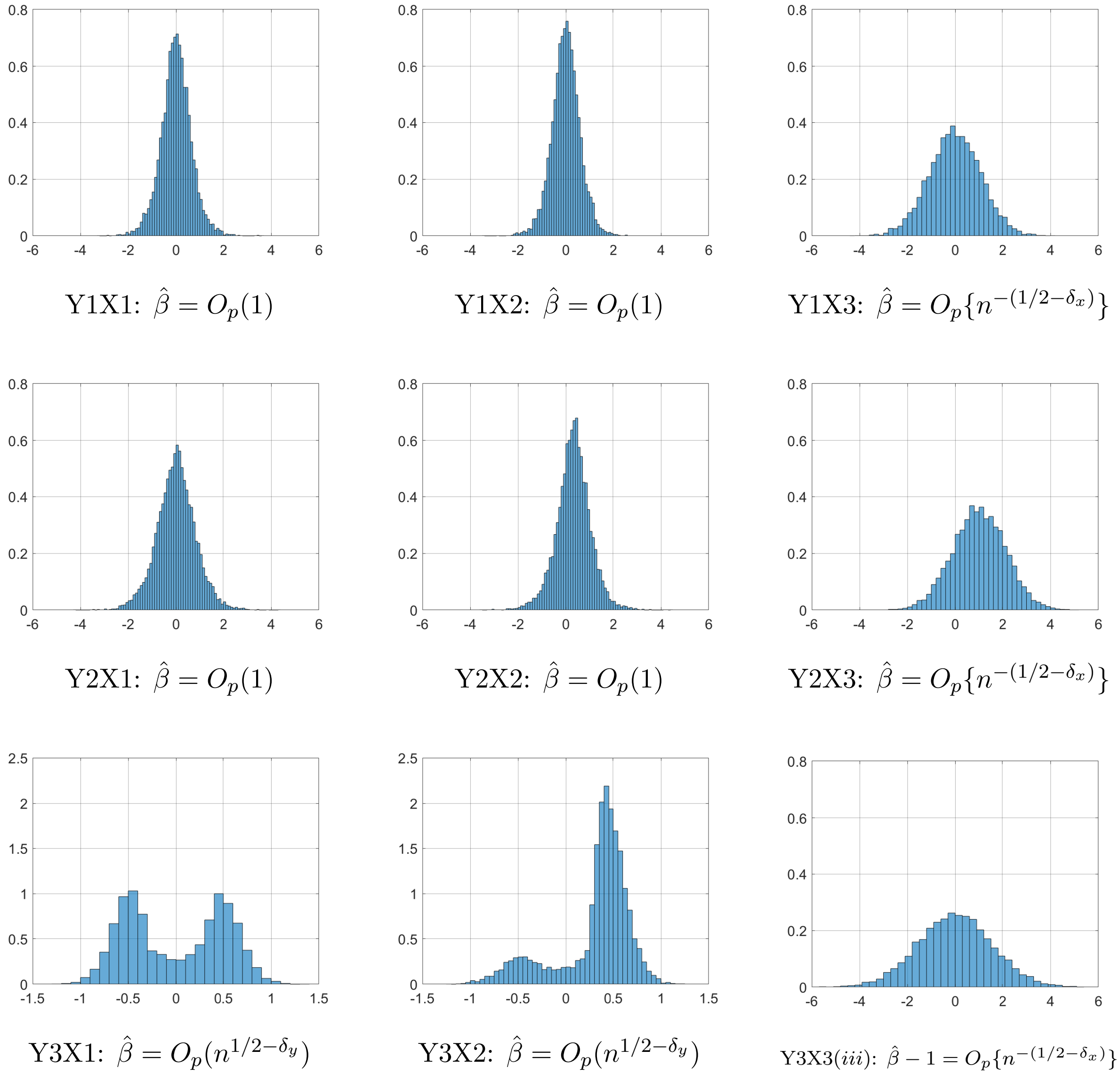
Another interesting finding is that (d_y, d_x) do not affect the order of $\hat{\beta}$ for any cases considered. Besides, d_y does not appear in the asymptotic distribution of scaled $\hat{\beta}$ when y has a weak drift. Similarly, d_x does not appear in the asymptotic distribution of scaled $\hat{\beta}$ when x has a weak drift. These facts suggest that (δ_y, δ_x) play more important role than (d_y, d_x) in determining the asymptotic behavior of $\hat{\beta}$.

When x has a weak or semi-strong drift, the asymptotic distributions of scaled $\hat{\beta}$ take non-standard forms (Cases X1-X2 in Table 1). These non-standard distributions can be numerically approximated by simulating $(\xi_x, \xi_y, B_{xx}, B_{xy})$; see Appendix A.3 for a specific procedure of the simulation. In Figure 3, we present numerical approximations of the asymptotic distributions of scaled $\hat{\beta}$ for all cases except Y3X3(i) and (ii), where the scaled $\hat{\beta}$ converges in probability to the non-stochastic constant d_y/d_x. For visual clarity, we set $d_y = d_x = \sigma_{\epsilon y} = \sigma_{\epsilon x} = 1$ for all cases considered.

Several remarks on Figure 3 are in order. First, the symmetric and bimodal distribution arises in Case Y3X1. The source of the bimodality can be seen from the functional form that $n^{-(1/2-\delta_y)}\hat{\beta} \Rightarrow d_y \xi_x / B_{xx}$ (Table 1). By (15), ξ_x is normally distributed. The Brownian motion $W_x(t)$ is normally distributed and B_{xx} is a function of $W_x^2(t)$, hence B_{xx} is related to a chi-squared random variable; recall (14). These properties imply that $d_y \xi_x / B_{xx}$ is related to the ratio of a normal random variable to a chi-squared random variable, confirming the bimodality in Case Y3X1. While it is well known that some ratio distributions exhibit bimodality (e.g., Marsaglia, 1965, Pham-Gia, Turkkan, and Marchand, 2006), we are unaware of the appearance of a bimodal distribution in any previous work on spurious regression.

An intuitive explanation of the symmetry observed in Case Y3X1 is as follows. Since $d_y = 1$ and $\delta_y < 1/2$, y has a positive, strong drift. Since $d_x = 1$ and $\delta_x > 1/2$, x has a positive, weak drift. The strong drift in y dominates the weak drift in x, hence there is

Figure 3: Asymptotic distribution of the scaled least squares estimator $\hat{\beta}$



Y1X1: $\hat{\beta} = O_p(1)$ Y1X2: $\hat{\beta} = O_p(1)$ Y1X3: $\hat{\beta} = O_p\{n^{-(1/2-\delta_x)}\}$

Y2X1: $\hat{\beta} = O_p(1)$ Y2X2: $\hat{\beta} = O_p(1)$ Y2X3: $\hat{\beta} = O_p\{n^{-(1/2-\delta_x)}\}$

Y3X1: $\hat{\beta} = O_p(n^{1/2-\delta_y})$ Y3X2: $\hat{\beta} = O_p(n^{1/2-\delta_y})$ Y3X3(iii): $\hat{\beta} - 1 = O_p\{n^{-(1/2-\delta_x)}\}$

The DGP is $y_t = n^{-\delta_y} + y_{t-1} + \epsilon_{yt}$ and $x_t = n^{-\delta_x} + x_{t-1} + \epsilon_{xt}$, where the error variances are $\sigma^2_{\epsilon y} = \sigma^2_{\epsilon x} = 1$. Y1: $\delta_y \in (1/2, \infty)$. Y2: $\delta_y = 1/2$. Y3: $\delta_y \in [0, 1/2)$. X1: $\delta_x \in (1/2, \infty)$. X2: $\delta_x = 1/2$. X3: $\delta_x \in [0, 1/2)$. Y3X3(iii): $\delta_y = \delta_x \in [0, 1/2)$. The regression model is $y_t = \alpha + \beta x_t + u_t$. $\hat{\beta}$ is the least squares estimator for β. This figure presents numerical approximations of the asymptotic distributions of properly scaled $\hat{\beta}$.

effectively the zero drift in x. It is therefore not surprising that $\hat{\beta}$ takes a positive value and a negative value of the same magnitude with equal probability.

