This is to certify that the

dissertation entitled

SPURIOUS REGRESSION WITH
FRACTIONALLY INTEGRATED PROCESSES

presented by

WEN-JEN TSAY

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

R. Baillie
Major professor

Date 10/25/1995

SPURIOUS REGRESSION WITH FRACTIONALLY INTEGRATED PROCESSES

By

Wen-Jen Tsay

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1995

ABSTRACT

SPURIOUS REGRESSION WITH
FRACTIONALLY INTEGRATED PROCESSES

By

Wen-Jen Tsay

This dissertation considers the spurious effect in a simple linear regression model of $I(d)$ processes. In Chapter 2 we find that when we regress a fractionally integrated process on a constant and another independent fractionally integrated process, spurious effects could arise. The most interesting finding is that the spurious effect could occur when both the dependent variable and regressor are stationary fractionally integrated processes. This implies the usual procedure to avoid the spurious effect by differencing may be questionable and empirical results based on such time series regressions may be misleading. In Chapter 3 we consider the asymptotic distributions of the regression coefficient estimators and the corresponding test statistics when the regressor and disturbance term are independent fractionally integrated processes. The main finding is that a long memory disturbance term could cause the null hypothesis to be overly rejected. The main conclusion of this dissertation is that careful study of the properties of the regressor and residuals are necessary before a regression is used. Otherwise, we could incorrectly find two independent time series to be correlated or two correlated time series to be independent, due to the persistence in data series.

ACKNOWLEDGMENTS

First of all, I would like to thank the members of my dissertation committee for their precious advice to help me prepare and complete this dissertation. Without their assistance the completion of this dissertation will not be possible. Especially I would like to thank Professor Richard T. Baillie, the committee chair, who inspires me to work on the subjects of this dissertation and provides me invaluable comments. Also I would like to thank Professor Peter Schmidt who reads my manuscript patiently and suggests numerous corrections and improvement. Moreover, my greatest debt of gratitude goes to Professor Ching-Fan Chung, who reads much of the near-final draft and offers substantial mathematical and editorial comments on it.

There are so many people who have been important for me during the whole program. I gratefully acknowledge Professor Christine E. Amsler who is especially generous to me when I was her Teaching Assistant. Also I am glad that I met Hailong Qian and Keshin Tswei who entertain my life in Michigan State University. Especially I am indebted to Hailong Qian for his patience and wisdom to help me solve many problems. I also want to thank Dan Hansen, Bill Horrace, Jess Reaser and Jim Zolnierek who have been my officemate and classmate. They are kind and teach me a lot of American culture. Without them, studying in my office room would not have been fun anymore.

Special appreciation also goes to the administrative staff in the department. Especially I acknowledge Mrs. Ann Feldman who is kind and offers me helpful advice to help me get through my program smoothly.

Finally, I would like to thank my wife, Ching-Yi Hsieh, for her sacrifice, understanding and encouragement. Without her help, it is impossible for me to finish the whole program. I also want to thank my parents for their love and support. Unfortunately, my father died two years ago. I could never repay the debt of gratitude I owe him. Also I wish to express my thanks to the family of my parents in law, they offer me a lot of help and I am indebted to them forever. My son, Kevin Tsay, who gives me the most wonderful joy, I would like to thank him for his presence in my life.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

It is a widely held belief that many data series in economics are $I(1)$ processes, or near $I(1)$ processes, as argued by Nelson and Plosser (1982). In recent years we also witness fast growing studies on fractionally integrated processes, or the $I(d)$ processes with the differencing parameter d being a fractional number. The $I(d)$ processes are natural generalizations of the $I(1)$ process that exhibit broader long-run characteristics. More specifically, the $I(d)$ processes can be either stationary or nonstationary, depending on the value of the fractional differencing parameter. The major characteristic of a stationary $I(d)$ process is its long memory which is reflected by the hyperbolic decay in its autocorrelations. A number of economic and financial series have been shown to posses long memory.

The $I(d)$ process is not the only model that displays the hyperbolic decay in its autocorrelations. The long range dependence in time series data can be traced back to Hurst (1951) who found the long term persistence in hydrology data which were referred to as a Hurst effect. The work of Hurst has attracted a lot of attention while the phenomenon of the hyperbolic decay in autocorrelations can be observed in many other fields. Many models have thus been proposed to characterize the long range persistence in time series. The two most famous models are the fractional Gaussian noise model proposed by Mandelbrot and Van Ness (1968) and the fractionally integrated process which is the subject of this dissertation.

A process Y_t is said to be a zero mean fractionally integrated autoregressive-moving average process of order p, d, q or ARFIMA (p, d, q) if it is defined by

$$\Phi(L)(1-L)^d(Y_t) = \Theta(L)\epsilon_t \tag{1.1}$$

where L is the lag operator, $\Phi(.)$ is a p^{th} order polynomial, d is the differencing parameter, $\Theta(.)$ is a q^{th} order polynomial, the roots of $\Phi(.)$ and $\Theta(.)$ are outside the unit circle, $\Phi(.)$ and $\Theta(.)$ have no common roots, ϵ_t is a white noise process with finite variance σ_ϵ^2 and the fractional differencing operator $(1-L)^d$ has the following binomial series expansion:

$$(1-L)^d = \sum_{j=0}^{\infty} \psi_j L^j \tag{1.2}$$

where

$$\psi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<\ell\leq j} \frac{\ell-1-d}{\ell}, \qquad j = 0, 1, 2, \ldots, \tag{1.3}$$

and $\Gamma(.)$ is the gamma function.