Second, the *asymmetric* and bimodal distribution arises in Case Y3X2; the right peak around $\hat{\beta} = 0.5$ is taller than the left peak around $\hat{\beta} = -0.5$. The crucial difference from Case Y3X1 is that the semi-strong drift in x is non-negligible relative to the strong drift in y. Hence, there is an upward pressure on the spurious correlation between y and x, producing the asymmetric distribution. In fact, the right peak becomes so dominant that the distribution becomes asymmetric and *unimodal* when $d_x = 3$, for example.[4]

Third, the approximated distributions under the strong drift in x all resemble the normal distributions, confirming the asymptotic normality in Cases Y1X3, Y2X3, and Y3X3(iii) (Table 1). In particular, the distributions in Cases Y1X3 and Y2X3 are respectively centered at 0 and $d_y/d_x = 1$, keeping the same variance $6\sigma_{\epsilon y}^2/5d_x^2 = 1.2$ as expected.

4 Asymptotic properties of the squared t-statistic

In this section, we derive the asymptotic properties of the squared t-statistic $\hat{t}_\beta^2$ under the generalized drifts (2). We focus on $\hat{t}_\beta^2$ instead of $\hat{t}_\beta$ for analytical convenience. A key insight is that the t-statistic appearing in (6) can be rewritten as follows.

Lemma 4. *It follows that*

$$\hat{t}_\beta^2 = \frac{n(\lambda_n^{xy})^2}{\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2}, \tag{20}$$

where $(\lambda_n^{xx}, \lambda_n^{xy})$ *are defined in* (9) *and*

$$\lambda_n^{yy} = n^{-3}\sum_{t=1}^{n}(y_t - \bar{y})^2. \tag{21}$$

See Appendix A.4 for the proof of Lemma 4.

In view of Lemma 4, characterizing the asymptotic properties of $\hat{t}_\beta^2$ requires the asymptotic expansion of λ_n^{yy}. By the analogy of (10), it follows that

$$\lambda_n^{yy} = \frac{1}{12}n^{-2\delta_y}d_y^2 + 2n^{-(1/2+\delta_y)}d_y\hat{\xi}_y + n^{-1}\hat{B}_{yy} + o_p(n^{-1}),$$

where

$$\hat{B}_{yy} = n^{-2}\sum_{t=1}^{n}s_{yt}^2 - \left(n^{-3/2}\sum_{t=1}^{n}s_{yt}\right)^2.$$

[4] The figure for Case Y3X2 with $(d_y, d_x) = (1, 3)$ is omitted to save space, but available upon request.

As in Lemma 2, we have that $\hat{B}_{yy} \Rightarrow B_{yy}$, where

$$B_{yy} = \int_0^1 \sigma_{\epsilon y}^2 W_y^2(t)dt - \left\{\int_0^1 \sigma_{\epsilon y} W_y(t)dt\right\}^2.$$

The definitions of $(\hat{B}_{yy}, B_{yy})$ are analogous to those of $(\hat{B}_{xx}, B_{xx})$; see (13)-(14).

By the analogy of (17), the dominant terms of λ_n^{yy} are summarized as follows:

$$\lambda_n^{yy} = \begin{cases} n^{-1}\hat{B}_{yy} + o_p(n^{-1}) & \text{(Case Y1)}, \\ n^{-1}\hat{Z}_{yy} + o_p(n^{-1}) & \text{(Case Y2)}, \\ \frac{1}{12}n^{-2\delta_y}d_y^2 + 2n^{-(1/2+\delta_y)}d_y\hat{\xi}_y + n^{-1}\hat{B}_{yy} + o_p(n^{-1}) & \text{(Case Y3)}, \end{cases} \tag{22}$$

where

$$\hat{Z}_{yy} = \frac{1}{12}d_y^2 + 2d_y\hat{\xi}_y + \hat{B}_{yy}.$$

By Lemma 2, we have that $\hat{Z}_{yy} \Rightarrow Z_{yy}$, where

$$Z_{yy} = \frac{1}{12}d_y^2 + 2d_y\xi_y + B_{yy}.$$

To characterize the asymptotic behavior of $\hat{t}_\beta^2$, we expand the numerator and denominator of (20) separately. Taking the square of (18), the numerator of (20) is expanded as follows:

$$n(\lambda_n^{xy})^2 = \begin{cases} n^{-1}\hat{B}_{xy}^2 + o_p(n^{-1}) & \text{(Case Y1X1)}, \\ n^{-1}(d_x\hat{\xi}_y + \hat{B}_{xy})^2 + o_p(n^{-1}) & \text{(Case Y1X2)}, \\ n^{-2\delta_x}d_x^2\hat{\xi}_y^2 + o_p(n^{-2\delta_x}) & \text{(Case Y1X3)}, \\ n^{-1}(d_y\hat{\xi}_x + \hat{B}_{xy})^2 + o_p(n^{-1}) & \text{(Case Y2X1)}, \\ n^{-1}\hat{Z}_{xy}^2 + o_p(n^{-1}) & \text{(Case Y2X2)}, \\ n^{-2\delta_x}d_x^2\left(\frac{1}{12}d_y + \hat{\xi}_y\right)^2 + o_p(n^{-2\delta_x}) & \text{(Case Y2X3)}, \\ n^{-2\delta_y}d_y^2\hat{\xi}_x^2 + o_p(n^{-2\delta_y}) & \text{(Case Y3X1)}, \\ n^{-2\delta_y}d_y^2\left(\frac{1}{12}d_x + \hat{\xi}_x\right)^2 + o_p(n^{-2\delta_y}) & \text{(Case Y3X2)}, \\ \frac{1}{144}n^{-2(\delta_x+\delta_y-1/2)}d_x^2d_y^2 + o_p\left\{n^{-2(\delta_x+\delta_y-1/2)}\right\} & \text{(Case Y3X3)}. \end{cases} \tag{23}$$

Combining (17), (18), and (22), the denominator of (20) is expanded as follows:

$$\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2 = \begin{cases} n^{-2}(\hat{B}_{xx}\hat{B}_{yy} - \hat{B}_{xy}^2) + o_p(n^{-2}) & \text{(Case Y1X1)}, \\ n^{-2}\{\hat{B}_{yy}\hat{Z}_{xx} - (d_x\hat{\xi}_y + \hat{B}_{xy})^2\} + o_p(n^{-2}) & \text{(Case Y1X2)}, \\ n^{-(1+2\delta_x)}d_x^2\left(\frac{1}{12}\hat{B}_{yy} - \hat{\xi}_y^2\right) + o_p\left\{n^{-(1+2\delta_x)}\right\} & \text{(Case Y1X3)}, \\ n^{-2}\{\hat{B}_{xx}\hat{Z}_{yy} - (d_y\hat{\xi}_x + \hat{B}_{xy})^2\} + o_p(n^{-2}) & \text{(Case Y2X1)}, \\ n^{-2}(\hat{Z}_{xx}\hat{Z}_{yy} - \hat{Z}_{xy}^2) + o_p(n^{-2}) & \text{(Case Y2X2)}, \\ n^{-(1+2\delta_x)}d_x^2\left(\frac{1}{12}\hat{B}_{yy} - \hat{\xi}_y^2\right) + o_p\left\{n^{-(1+2\delta_x)}\right\} & \text{(Case Y2X3)}, \\ n^{-(1+2\delta_y)}d_y^2\left(\frac{1}{12}\hat{B}_{xx} - \hat{\xi}_x^2\right) + o_p\left\{n^{-(1+2\delta_y)}\right\} & \text{(Case Y3X1)}, \\ n^{-(1+2\delta_y)}d_y^2\left(\frac{1}{12}\hat{B}_{xx} - \hat{\xi}_x^2\right) + o_p\left\{n^{-(1+2\delta_y)}\right\} & \text{(Case Y3X2)}, \\ n^{-(1+2\delta_x)}d_x^2\left(\frac{1}{12}\hat{B}_{yy} - \hat{\xi}_y^2\right) + n^{-(1+2\delta_y)}d_y^2\left(\frac{1}{12}\hat{B}_{xx} - \hat{\xi}_x^2\right) & \\ \qquad -2n^{-(1+\delta_x+\delta_y)}d_xd_y\left(\frac{1}{12}\hat{B}_{xy} - \hat{\xi}_x\hat{\xi}_y\right) + o_p\left\{n^{-(1+\delta_x+\delta_y)}\right\} & \text{(Case Y3X3)}. \end{cases} \tag{24}$$

Case Y3X3 requires extra caution when computing (20). As shown in (23), the dominant term of $n(\lambda_n^{xy})^2$ is of order $n^{-2(\delta_x+\delta_y-1/2)}$. Equation (24) implies that the dominant term of $\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2$ is of order $n^{-(1+2\min\{\delta_x,\delta_y\})}$. The relative magnitude of these terms depends on the values of $\delta_x \in [0, 1/2)$ and $\delta_y \in [0, 1/2)$. Hence, Case Y3X3 should be divided into the three subcases defined in (19). For each subcase, (24) is rewritten as follows:

$$\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2 = \begin{cases} n^{-(1+2\delta_y)}d_y^2\left(\frac{1}{12}\hat{B}_{xx} - \hat{\xi}_x^2\right) + o_p\left\{n^{-(1+2\delta_y)}\right\} & \text{(Y3X3}(i)), \\ n^{-(1+2\delta_x)}d_x^2\left(\frac{1}{12}\hat{B}_{yy} - \hat{\xi}_y^2\right) + o_p\left\{n^{-(1+2\delta_x)}\right\} & \text{(Y3X3}(ii)), \\ n^{-(1+2\delta_x)}\left\{\frac{1}{12}\hat{B}^* - (\hat{\xi}^*)^2\right\} + o_p\left\{n^{-(1+2\delta_x)}\right\} & \text{(Y3X3}(iii)), \end{cases} \tag{25}$$

where

$$\hat{B}^* = d_y^2\hat{B}_{xx} - 2d_xd_y\hat{B}_{xy} + d_x^2\hat{B}_{yy}, \qquad \hat{\xi}^* = d_y\hat{\xi}_x - d_x\hat{\xi}_y.$$

By Lemma 2, we have that $\hat{B}^* \Rightarrow B^*$ and $\hat{\xi}^* \Rightarrow \xi^*$, where

$$B^* = d_y^2B_{xx} - 2d_xd_yB_{xy} + d_x^2B_{yy}, \qquad \xi^* = d_y\xi_x - d_x\xi_y. \tag{26}$$

Substitute (23)-(25) into (20) to characterize the asymptotic properties of $\hat{t}^2_\beta$.

Theorem 5. *Let the DGP (1) and Assumption 1 hold, then for each drift specification, the order of stochastic divergence of $\hat{t}^2_\beta$ and the asymptotic distribution of scaled $\hat{t}^2_\beta$ are characterized in Table 2.*

The proof of Theorem 5 is omitted, since for all cases the results follow directly from substitution of (23)-(25) into (20) and application of Lemma 2.

As shown in Table 2, $\hat{t}^2_\beta = O_p(n)$ unless both y and x have strong drifts. In Case Y3X3, the squared t-statistic diverges at a faster rate than n:

$$\hat{t}^2_\beta = O_p\left\{n^{2(1-\max\{\delta_x,\delta_y\})}\right\}, \qquad \delta_y \in [0,1/2), \quad \delta_x \in [0,1/2).$$

These results indicate that the probability of making Type I Error approaches 1 for all cases, and the symptom is particularly serious in Case Y3X3. Another implication is that (d_y, d_x) do not affect the rate of divergence for any cases considered.

Our divergence results contain two existing results as special cases. First, Phillips (1986) found that $\hat{t}^2_\beta = O_p(n)$ when both y and x have zero drifts; this result can be replicated by setting $\delta_x \to \infty$ and $\delta_y \to \infty$ in Case Y1X1. Second, Entorf (1997) found that $\hat{t}^2_\beta = O_p(n^2)$ when both y and x have nonzero constant drifts; this result can be replicated by setting $\delta_x = \delta_y = 0$ in Case Y3X3(iii).

We now focus on the asymptotic distributions of the properly scaled $\hat{t}^2_\beta$ (Table 2). The asymptotic distributions have an interchangeable structure between Cases Y1X2 and Y2X1; between Cases Y1X3 and Y3X1; between Cases Y2X3 and Y3X2; between Cases Y3X3(i) and Y3X3(ii). We have found that $n^{-1}\hat{t}^2_\beta \Rightarrow B^2_{xy}/(B_{xx}B_{yy} - B^2_{xy})$ in Case Y1X1, generalizing the existing result of Phillips (1986) from the zero drifts to the weak drifts. Entorf (1997) derived the n^2-divergence property under the nonzero constant drifts, but did not derive the asymptotic distribution of $n^{-2}\hat{t}^2_\beta$. We have filled this gap by discovering that:

$$n^{-2}\hat{t}^2_\beta \Rightarrow \frac{\frac{1}{144}d_x^2 d_y^2}{\frac{1}{12}B^* - (\xi^*)^2} \qquad (\delta_y = \delta_x = 0),$$

where B^* and ξ^* are defined in (26).