Given the above definition, we note a random walk is ARFIMA (0,1,0) and a zero mean fractionally integrated process of order d, I(d) is ARFIMA(0,d,0) and has the following form:

$$(1-L)^d Y_t = \epsilon_t. \tag{1.4}$$

This process is first introduced by Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981). They show that $I(d)$ is stationary and invertible when $d \in (-0.5, 0.5)$.

The main feature of the $I(d)$ process is that its autocovarance function declines at a slower hyperbolic rate (instead of the geometric rate found in the conventional ARMA models). When $d > 0$, the series are positively correlated and we call it a long memory process since it exhibits long range dependence in the sense of $\sum_{j=-\infty}^{\infty} r(j) = \infty$, where $r(j)$ is the autocovariance function of $I(d)$ at lag j. When $-0.5 < d < 0$, the series are negatively correlated and such that $\sum_{j=-\infty}^{\infty} | r(j) | < \infty$, and the process is sometime referred to as an intermediate memory process.

If $d \in (-0.5, 0.5)$, then an $I(d)$ process has the following AR(∞) and MA(∞) representations:

$$Y_t = \sum_{j=1}^{\infty} \phi_j Y_{t-j} + \epsilon_t \qquad j = 1, 2, 3, \ldots \tag{1.5}$$

where

$$\phi_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = -\psi_j, \tag{1.6}$$

and

$$Y_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} = (1-L)^{-d} \epsilon_t \tag{1.7}$$

where

$$\theta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{0<\ell\leq j} \frac{\ell - 1 + d}{\ell}, \qquad j = 0, 1, 2, \ldots. \tag{1.8}$$

Let us denote the spectral density, autocorrelation function and partial autocorrelation of a stationary $I(d)$ process by $f(.)$, $\rho(.)$ and $\alpha(.)$, respectively, then

$$f(\lambda) = | 1 - e^{-i\lambda} |^{-2d} \frac{\sigma_\epsilon^2}{2\pi} = | 2\sin(\lambda/2) |^{-2d} \frac{\sigma_\epsilon^2}{2\pi}, \quad -\pi \leq \lambda \leq \pi, \tag{1.9}$$

$$\rho(j) = \frac{\Gamma(j+d)\Gamma(1-d)}{\Gamma(j-d+1)\Gamma(d)} = \prod_{0<\ell\leq j} \frac{\ell - 1 + d}{\ell - d}, \qquad j = 1, 2, \ldots, \tag{1.10}$$

and

$$\alpha(j) = \frac{d}{j-d}, \qquad j = 1, 2, \ldots. \tag{1.11}$$

Applying Stirling's formula to (6), (8) and (11) gives that, as $j \to \infty$,

$$\phi_j \sim \frac{1}{\Gamma(-d)} j^{-d-1}, \tag{1.12}$$

$$\theta_j \sim \frac{1}{\Gamma(d)} j^{d-1}, \tag{1.13}$$

and

$$\rho(j) \sim \frac{\Gamma(1-d)}{\Gamma(d)} j^{2d-1}, \tag{1.14}$$

where $a_j \sim b_j$ means $\lim_{j\to\infty} a_j/b_j = 1$. Moreover, we note $\sin\lambda \sim \lambda$ as $\lambda \to 0$. Therefore,

$$f(\lambda) \sim \lambda^{-2d}, \tag{1.15}$$

as $\lambda \to 0$. This result suggests that the spectral density of a long memory process has a singularity at 0 frequency.

This dissertation considers the spurious effect in a simple linear regression of $I(d)$ processes. We first study the spurious effect when we regress a fractionally integrated process on a constant and another independent fractionally integrated process and then investigate the spurious effect when the disturbance term and regressor are fractionally integrated processes.

This dissertation is organized as follows: In Chapter 2 we extend Granger and Newbold's (1974) work to study the spurious effect of regressing a fractionally integrated process on a constant and another independent fractionally integrated process. We also examine the spurious effect of detrending an $I(d)$ process. We specify the conditions under which the spurious effect can occur. An important implication from our results is that the usual procedure to avoid spurious effect by differencing may be questionable.

In Chapter 3 we show the spurious effect when the dependent variable and the regressor are correlated while both regressor and disturbance term are $I(d)$ processes. We consider the issue of the statistical inference regarding the regression coefficient estimators. The conclusion is that many empirical testing results should be interpreted more carefully because the usual t tests may not be valid when long memory is present in the disturbance term.

The Conclusion and some extensions are presented in Chapter 4.