In Figures 4-5, we present numerical approximations of the asymptotic distributions of scaled $\hat{t}^2_\beta$ for all cases, where $d_y = d_x = \sigma_{\epsilon y} = \sigma_{\epsilon x} = 1$. All cases except for Y3X3 are shown

Table 2: Asymptotic properties of the squared t-statistic $\hat{t}_\beta^2$

Case	δ_y	δ_x	Subcase	Order of $\hat{t}_\beta^2$	Asy. dist. of scaled $\hat{t}_\beta^2$
Y1X1	$(1/2, \infty)$	$(1/2, \infty)$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{B_{xy}^2}{B_{xx}B_{yy} - B_{xy}^2}$
Y1X2	$(1/2, \infty)$	$1/2$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{(d_x\xi_y + B_{xy})^2}{B_{yy}Z_{xx} - (d_x\xi_y + B_{xy})^2}$
Y1X3	$(1/2, \infty)$	$[0, 1/2)$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{\xi_y^2}{\frac{1}{12}B_{yy} - \xi_y^2}$
Y2X1	$1/2$	$(1/2, \infty)$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{(d_y\xi_x + B_{xy})^2}{B_{xx}Z_{yy} - (d_y\xi_x + B_{xy})^2}$
Y2X2	$1/2$	$1/2$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{Z_{xy}^2}{Z_{xx}Z_{yy} - Z_{xy}^2}$
Y2X3	$1/2$	$[0, 1/2)$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{(\frac{1}{12}d_y + \xi_y)^2}{\frac{1}{12}B_{yy} - \xi_y^2}$
Y3X1	$[0, 1/2)$	$(1/2, \infty)$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{\xi_x^2}{\frac{1}{12}B_{xx} - \xi_x^2}$
Y3X2	$[0, 1/2)$	$1/2$	-	$\hat{t}_\beta^2 = O_p(n)$	$\frac{(\frac{1}{12}d_x + \xi_x)^2}{\frac{1}{12}B_{xx} - \xi_x^2}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(i) $\delta_y < \delta_x$	$\hat{t}_\beta^2 = O_p\left\{n^{2(1-\delta_x)}\right\}$	$\frac{\frac{1}{144}d_x^2}{\frac{1}{12}B_{xx} - \xi_x^2}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(ii) $\delta_y > \delta_x$	$\hat{t}_\beta^2 = O_p\left\{n^{2(1-\delta_y)}\right\}$	$\frac{\frac{1}{144}d_y^2}{\frac{1}{12}B_{yy} - \xi_y^2}$
Y3X3	$[0, 1/2)$	$[0, 1/2)$	(iii) $\delta_y = \delta_x$	$\hat{t}_\beta^2 = O_p\left\{n^{2(1-\delta_x)}\right\}$	$\frac{\frac{1}{144}d_x^2 d_y^2}{\frac{1}{12}B^* - (\xi^*)^2}$

DGP: $y_t = d_y n^{-\delta_y} + y_{t-1} + \epsilon_{yt}$ and $x_t = d_x n^{-\delta_x} + x_{t-1} + \epsilon_{xt}$, where $d_y \neq 0$ and $d_x \neq 0$. Model: $y_t = \alpha + \beta x_t + u_t$. $\hat{t}_\beta^2$ is the squared t-statistic associated with β. This table summarizes the order of stochastic divergence of $\hat{t}_\beta^2$ and the asymptotic distribution of properly scaled $\hat{t}_\beta^2$ for each drift specification. Taking Case Y1X1 as an example, the last column should be interpreted as $n^{-1}\hat{t}_\beta^2 \Rightarrow B_{xy}^2/(B_{xx}B_{yy} - B_{xy}^2)$.

in Figure 4, and Cases Y3X3(i)-(iii) are shown in Figure 5. In Figure 4, $n^{-1}\hat{t}_{\beta}^{2}$ follows a positively skewed distribution whose peak is located at 0. We confirm that the asymptotic distributions are identical to each other between Cases Y1X2 and Y2X1; between Cases Y1X3 and Y3X1; between Cases Y2X3 and Y3X2. The structure of Figure 4 can be thought of as a 3×3 symmetric matrix which is invariant under "transposition".

For Case Y3X3(i), $n^{-2(1-\delta_x)}\hat{t}_{\beta}^{2}$ follows a positively skewed distribution whose peak is located around 1 (Figure 5). For Case Y3X3(ii), $n^{-2(1-\delta_y)}\hat{t}_{\beta}^{2}$ follows the same distribution as in (i). The equivalence of the two distributions confirms the interchangeability between the two subcases observed in Table 2. For Case Y3X3(iii), $n^{-2(1-\delta_x)}\hat{t}_{\beta}^{2}$ follows a positively skewed distribution whose peak is located around 0.5. Summarizing Table 2 and Figures 4-5, the asymptotic properties of $\hat{t}_{\beta}^{2}$ differ starkly between the case where both y and x have strong drifts and the other cases.