CHAPTER 2

THE SPURIOUS REGRESSION OF FRACTIONALLY INTEGRATED PROCESSES

2.1. Introduction

The spurious regression was first studied by Granger and Newbold (1974) using simulation. They show that when unrelated data series are integrated processes of order 1 or the $I(1)$ processes, then running a regression with this type of data will yield spurious effects. That is, the null hypothesis of no relationship among the unrelated $I(1)$ processes will be rejected much too often. Furthermore, the spurious regression tends to yield a high coefficient of determination (R^2) as well as highly autocorrelated residuals, indicated by a very low value of Durbin-Watson (DW) statistic. Granger and Newbold's simulation results are later supported by Phillips's (1986) theoretical analysis. Phillips proves that the usual t test statistic in a spurious regression does not have a limiting distribution but diverges as the sample size (T) approaches infinity. He also shows that R^2 has a non-degenerate limiting distribution while the DW statistic converges in probability to zero.

The history of the research on spurious detrending follows a similar thread. Nelson and Kang (1981, 1984) first employ simulation to demonstrate that the regression of a driftless $I(1)$ process on a time trend produces an incorrect result of a significant trend. Extending the Phillips's (1986) approach, Durlauf and Phillips (1988) derive the asymptotic distributions for the least squares estimators in such a regression. In particular, the latter authors show that the t test statistics diverge and there are no correct critical values for the conventional significance tests.

All these studies of the spurious regression concentrate on the nonstationary $I(1)$ processes. It reflects the widely held belief that many data series in economics are $I(1)$ processes, or near $I(1)$ processes, as argued by Nelson and Plosser (1982). Against this backdrop, we also witness in recent years fast growing studies on fractionally integrate processes, or the $I(d)$ processes with the differencing parameter d being a fractional number. The $I(d)$ processes are natural generalization of the $I(1)$ processes that exhibit a broader long-run characteristics. More specifically, the $I(d)$ processes can be either stationary or nonstationary, depending on the value of the fractional differencing parameter. The major characteristic of a stationary $I(d)$ process is its long memory which is reflected by the hyperbolic decay in its autocorrelations. A number of economic and financial series have been shown to possess long memory. See Baillie (1995) for an updated survey on the applications of the $I(d)$ processes in economics and finance.

The objective of this chapter is to extend the theoretical analysis of the spurious regression from $I(1)$ processes to the class of long memory $I(d)$ processes. We establish and analyze conditions on the $I(d)$ processes that inflict the spurious effect in a simple linear regression model. The nonstandard asymptotic distributions of various coefficient estimators and test statistics are then derived.

The main finding from our study is that the spurious regression can arise among a wide range of long memory $I(d)$ processes, even in cases where both dependent variable and regressor are stationary. A few conclusions may then be drawn. First, different from what

Phillips (1986) and Durlauf and Phillips (1988) have suggested, the cause for spurious effects seems to be neither nonstationarity nor lack of ergodicity but the strong long memory in the data series. As a result, spurious effects might occur more often than we previously believed as they can arise even among stationary series. Furthermore, the usual first-differencing procedure may not be able to completely eliminate spurious effects if the data series are not only nonstationary but possess strong long memory (such as in the case where they are $I(d)$ processes with $d > 1$).

2.2. A General Theory of Spurious Effects

Our analysis of the spurious effects are based on several simple linear regression models in which the dependent variable and the single non-constant regressor are independent $I(d)$ processes with d lying in different ranges. Before presenting these models, let's first briefly review some basic properties of the $I(d)$ processes.

A process Y_t is said to be a fractionally integrated process of order d, denoted as $I(d)$, if it is defined by

$$(1 - L)^d Y_t = \epsilon_t,$$

where L is the usual lag operator, d is the differencing parameter which can be a fractional number, and the innovation sequence ϵ_t is white noise with a zero mean and finite variance. The fractional differencing operator $(1 - L)^d$ is defined as follows:

$$(1 - L)^d = \sum_{j=0}^{\infty} \psi_j L^j, \qquad \text{where} \quad \psi_j = \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)},$$

and $\Gamma(\cdot)$ is the gamma function. This process is first introduced by Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981). They show that Y_t is stationary when $d < 0.5$ and is invertible when $d > -0.5$.

The main feature of the $I(d)$ process is that its autocovariance function declines at a slower hyperbolic rate (instead of the geometric rate found in the conventional ARMA models):

$$\gamma(j) = O(j^{2d-1}),$$

where $\gamma(j)$ is the autocovariance function at lag j. When $d > 0$, the $I(d)$ process is said to have long memory since it exhibits long range dependence in the sense that $\sum_{j=-\infty}^{\infty} \gamma(j) = \infty$. When $d < 0$, then $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$ and the process is sometimes referred to as an intermediate memory process.

Our analysis focuses on the class of long memory $I(d)$ processes with $d > 0$. We are particularly interested in the distinction between the nonstationary subclass of $I(d)$ processes with $d \geq 0.5$ and the stationary subclass with $d < 0.5$. To examine potentially different types of spurious effects, we propose six regression models for different classes of $I(d)$ processes, mainly based on whether the fractional differencing parameter d is greater than 0.5 or not. The exact specifications of these models can be conveniently expressed with four $I(d)$ processes. Let's first define two stationary ones with different differencing parameters d_1 and d_2 whose values lie between -0.5 and 0.5:

$$(1-L)^{d_1} v_t = a_t \qquad \text{and} \qquad (1-L)^{d_2} w_t = b_t,$$

where a_t and b_t are two white noises with zero mean and finite variances σ_a^2 and σ_b^2, respectively; that is, v_t and w_t are $I(d_1)$ and $I(d_2)$

processes, respectively, and both of them are stationary and invertible. When these two processes are employed in our later analysis, the values of their differencing parameters are mostly assumed to be in (0, 0.5); i.e., the stationary processes v_t and w_t are often assumed to have long memory. We can also define two nonstationary $I(1+d_1)$ and $I(1+d_2)$ processes by integrating v_t and w_t:

$$y_t = y_{t-1} + v_t \qquad \text{and} \qquad x_t = x_{t-1} + w_t.$$

Obviously, the orders of integration of these two nonstationary fractionally integrated processes lie between 0.5 and 1.5. Given these four fractionally integrated processes, we consider the following six simple linear regression models:

Model 1: $y_t = \alpha + \beta x_t + u_t,$

Model 2: $v_t = \alpha + \beta w_t + u_t,$ where $d_1 + d_2 > 0.5,$

Model 3: $y_t = \alpha + \beta w_t + u_t,$ where $d_2 > 0,$

Model 4: $v_t = \alpha + \beta x_t + u_t,$ where $d_1 > 0,$

Model 5: $y_t = \alpha + \beta t + u_t,$

Model 6: $v_t = \alpha + \beta t + u_t,$ where $d_1 > 0.$

In Model 1 the orders of integration of both the dependent variable and the regressor lie between 0.5 and 1.5, and can be equal to 1. So Model 1 may be considered a generalization of Phillips' (1986) spurious regression to the case of fractionally integrated processes. Model 2 presents the most interesting case in our analysis. In it both the dependent variable and the regressor are assumed to be stationary, ergodic, and strongly persistent in the sense that their fractional differencing parameters sum up to a value greater than 0.5. Following Phillips' arguments, we tend to think no spurious effect should occur in such a model where variables are ergodic. But our analysis of Model 2 presents a result to the contrary. The analysis of Model 2 seems to go beyond the previous study of spurious effects and allow us to gain new insight into the problem.

Models 3 and 4 differ from Model 1 in that the order of integration in one of the dependent variable and the regressor is reduced to the stationary range between 0 and 0.5. We can conveniently view Models 3 and 4 as two intermediate cases between Model 1 of nonstationary fractionally integrated processes and Model 2 of stationary fractionally integrated processes. We thus expect the analysis of these two new models to be a mixture of those of Models 1 and 2.

In Models 5 and 6 we consider the effect of detrending the nonstationary and stationary fractionally integrated processes, respectively. Through these two models, we generalize the results of Durlauf and Phillips (1988). Also, Models 5 and 6 can be regarded as variants of Models 1 and 4, respectively, with the nonstationary regressor x_t replaced by the time trend. This similarity in the model specifications will also be reflected in their analytic results.

The following assumption on the two white noise processes a_t and b_t are made throughout this paper to simplify our analysis.

Assumption 1. *The two white noises a_t and b_t are independently and identically distributed with zero means, and their moments satisfy the following conditions:* $\mathrm{E}|a_t|^p < \infty$, *with* $p \geq \max\{4,\ -8d_1/(1+2d_1)\}$;

and $\mathrm{E}|\mathrm{b_t}|^{\mathrm{q}} < \infty$ *with* $q \geq \max\{4,\ -8d_2/(1+2d_2)\}$*. Moreover,* a_t *and* b_t *are independent of each other.*

We also assume, without loss of generality, that the initial values of the fractionally integrated processes v_0, w_0, y_0, and x_0 are all zero. Hence, y_t and x_t can be considered as the partial sums of v_t and w_t, respectively; i.e., $y_T = \sum_{t=1}^{T} v_t$ and $x_T = \sum_{t=1}^{T} w_t$.

Before presenting Lemma 1, which is the cornerstone of our analysis, let's summarize two important asymptotic results on the partial sums y_t and x_t. First, given the variances of y_T and x_T:

$$\sigma_y^2 = \mathrm{Var}(\mathrm{y_T}) = \mathrm{Var}\left(\sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{v_t}\right) \quad \text{and} \quad \sigma_{\mathrm{x}}^2 = \mathrm{Var}(\mathrm{x_T}) = \mathrm{Var}\left(\sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{w_t}\right)$$

Sowell (1990, Theorem 1) proves that

$$\sigma_y^2 = O(T^{1+2d_1}) \qquad \text{and} \qquad \sigma_x^2 = O(T^{1+2d_2}).$$

Furthermore, Davydov (1970) shows that as $T \to \infty$,

$$\frac{1}{\sigma_y} y_{[Tr]} \Rightarrow B_{0.5+d_1}(r) \qquad \text{and} \qquad \frac{1}{\sigma_x} x_{[Tr]} \Rightarrow B_{0.5+d_2}(r),$$

for $r \in [0,\ 1]$, where $[Tr]$ denotes the integer part of Tr, the notation $\Rightarrow$ denotes weak convergence, and $B_{0.5+d}(t)$ is the fractional Brownian motion which is defined by the following stochastic integral

$$B_{0.5+d}(t) \equiv \frac{1}{\Gamma(1+d)} \int_0^t (t-s)^d \, dB_{0.5}(s), \qquad \text{for } d \in (-0.5,\ 0.5),$$

where $B_{0.5}(t)$ is the standard Brownian motion. See Mandelbrot and Van Ness (1968). Our notation for the standard and the fractional versions of Brownian motions suggests that the former is a special case of the latter with $d = 0$. These two well-known results help establishing the following lemma.