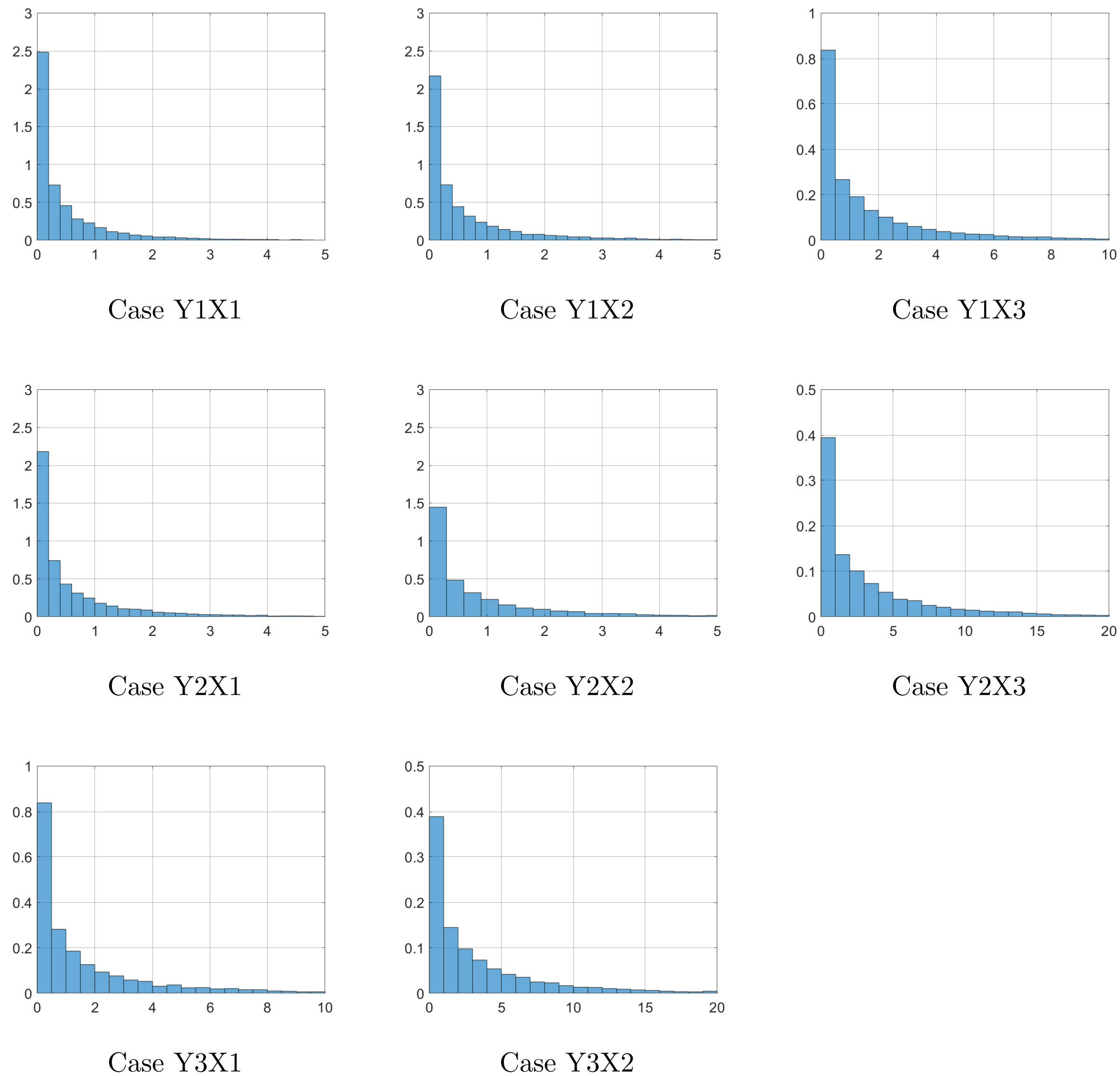
5 Conclusion

Spurious regression is one of the most fundamental topics in time series analysis. In the literature, the asymptotic behavior of the estimated slope parameter $\hat{\beta}$ and the associated t-statistic $\hat{t}_{\beta}$ is studied only when both y and x have zero drifts (Phillips, 1986) or nonzero constant drifts (Entorf, 1997). In this paper, we have vastly extended these scenarios by allowing for each of y and x to have the generalized drift: $d_{yn} = d_y n^{-\delta_y}$ and $d_{xn} = d_x n^{-\delta_x}$. Characterizing the consequence of the generalized drift is not only of theoretical interest, but also of practical relevance since in empirical applications it is often hard to judge if target variables have zero or nonzero constant drifts.

Given the generalized drift, we have derived the order of asymptotic convergence or divergence of $\hat{\beta}$ and $\hat{t}_{\beta}^{2}$ as well as their asymptotic distributions. We have found that $\hat{\beta}$ may converge, diverge, or neither depending on the values of (δ_y, δ_x). Further, the asymptotic distribution of the properly scaled $\hat{\beta}$ takes on various interesting shapes such as a bimodal and asymmetric distribution. We have also revealed that $\hat{t}_{\beta}^{2}$ diverges at rate n if $\delta_y \geq 1/2$ or $\delta_x \geq 1/2$. If both y and x have strong drifts (i.e., $\delta_y < 1/2$ and $\delta_x < 1/2$), then $\hat{t}_{\beta}^{2}$ diverges at the faster rate than n. Hence, the conventional t-test for the zero hypothesis of β fails for all cases considered, with the size of the test converging to 1.

To avoid the spurious regression involving random walks with generalized drifts, the

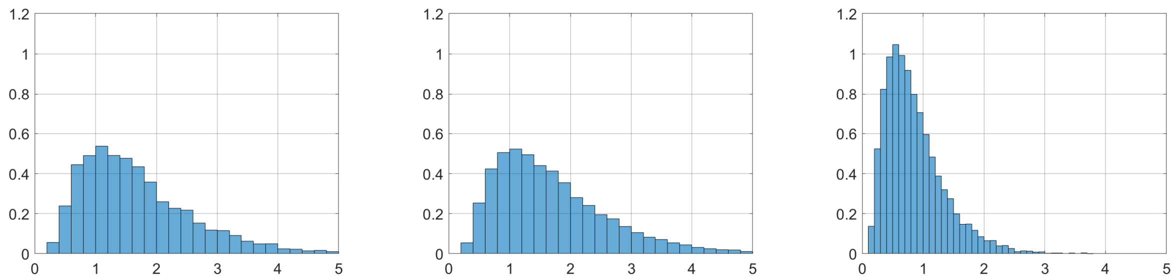
Figure 4: Asymptotic distribution of the scaled, squared t-statistic $n^{-1}\hat{t}_\beta^2$



The DGP is $y_t = n^{-\delta_y} + y_{t-1} + \epsilon_{yt}$ and $x_t = n^{-\delta_x} + x_{t-1} + \epsilon_{xt}$, where the error variances are $\sigma_{\epsilon y}^2 = \sigma_{\epsilon x}^2 = 1$. Y1: $\delta_y \in (1/2, \infty)$. Y2: $\delta_y = 1/2$. Y3: $\delta_y \in [0, 1/2)$. X1: $\delta_x \in (1/2, \infty)$. X2: $\delta_x = 1/2$. X3: $\delta_x \in [0, 1/2)$. The regression model is $y_t = \alpha + \beta x_t + u_t$. $\hat{t}_\beta$ is the t-statistic associated with $H_0 : \beta = 0$. This figure presents numerical approximations of the asymptotic distributions of properly scaled $\hat{t}_\beta^2$. For all cases considered in this figure, $\hat{t}_\beta^2 = O_p(n)$ and hence the proper scaling factor is n^{-1}.

standard approach of differencing or detrending should be taken. As is well known, the key condition for avoiding the spurious regression is that the residual should be stationary. To

Figure 5: Asymptotic distribution of the scaled, squared t-statistic $\hat{t}^2_\beta$ (Case Y3X3)



$(i)\ \hat{t}^2_\beta = O_p\left\{n^{2(1-\delta_x)}\right\}$ $(ii)\ \hat{t}^2_\beta = O_p\left\{n^{2(1-\delta_y)}\right\}$ $(iii)\ \hat{t}^2_\beta = O_p\left\{n^{2(1-\delta_x)}\right\}$

The DGP is $y_t = n^{-\delta_y} + y_{t-1} + \epsilon_{yt}$ and $x_t = n^{-\delta_x} + x_{t-1} + \epsilon_{xt}$, where the error variances are $\sigma^2_{\epsilon y} = \sigma^2_{\epsilon x} = 1$. Y3: $\delta_y \in [0, 1/2)$. X3: $\delta_x \in [0, 1/2)$. The subcases are organized as follows: (i) $\delta_y < \delta_x$, (ii) $\delta_y > \delta_x$, and (iii) $\delta_y = \delta_x$. The regression model is $y_t = \alpha + \beta x_t + u_t$. $\hat{t}_\beta$ is the t-statistic associated with $H_0 : \beta = 0$. This figure presents numerical approximations of the asymptotic distributions of properly scaled $\hat{t}^2_\beta$.

ensure the stationarity of the residual, the researcher should sufficiently difference or detrend y and x. This is a well-established remedy to the spurious regression, and it should operate well for the case of generalized drifts.

Acknowledgements

We thank Marcus Chambers, Shigeyuki Hamori, Serena Ng, and Zheng Zhang for helpful comments and discussions. The second author, Kaiji Motegi, is grateful for the financial support of the Organization for Advanced and Integrated Research, Kobe University.

Declaration of Interest Statement

The authors declare that there is no conflict of interest.

Data Availability Statement

In Section 1, Nikkei 225 (also known as Nikkei Stock Average) and the land temperature anomalies of North America are analyzed. The annual average data of Nikkei 225 are retrieved at Federal Reserve Economic Data maintained by the Federal Reserve Bank of St. Louis (series ID: `NIKKEI225`). The annual temperature data are retrieved at Climate at a Glance, NOAA National Centers for Environmental Information. The data and the Matlab codes which are sufficient for replicating all the empirical and numerical results of this paper are publicly available at the personal website of the second author, Kaiji Motegi:

`http://www2.kobe-u.ac.jp/~motegi/research.html`

References

Billingsley, P. (1968): *Convergence of Probability Measures.* Wiley, New York.

Bradley, R. C. (2005): "Basic properties of strong mixing conditions: A survey and some open questions," *Probability Surveys*, 2, 107–144.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford, U.K.