Lemma 1. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results:*

1. $$\frac{1}{T}\sum_{t=1}^{T}\frac{y_t}{\sigma_y} \Rightarrow \int_0^1 B_{0.5+d_1}(s)\,ds \qquad \text{and}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_t}{\sigma_x} \Rightarrow \int_0^1 B_{0.5+d_2}(s)\,ds.$$

2. $$\frac{1}{T}\sum_{t=1}^{T}\frac{y_t^2}{\sigma_y^2} \Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\,ds \qquad \text{and}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_t^2}{\sigma_x^2} \Rightarrow \int_0^1 [B_{0.5+d_2}(s)]^2\,ds.$$

3. $$\frac{1}{T}\sum_{t=1}^{T}\frac{(y_t-\overline{y})^2}{\sigma_y^2} \Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_1}(s)\,ds\right]^2,$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{(x_t-\overline{x})^2}{\sigma_x^2} \Rightarrow \int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2.$$

4. $$\sum_{t=1}^{T}\frac{v_t}{\sigma_y} \Rightarrow B_{0.5+d_1}(1) \qquad \text{and} \qquad \sum_{t=1}^{T}\frac{w_t}{\sigma_x} \Rightarrow B_{0.5+d_2}(1).$$

5. $$\frac{1}{T}\sum_{t=1}^{T}v_t^2 \xrightarrow{p} \gamma_v(0) = \frac{\Gamma(1-2d_1)}{\Gamma(1-d_1)^2}\sigma_a^2 . \qquad \text{and}$$

$$\frac{1}{T}\sum_{t=1}^{T} w_t^2 \xrightarrow{p} \gamma_w(0) = \frac{\Gamma(1-2d_2)}{\Gamma(1-d_2)^2}\sigma_b^2.$$

6. $$\frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2 \xrightarrow{p} \gamma_v(0) \qquad \text{and}$$

$$\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2 \xrightarrow{p} \gamma_w(0).$$

7. $$\sum_{t=1}^{T} v_t w_t \begin{cases} = O_p(T^{d_1+d_2}), & \text{if } d_1 + d_2 \geq 0.5, \\ \leq O_p(T^{0.5}), & \text{otherwise,} \end{cases}$$

or, equivalently,

$$\sum_{t=1}^{T} \frac{v_t}{\sigma_y} \cdot \frac{w_t}{\sigma_x} = \begin{cases} o_p(T^{\epsilon-1}), & \text{if } d_1 + d_2 \geq 0.5, \\ o_p(T^{\epsilon-0.5-d_1-d_2}), & \text{otherwise,} \end{cases}$$

for any $\epsilon > 0$.

8. $$\frac{1}{T}\sum_{t=1}^{T} \frac{y_t}{\sigma_y} \cdot \frac{x_t}{\sigma_x} \Rightarrow \int_0^1 B_{0.5+d_1}(s) \cdot B_{0.5+d_2}(s)\, ds.$$

9. $$\sum_{t=1}^{T} \frac{y_t}{\sigma_y} \cdot \frac{w_t}{\sigma_x} \Rightarrow \int_0^1 B_{0.5+d_1}(s)\, dB_{0.5+d_2}(s), \quad \text{for } d_2 > 0,$$

$$\sum_{t=1}^{T} \frac{v_t}{\sigma_y} \cdot \frac{x_t}{\sigma_x} \Rightarrow \int_0^1 B_{0.5+d_2}(s)\, dB_{0.5+d_1}(s), \quad \textit{for } d_1 > 0.$$

10. $$\sum_{t=1}^{T} \frac{t}{T} \cdot \frac{v_t}{\sigma_y} \Rightarrow B_{0.5+d_1}(1) - \int_0^1 B_{0.5+d_1}(s)\, ds \qquad \text{and}$$

$$\sum_{t=1}^{T} \frac{t}{T} \cdot \frac{w_t}{\sigma_x} \Rightarrow B_{0.5+d_2}(1) - \int_0^1 B_{0.5+d_2}(s)\, ds.$$

11. $$\frac{1}{T} \sum_{t=1}^{T} \frac{t}{T} \cdot \frac{y_t}{\sigma_y} \Rightarrow \int_0^1 s \cdot B_{0.5+d_1}(s)\, ds \qquad \text{and}$$

$$\frac{1}{T} \sum_{t=1}^{T} \frac{t}{T} \cdot \frac{x_t}{\sigma_x} \Rightarrow \int_0^1 s \cdot B_{0.5+d_2}(s)\, ds.$$

Here, $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions, $\gamma_v(j)$ and $\gamma_w(j)$ are the autocovariance functions of v_t and w_t, respectively, at lag j, and σ_a^2 and σ_b^2 are the variances of the underlying white noises a_t and b_t, respectively. The notation $\xrightarrow{p}$ means convergence in probability.