Donsker, M. D. (1951): "An Invariance Principle for Certain Probability Limit Theorems," *Memoirs of the American Mathematical Society*, 6, 1–12.

Durlauf, S. N., and P. C. B. Phillips (1988): "Trends versus Random Walks in Time Series Analysis," *Econometrica*, 56, 1333–1354.

Entorf, H. (1997): "Random walks with drifts: Nonsense regression and spurious fixed-effect estimation," *Journal of Econometrics*, 80, 287–296.

Ernst, P. A., L. A. Shepp, and A. J. Wyner (2017): "Yule's "nonsense correlation" solved!," *The Annals of Statistics*, 45, 1789–1809.

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

Marsaglia, G. (1965): "Ratios of Normal Variables and Ratios of Sums of Uniform Variables," *Journal of the American Statistical Association*, 60, 193–204.

Park, J. Y., and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.

——— (1989): "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–131.

Pham-Gia, T., N. Turkkan, and E. Marchand (2006): "Density of the Ratio of Two Normal Random Variables and Applications," *Communications in Statistics - Theory and Methods*, 35, 1569–1591.

Phillips, P. C. B. (1986): "Understanding Spurious Regression in Econometrics," *Journal of Econometrics*, 33, 311–340.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models With Some Unit Roots," *Econometrica*, 58, 113–144.

Ventosa-Santaulària, D. (2009): "Spurious Regression," *Journal of Probability and Statistics*, 1, 1–27.

Vigen, T. (2015): *Spurious Correlations*. Hachette Books.

Appendices

A.1 Proof of Lemma 1

Proof of Lemma 1. We prove (10) only; the proof of (11) is completely analogous and hence omitted. Toward proving (10), we first observe the basic identities:

$$\sum_{t=1}^{n} t = \frac{1}{2}n(n+1), \tag{A.1}$$

$$\sum_{t=1}^{n} \left\{t - \frac{1}{2}(n+1)\right\} = 0, \tag{A.2}$$

$$\sum_{t=1}^{n} \left\{t - \frac{1}{2}(n+1)\right\}^2 = \frac{1}{12}n(n+1)(n-1). \tag{A.3}$$

Equations (A.1)-(A.3) can be verified by elementary algebra. Implement the backward iteration to (1):

$$\begin{aligned} x_t &= d_{xn} + x_{t-1} + \epsilon_{xt} = d_{xn} + (d_{xn} + x_{t-2} + \epsilon_{x,t-1}) + \epsilon_{xt} \\ &= 2d_{xn} + x_{t-2} + \epsilon_{x,t-1} + \epsilon_{xt} = \cdots = td_{xn} + x_{t-t} + \epsilon_{x,t-(t-1)} + \cdots + \epsilon_{x,t-1} + \epsilon_{xt} \\ &= td_{xn} + x_0 + \sum_{\tau=1}^{t} \epsilon_{x\tau} = td_{xn} + s_{xt}, \end{aligned} \tag{A.4}$$

where the last equality follows by (3). By (A.4), we have that

$$\bar{x} \equiv n^{-1}\sum_{t=1}^{n} x_t = n^{-1}\sum_{t=1}^{n}(td_{xn} + s_{xt}) = \frac{1}{2}d_{xn}(n+1) + n^{-1}\sum_{t=1}^{n} s_{xt}, \tag{A.5}$$

where the last equality follows by (A.1). By (A.4) and (A.5), we have that

$$x_t - \bar{x} = d_{xn}\left\{t - \frac{1}{2}(n+1)\right\} + s_{xt} - n^{-1}\sum_{\tau=1}^{n} s_{x\tau}.$$

Take the square of both sides to get

$$(x_t - \bar{x})^2 = d_{xn}^2 \left\{ t - \frac{1}{2}(n+1) \right\}^2 + s_{xt}^2 + \left(n^{-1} \sum_{\tau=1}^{n} s_{x\tau} \right)^2 + 2d_{xn} s_{xt} \left\{ t - \frac{1}{2}(n+1) \right\}$$
$$- 2n^{-1} d_{xn} \left\{ t - \frac{1}{2}(n+1) \right\} \sum_{\tau=1}^{n} s_{x\tau} - 2n^{-1} s_{xt} \sum_{\tau=1}^{n} s_{x\tau}.$$

Take the sum of each side for $t = 1, \ldots, n$ to get

$$\begin{aligned}
\sum_{t=1}^{n} (x_t - \bar{x})^2 = d_{xn}^2 \sum_{t=1}^{n} \left\{ t - \frac{1}{2}(n+1) \right\}^2 + \sum_{t=1}^{n} s_{xt}^2 + n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 \\
+ 2d_{xn} \sum_{t=1}^{n} s_{xt} \left\{ t - \frac{1}{2}(n+1) \right\} \\
- 2n^{-1} d_{xn} \left[\sum_{t=1}^{n} \left\{ t - \frac{1}{2}(n+1) \right\} \right] \left(\sum_{t=1}^{n} s_{xt} \right) - 2n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 \\
= \frac{1}{12} d_{xn}^2 n(n+1)(n-1) + \sum_{t=1}^{n} s_{xt}^2 + n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 \\
+ 2d_{xn} \sum_{t=1}^{n} s_{xt} \left\{ t - \frac{1}{2}(n+1) \right\} - 2n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 \\
= \frac{1}{12} d_{xn}^2 n(n+1)(n-1) + \sum_{t=1}^{n} s_{xt}^2 - n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 \\
+ 2d_{xn} \sum_{t=1}^{n} s_{xt} \left\{ t - \frac{1}{2}(n+1) \right\},
\end{aligned} \tag{A.6}$$

where the second equality follows by (A.2) and (A.3). Note that

$$\sum_{t=1}^{n} s_{xt}^2 - n^{-1} \left(\sum_{t=1}^{n} s_{xt} \right)^2 = n^2 \left\{ n^{-2} \sum_{t=1}^{n} s_{xt}^2 - \left(n^{-3/2} \sum_{t=1}^{n} s_{xt} \right)^2 \right\} = n^2 \hat{B}_{xx}, \tag{A.7}$$

where $\hat{B}_{xx}$ is defined in (13). Note also that

$$\sum_{t=1}^{n} s_{xt} \left\{ t - \frac{1}{2}(n+1) \right\} = \sum_{t=1}^{n} t s_{xt} - \frac{1}{2} n \sum_{t=1}^{n} s_{xt} - \frac{1}{2} \sum_{t=1}^{n} s_{xt}$$

$$
\begin{aligned}
&= n^{5/2}\left(n^{-5/2}\sum_{t=1}^{n} ts_{xt} - \frac{1}{2}n^{-3/2}\sum_{t=1}^{n} s_{xt}\right) - \frac{1}{2}\sum_{t=1}^{n} s_{xt} \\
&= n^{5/2}\hat{\xi}_x - \frac{1}{2}\sum_{t=1}^{n} s_{xt},
\end{aligned} \tag{A.8}
$$

where $\hat{\xi}_x$ is defined in (12). Substitute (A.7) and (A.8) into (A.6) to get

$$
\begin{aligned}
\sum_{t=1}^{n}(x_t - \bar{x})^2 &= \frac{1}{12}d_{xn}^2 n(n+1)(n-1) + n^2\hat{B}_{xx} + 2d_{xn}\left(n^{5/2}\hat{\xi}_x - \frac{1}{2}\sum_{t=1}^{n} s_{xt}\right) \\
&= \frac{1}{12}d_{xn}^2 n(n+1)(n-1) + 2n^{5/2}d_{xn}\hat{\xi}_x + n^2\hat{B}_{xx} - d_{xn}\sum_{t=1}^{n} s_{xt}.
\end{aligned}
$$

Thus, we have that

$$
\begin{aligned}
\lambda_n^{xx} &\equiv n^{-3}\sum_{t=1}^{n}(x_t-\bar{x})^2 = \frac{1}{12}d_{xn}^2\left(1-n^{-2}\right) + 2n^{-1/2}d_{xn}\hat{\xi}_x + n^{-1}\hat{B}_{xx} - n^{-3}d_{xn}\sum_{t=1}^{n} s_{xt} \\
&= \frac{1}{12}d_{xn}^2 + 2n^{-1/2}d_{xn}\hat{\xi}_x + n^{-1}\hat{B}_{xx} - \frac{1}{12}n^{-2}d_{xn}^2 - n^{-3}d_{xn}\sum_{t=1}^{n} s_{xt}. \\
&= \frac{1}{12}(d_x n^{-\delta_x})^2 + 2n^{-1/2}d_x n^{-\delta_x}\hat{\xi}_x + n^{-1}\hat{B}_{xx} - \frac{1}{12}n^{-2}(d_x n^{-\delta_x})^2 - n^{-3}d_x n^{-\delta_x}\sum_{t=1}^{n} s_{xt} \\
&= \frac{1}{12}n^{-2\delta_x}d_x^2 + 2n^{-(1/2+\delta_x)}d_x\hat{\xi}_x + n^{-1}\hat{B}_{xx} - \frac{1}{12}n^{-2(1+\delta_x)}d_x^2 - n^{-(3+\delta_x)}d_x\sum_{t=1}^{n} s_{xt} \\
&= \frac{1}{12}n^{-2\delta_x}d_x^2 + 2n^{-(1/2+\delta_x)}d_x\hat{\xi}_x + n^{-1}\hat{B}_{xx} + o_p(n^{-1}).
\end{aligned}
$$

□

A.2 Proof of Theorem 3

For Cases X1, X2, and Y3X3(i)-(ii), Theorem 3 and Table 1 follow directly by substituting (17) and (18) into (8) and by applying Lemma 2. The proof of these cases is omitted to save space. For Cases Y1X3, Y2X3, and Y3X3(iii), the asymptotic distribution of scaled $\hat{\beta}$ is the normal distribution (Table 1). In this section, we provide a complete proof of these cases.

Proof of Case Y1X3. Equations (8), (17), and (18) imply that

$$
\hat{\beta} = \frac{\lambda_n^{xy}}{\lambda_n^{xx}} = \frac{n^{-(1/2+\delta_x)}d_x\hat{\xi}_y + o_p\left\{n^{-(1/2+\delta_x)}\right\}}{\frac{1}{12}n^{-2\delta_x}d_x^2 + o_p(n^{-2\delta_x})} = \frac{n^{-(1/2-\delta_x)}\hat{\xi}_y + o_p\left\{n^{-(1/2-\delta_x)}\right\}}{\frac{1}{12}d_x + o_p(1)}.
$$

Hence, we have that

$$n^{1/2-\delta_x}\hat{\beta} = \frac{\hat{\xi}_y + o_p(1)}{\frac{1}{12}d_x + o_p(1)}.$$

Equation (16) implies that $\hat{\xi}_y \Rightarrow \mathcal{N}(0,\, \sigma_{\epsilon y}^2/120)$. Hence, we have that

$$\frac{\hat{\xi}_y}{\frac{1}{12}d_x} \Rightarrow \mathcal{N}\left(0,\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right).$$

Thus, we obtain the desired result:

$$\hat{\beta} = O_p\left\{n^{-(1/2-\delta_x)}\right\}, \qquad n^{1/2-\delta_x}\hat{\beta} \Rightarrow \mathcal{N}\left(0,\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right).$$

□

Proof of Case Y2X3. Similar to the proof of Case Y1X3, we have by (8), (17), and (18) that

$$n^{1/2-\delta_x}\hat{\beta} = \frac{\frac{1}{12}d_y + \hat{\xi}_y + o_p(1)}{\frac{1}{12}d_x + o_p(1)}.$$

Equation (16) implies that $\hat{\xi}_y \Rightarrow \mathcal{N}(0,\, \sigma_{\epsilon y}^2/120)$. Hence, we have that

$$\frac{\frac{1}{12}d_y + \hat{\xi}_y}{\frac{1}{12}d_x} \Rightarrow \mathcal{N}\left(\frac{d_y}{d_x},\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right).$$

Thus, we obtain the desired result:

$$\hat{\beta} = O_p\left\{n^{-(1/2-\delta_x)}\right\}, \qquad n^{1/2-\delta_x}\hat{\beta} \Rightarrow \mathcal{N}\left(\frac{d_y}{d_x},\ \frac{6\sigma_{\epsilon y}^2}{5d_x^2}\right).$$

□

Proof of Case Y3X3(iii). Extract the relevant terms from (17):

$$\lambda_n^{xx} = \frac{1}{12}n^{-2\delta_x}d_x^2 + 2n^{-(1/2+\delta_x)}d_x\hat{\xi}_x + o_p\left\{n^{-(1/2+\delta_x)}\right\}. \tag{A.9}$$

For Case Y3X3(*iii*), $\delta_y = \delta_x \in [0, 1/2)$ and hence (18) is rewritten as

$$\lambda_n^{xy} = \frac{1}{12}n^{-2\delta_x}d_xd_y + n^{-(1/2+\delta_x)}(d_y\hat{\xi}_x + d_x\hat{\xi}_y) + o_p\left\{n^{-(1/2+\delta_x)}\right\}. \tag{A.10}$$

Divide (A.10) by (A.9) to get

$$\lambda_n^{xy} = \frac{d_y}{d_x}\lambda_n^{xx} + n^{-(1/2+\delta_x)}(d_x\hat{\xi}_y - d_y\hat{\xi}_x) + o_p\left\{n^{-(1/2+\delta_x)}\right\}.$$

Hence, we have that

$$\begin{aligned}\hat{\beta} = \frac{\lambda_n^{xy}}{\lambda_n^{xx}} &= \frac{d_y}{d_x} + \frac{n^{-(1/2+\delta_x)}(d_x\hat{\xi}_y - d_y\hat{\xi}_x) + o_p\left\{n^{-(1/2+\delta_x)}\right\}}{\frac{1}{12}n^{-2\delta_x}d_x^2 + o_p(n^{-2\delta_x})} \\ &= \frac{d_y}{d_x} + \frac{n^{-(1/2-\delta_x)}(d_x\hat{\xi}_y - d_y\hat{\xi}_x) + o_p\left\{n^{-(1/2-\delta_x)}\right\}}{\frac{1}{12}d_x^2 + o_p(1)}.\end{aligned}$$

Hence, we have that

$$n^{1/2-\delta_x}\left(\hat{\beta} - \frac{d_y}{d_x}\right) = \frac{d_x\hat{\xi}_y - d_y\hat{\xi}_x + o_p(1)}{\frac{1}{12}d_x^2 + o_p(1)}. \tag{A.11}$$