All the theorem proofs are in the Mathematical Proof. In the rest of this section the results of Lemma 1 will be used to develop the theory of spurious effects, presented in a series of theorems and corollaries, for the proposed six models. The first two models will be discussed separately in subsections 2.1 and 2.2. These two models provide us with a framework which facilitates the explanations of the other four models in subsections 2.3 and 2.5. One subsection – subsection 2.4 – will be devoted to the analysis of an important issue about how the orders of integration of the fractionally integrated processes are directly related to the spurious effects.

We will adopt the following notation for the various statistics from the Ordinary Least Squares (OLS) estimation. Let $\widehat{\alpha}$ and $\widehat{\beta}$ denote the usual OLS estimators of the intercept and the slope. Their respective variances are estimated by s_β^2 and s_α^2, from which we have the t ratios $t_\beta = \widehat{\beta}/s_\beta$ and $t_\alpha = \widehat{\alpha}/s_\alpha$. Also, let s^2 denote the estimated variance of the OLS residuals, R^2 the coefficient of determination, and DW the Durbin-Watson statistic. Finally, in addition to the autocovariance functions $\gamma_v(j)$ and $\gamma_w(j)$ of v_t and w_t, let $\rho_v(j)$ and $\rho_w(j)$ be their respective autocorrelations at lag j.

2.2.1. Model 1 of Nonstationary Fractionally Integrated Processes

In Model 1 a nonstationary $I(1+d_1)$ process y_t is regressed on another independent and nonstationary $I(1+d_2)$ process x_t:

$$y_t = \alpha + \beta x_t + u_t.$$

Since the permissible range for the values of the fractional differencing parameters d_1 and d_2 is $(-0.5, 0.5)$, Model 1 generalizes Phillips' (1986) model of integrated processes in which $d_1 = d_2 = 0$. Unsurprisingly, all the results we derive for Model 1 are straightforward generalization of Phillips' theory of the spurious effects. The results for Model 1 are presented in the following theorem:

Theorem 1. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results:*

1. $\frac{\sigma_x}{\sigma_y}\widehat{\beta} \Rightarrow$

$$\frac{\int_0^1 B_{0.5+d_1}(s)\cdot B_{0.5+d_2}(s)\,ds - \left[\int_0^1 B_{0.5+d_1}(s)ds\right]\left[\int_0^1 B_{0.5+d_2}(s)ds\right]}{\int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2}$$

$\equiv \beta_*.$

Note that $\sigma_y/\sigma_x = O(T^{d_1-d_2})$.

2. $$\frac{1}{\sigma_y}\widehat{\alpha} \Rightarrow \int_0^1 B_{0.5+d_1}(s)\,ds - \beta_*\int_0^1 B_{0.5+d_2}(s)\,ds \equiv \alpha_*,$$

where β_* *is defined in 1. Note that* $\sigma_y = O(T^{0.5+d_1})$.

3. $$\frac{1}{\sigma_y^2}s^2 \Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_1}(s)\,ds\right]^2$$

$$-\beta_*^2\left\{\int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2\right\} \equiv \sigma_*^2,$$

where β_* *is defined in 1. Note that* $\sigma_y^2 = O(T^{1+2d_1})$.

4. $$\frac{T\sigma_x^2}{\sigma_y^2}s_\beta^2 \Rightarrow \frac{\sigma_*^2}{\int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2} \equiv \sigma_{*\beta}^2,$$

where σ_*^2 is defined in 3. Note that $\sigma_y^2/T\sigma_x^2 = O(T^{2d_1-2d_2-1})$.

5. $$\frac{T}{\sigma_y^2} s_\alpha^2 \Rightarrow \sigma_*^2 \left\{ 1 + \frac{\left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2} \right\}$$

$$\equiv \sigma_{*\alpha}^2,$$

where σ_*^2 is defined in 3. Note that $\sigma_y^2/T = O(T^{2d_1})$.

6. $$\frac{1}{\sqrt{T}} t_\beta \Rightarrow \frac{\beta_*}{\sigma_{*\beta}},$$

where β_* is defined in 1 and $\sigma_{*\beta}^2$ is defined in 4.

7. $$\frac{1}{\sqrt{T}} t_\alpha \Rightarrow \frac{\alpha_*}{\sigma_{*\alpha}},$$

where α_* is defined in 2 and $\sigma_{*\alpha}^2$ is defined in 5.

8. $$R^2 \Rightarrow \frac{\beta_*^2 \left\{ \int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2 \right\}}{\int_0^1 [B_{0.5+d_1}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_1}(s)\, ds\right]^2},$$

where β_* is defined in 1.

9. $DW \xrightarrow{p} 0.$

Here, $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions.

The most important result in Theorem 1 is that, as the sample size T increases, the two t ratios t_β and t_α diverge at the same rate of $\sqrt{T}$, which is independent of the magnitudes of the fractional differencing parameters d_1 and d_2. This result is exactly the same as what Phillips (1986) has obtained for the case where $d_1 = d_2 = 0$. So even when the orders of integration in the dependent variable and the regressor differ from 1 by as much as 0.5, the usual problem in using the t tests remains: the probability of rejecting the null hypothesis of $\beta = 0$ or $\alpha = 0$ based on t tests increases monotonically as the sample size increases.

The limiting distributions of the t ratios, after normalized by $\sqrt{T}$, are direct generalization of those derived by Phillips. The same conclusion also holds for R^2 and the DW statistics. In other words, when we compare our results with Phillips', we observe a common feature in these four statistics; namely, the nonzero values of d_1 and d_2 do not affect their convergence rates while the effects on their limiting distributions are quite straightforward: all the standard Brownian motions in Phillips' theory are replaced by fractional Brownian motions. That the fractional differencing parameters d_1 and d_2 play a relative minor role here is mainly because the four statistics are all ratios so that the effects of d_1 and d_2 are canceled out. In contrast, the results on the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$ are a different story. In Phillips' theory both $\widehat{\beta}$ and $\widehat{\alpha}/\sqrt{T}$ converge to some non-normal non-degenerate limiting distributions. But for the present model of the fractionally integrated processes, the orders of $\widehat{\beta}$ and $\widehat{\alpha}$ are $T^{d_1-d_2}$ and

$T^{d_1+0.5}$, respectively. So while $\widehat{\alpha}$ always diverges (though the rate can be slow if d_1 is close to -0.5), $\widehat{\beta}$ can be either divergent or convergent, depending on the relative magnitudes of d_1 and d_2. For example, if the order of integration in the dependent variable y_t is smaller than that of the regressor x_t; i.e., $d_1 < d_2$, then $\widehat{\beta}$ converges to zero, just like the conventional case of no spurious effects. Moreover, if $d_1 - d_2 = -0.5$, then, similar to the case of no spurious effects, $\sqrt{T} \cdot \widehat{\beta}$ has a limiting distribution, though its limiting distribution is not normal.

2.2.2. Model 2 of Stationary Fractionally Integrated Processes

In this section we consider Model 2 in which a stationary fractionally integrated process v_t is regressed on an independent and stationary fractionally integrated process w_t:

$$v_t = \alpha + \beta w_t + u_t.$$

We show that, although both v_t and w_t are stationary, the spurious effect in terms of the t tests could still exist under an additional condition on the fractional differencing parameters: $d_1 + d_2 > 0.5$. Loosely speaking, this condition implies that the two processes v_t and w_t are both strongly persistent.

Our analysis begins with a special case where we assume a set of more stringent conditions which helps deriving the exact limiting distributions for the various OLS estimators. This theory is based on an important result of Fox and Taqqu (1987) who show that the product of two highly persistent but stationary Gaussian processes, if adequately normalized, can converge. After examining this special case, we then show how the spurious effects may still exist in a more general framework even though the exact limiting distributions cannot be readily defined in such a case.

Let's first reproduce Fox and Taqqu's (1987) Theorem 6.1 here as Lemma 2.

Lemma 2. *Let* (X_t, Y_t) *be a stationary jointly Gaussian sequence with* $\mathrm{E}(\mathrm{X_t}) = \mathrm{E}(\mathrm{Y_t}) = 0$, $\mathrm{E}(\mathrm{X_t^2}) = \mathrm{E}(\mathrm{Y_t^2}) = 1$, *and* $\mathrm{E}(\mathrm{X_t Y_t}) = \mathrm{r}$. *Suppose that* σ_1 *and* σ_2 *are two arbitrary real numbers and that there exist* $0 < \delta_1, \delta_2 < 0.5$, *such that as* $j \to \infty$

$$\mathrm{E}(\mathrm{X_t X_{t+j}}) \sim \sigma_1^2 \cdot j^{-\delta_1}, \qquad \mathrm{E}(\mathrm{X_t Y_{t+j}}) \sim \frac{\rho\sigma_1\sigma_2 \mathrm{b_1}}{\sqrt{\mathrm{a_1 a_2}}} j^{-(\delta_1+\delta_2)/2},$$

$$\mathrm{E}(\mathrm{Y_t Y_{t+j}}) \sim \sigma_2^2 \cdot j^{-\delta_2}, \qquad \mathrm{E}(\mathrm{Y_t X_{t+j}}) \sim \frac{\rho\sigma_1\sigma_2 \mathrm{b_2}}{\sqrt{\mathrm{a_1 a_2}}} j^{-(\delta_1+\delta_2)/2},$$

where ρ *is a constant between 1 and* -1, *while* $a_1 = A(\delta_1, \delta_1)$, $a_2 = A(\delta_2, \delta_2)$, $b_1 = A(\delta_1, \delta_2)$, *and* $b_2 = A(\delta_2, \delta_1)$ *are four constants with* $A(\delta_1, \delta_2)$ *being defined by* $\int_0^\infty x^{-(\delta_1+1)/2}(x+1)^{-(\delta_2+1)/2} dx$, *then*

$$\frac{1}{T^{1-(\delta_1+\delta_2)/2}} \sum_{t=1}^{[Ts]} (X_t Y_t - r) \;\Rightarrow\; Z(s),$$

where $Z(s)$ *is*

$$\frac{\sigma_1\sigma_2}{\sqrt{a_1 a_2}} \int_{\Re^2} \int_0^s \left[\prod_{i=1}^{2} (u - x_i)^{-(\delta_i+1)/2} I_{\{x_i < u\}} \right] du \, dM_1(x_1) \, dM_2(x_2).$$

Here, M_1 *and* M_2 *are two Gaussian random measures with respect to Lebesgue measure, having unit variances and covariance* ρ.