By (16), we have that

$$d_x\hat{\xi}_y - d_y\hat{\xi}_x \Rightarrow \mathcal{N}\left(0,\ \frac{d_y^2\sigma_{\epsilon x}^2 + d_x^2\sigma_{\epsilon y}^2}{120}\right)$$

and hence

$$\frac{d_x\hat{\xi}_y - d_y\hat{\xi}_x}{\frac{1}{12}d_x^2} \Rightarrow \mathcal{N}\left\{0,\ \frac{6d_y^2}{5d_x^2}\left(\frac{\sigma_{\epsilon x}^2}{d_x^2} + \frac{\sigma_{\epsilon y}^2}{d_y^2}\right)\right\}. \tag{A.12}$$

Apply (A.12) to (A.11) to conclude that

$$\hat{\beta} - \frac{d_y}{d_x} = O_p\left\{n^{-(1/2-\delta_x)}\right\}, \qquad n^{1/2-\delta_x}\left(\hat{\beta} - \frac{d_y}{d_x}\right) \Rightarrow \mathcal{N}\left\{0,\ \frac{6d_y^2}{5d_x^2}\left(\frac{\sigma_{\epsilon x}^2}{d_x^2} + \frac{\sigma_{\epsilon y}^2}{d_y^2}\right)\right\}.$$

□

A.3 Approximating the asymptotic distributions

As shown in Tables 1-2, the asymptotic distributions of scaled $\hat{\beta}$ and $\hat{t}_{\beta}^2$ involve ξ_x, ξ_y, B_{xx}, B_{yy}, and B_{xy}. In this section, we describe how to approximate these quantities numerically. Generate the following disturbances with sufficiently large sample size N (say $N = 10^5$):

$$\begin{bmatrix} v_{xt} \\ v_{yt} \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \qquad t \in \{1, \ldots, N\}.$$

Generate $w_{xt} = w_{xt-1} + v_{xt}$ and $w_{yt} = w_{yt-1} + v_{yt}$. Compute the following quantities:

$$\tilde{A}_x = N^{-5/2}\sum_{t=1}^{N} t w_{xt}, \quad \tilde{H}_x = N^{-3/2}\sum_{t=1}^{N} w_{xt}, \quad \tilde{C}_x = N^{-2}\sum_{t=1}^{N} w_{xt}^2,$$
$$\tilde{A}_y = N^{-5/2}\sum_{t=1}^{N} t w_{yt}, \quad \tilde{H}_y = N^{-3/2}\sum_{t=1}^{N} w_{yt}, \quad \tilde{C}_y = N^{-2}\sum_{t=1}^{N} w_{yt}^2, \quad \tilde{D} = N^{-2}\sum_{t=1}^{N} w_{xt}w_{yt}.$$

Assumption 1 and FCLT imply that

$$\tilde{A}_x \Rightarrow \int_0^1 tW_x(t)dt, \quad \tilde{H}_x \Rightarrow \int_0^1 W_x(t)dt, \quad \tilde{C}_x \Rightarrow \int_0^1 W_x^2(t)dt,$$
$$\tilde{A}_y \Rightarrow \int_0^1 tW_y(t)dt, \quad \tilde{H}_y \Rightarrow \int_0^1 W_y(t)dt, \quad \tilde{C}_y \Rightarrow \int_0^1 W_y^2(t)dt, \quad \tilde{D} \Rightarrow \int_0^1 W_x(t)W_y(t)dt.$$

Compute the following quantities:

$$\begin{aligned}
&\tilde{\xi}_x = \sigma_{\epsilon x}\left(\tilde{A}_x - \frac{1}{2}\tilde{H}_x\right), \quad \tilde{\xi}_y = \sigma_{\epsilon y}\left(\tilde{A}_y - \frac{1}{2}\tilde{H}_y\right), \\
&\tilde{B}_{xx} = \sigma_{\epsilon x}^2\left(\tilde{C}_x - \tilde{H}_x^2\right), \quad \tilde{B}_{yy} = \sigma_{\epsilon y}^2\left(\tilde{C}_y - \tilde{H}_y^2\right), \quad \tilde{B}_{xy} = \sigma_{\epsilon x}\sigma_{\epsilon y}\left(\tilde{D} - \tilde{H}_x\tilde{H}_y\right).
\end{aligned} \tag{A.13}$$

Assumption 1 and Lemma 2 imply that

$$\tilde{\xi}_x \Rightarrow \xi_x, \quad \tilde{\xi}_y \Rightarrow \xi_y, \quad \tilde{B}_{xx} \Rightarrow B_{xx}, \quad \tilde{B}_{yy} \Rightarrow B_{yy}, \quad \tilde{B}_{xy} \Rightarrow B_{xy}.$$

Thus, the target quantities ξ_x, ξ_y, B_{xx}, B_{yy}, and B_{xy} can be approximated by (A.13).

A.4 Proof of Lemma 4

Proof of Lemma 4. Recall from (7) that $\hat{\sigma}_u^2 = n^{-1}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$. Rewrite this expression:

$$\begin{aligned}
\hat{\sigma}_u^2 &= n^{-1}\sum_{t=1}^{n}(y_t - \hat{\alpha} - \hat{\beta}x_t)^2 = n^{-1}\sum_{t=1}^{n}\left\{y_t - (\bar{y} - \hat{\beta}\bar{x}) - \hat{\beta}x_t\right\}^2 \\
&= n^{-1}\sum_{t=1}^{n}\left\{(y_t - \bar{y}) - \hat{\beta}(x_t - \bar{x})\right\}^2 = n^{-1}\sum_{t=1}^{n}\left\{(y_t - \bar{y})^2 - 2\hat{\beta}(y_t - \bar{y})(x_t - \bar{x}) + \hat{\beta}^2(x_t - \bar{x})^2\right\} \\
&= n^2\left\{n^{-3}\sum_{t=1}^{n}(y_t - \bar{y})^2 - 2\hat{\beta} \times n^{-3}\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x}) + \hat{\beta}^2 \times n^{-3}\sum_{t=1}^{n}(x_t - \bar{x})^2\right\} \\
&= n^2\left\{\lambda_n^{yy} - 2 \times \frac{\lambda_n^{xy}}{\lambda_n^{xx}} \times \lambda_n^{xy} + \left(\frac{\lambda_n^{xy}}{\lambda_n^{xx}}\right)^2 \times \lambda_n^{xx}\right\}
\end{aligned} \tag{A.14}$$

$$= n^2 \left\{ \lambda_n^{yy} - \frac{(\lambda_n^{xy})^2}{\lambda_n^{xx}} \right\} = n^2 \times \frac{\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2}{\lambda_n^{xx}}, \tag{A.15}$$

where (A.14) follows from (8), (9), and (21). Further, we have that

$$\hat{V}_x = n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 = n^2 \times n^{-3} \sum_{t=1}^{n} (x_t - \bar{x})^2 = n^2 \lambda_n^{xx}, \tag{A.16}$$

where the first equality follows from (7); the third equality follows from (9). Substitute (8), (A.15), and (A.16) into (6) to get

$$\hat{t}_\beta = \frac{n^{1/2}\lambda_n^{xy}}{\{\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2\}^{1/2}}.$$

Thus, the desired result is obtained:

$$\hat{t}_\beta^2 = \frac{n(\lambda_n^{xy})^2}{\lambda_n^{xx}\lambda_n^{yy} - (\lambda_n^{xy})^2}.$$

□