Note that the two processes X_t and Y_t are not only strongly persistent, as indicated by the hyperbolic convergence rates δ_1 and δ_2

in their autocorrelations, but also highly correlated with each other, as indicated by the hyperbolic convergence rates in their covariances. However, in our application we are only interested in the case where X_t and Y_t are independent so that r and ρ in the above lemma are both zero. The above lemma offers us the convergence rate of $\sum_{t=1}^{T} X_t Y_t$ and its limiting process $Z(t)$ given the Gaussian assumption and a narrower range for the parameters δ_1 and δ_2. In order to apply this lemma, we make the following assumption in addition to Assumption 1 we have made earlier.

Assumption 2. *The two fractionally integrated processes v_t and w_t are both Gaussian and the corresponding fractional differencing parameters d_1 and d_2 are both in the range of* (0.25, 0.5).

Given the facts that

$$\rho_v(j) \sim \frac{\Gamma(1-d_1)}{\Gamma(d_1)} j^{2d_1-1} \qquad \text{and} \qquad \rho_w(j) \sim \frac{\Gamma(1-d_2)}{\Gamma(d_2)} j^{2d_2-1},$$

it is straightforward to prove the following corollary in which X_t and Y_t in Lemma 2 are replaced by $v_t/\sqrt{\gamma_v(0)}$ and $w_t/\sqrt{\gamma_w(0)}$, respectively.

Corollary 1. *Given that Assumptions 1 and 2 hold, then, as $T \to \infty$,*

$$\frac{1}{T^{d_1+d_2}} \sum_{t=1}^{T} \frac{v_t}{\sqrt{\gamma_v(0)}} \cdot \frac{w_t}{\sqrt{\gamma_w(0)}} \Rightarrow Z(1),$$

where the limiting random variable $Z(1)$ is defined in Lemma 2 with $\delta_1 = 1-2d_1$, $\delta_2 = 1-2d_2$, $\sigma_1^2 = \Gamma(1-d_1)/\Gamma(d_1)$, and $\sigma_2^2 = \Gamma(1-d_2)/\Gamma(d_2)$.

With Lemma 1 and Corollary 1, we can then establish the following theorem about the spurious effect in Model 2.

Theorem 2. *Given that Assumptions 1 and 2 hold, then, as* $T \to \infty$, *we have the following results:*

1. $$\frac{T^2}{\sigma_y \sigma_x} \widehat{\beta} \Rightarrow \sqrt{\frac{\gamma_v(0)}{\gamma_w(0)}} \left[Z(1) - \frac{1}{\sqrt{\gamma_v(0) \cdot \gamma_w(0)}} B_{0.5+d_1}(1) \cdot B_{0.5+d_2}(1) \right].$$

   Note that $\sigma_y \sigma_x / T^2 = O(T^{d_1+d_2-1})$.

2. $$\frac{T}{\sigma_y} \widehat{\alpha} \Rightarrow B_{0.5+d_1}(1).$$

   Note that $\sigma_y / T = O(T^{d_1-0.5})$.

3. $$s^2 \xrightarrow{p} \gamma_v(0).$$

4. $$T \cdot s_\beta^2 \xrightarrow{p} \frac{\gamma_v(0)}{\gamma_w(0)}.$$

5. $$T \cdot s_\alpha^2 \xrightarrow{p} \gamma_v(0).$$

6. $$\frac{T^{3/2}}{\sigma_y \sigma_x} t_\beta \Rightarrow Z(1) - \frac{1}{\sqrt{\gamma_v(0) \cdot \gamma_w(0)}} B_{0.5+d_1}(1) \cdot B_{0.5+d_2}(1).$$

   Note that $\sigma_y \sigma_x / T^{3/2} = O(T^{d_1+d_2-0.5})$.

7. $$\frac{\sqrt{T}}{\sigma_y} t_\alpha \Rightarrow \frac{1}{\sqrt{\gamma_v(0)}} B_{0.5+d_1}(1).$$

Note that $\sigma_y/\sqrt{T} = O(T^{d_1})$.

8. $$\frac{T^4}{\sigma_y^2 \sigma_x^2} R^2 \Rightarrow \left[Z(1) - \frac{1}{\sqrt{\gamma_v(0) \cdot \gamma_w(0)}} B_{0.5+d_1}(1) \cdot B_{0.5+d_2}(1) \right]^2.$$

Note that $\sigma_y^2 \sigma_x^2 / T^4 = O(T^{2d_1+2d_2-2})$.

9. $$DW \xrightarrow{p} 2 - 2\rho_v(1) = \frac{2(1-2d_1)}{1-d_1}.$$

Here $B_{0.5+d_1}(t)$ *and* $B_{0.5+d_2}(t)$ *are two independent fractional Brownian motions, and* $Z(1)$ *is a random variable defined in Corollary 1.*

The most important result from this theorem is the divergence rates of the two t ratios t_β and t_α, which are $T^{d_1+d_2-0.5}$ and T^{d_1}, respectively. Recall that $d_1 + d_2 - 0.5$ is necessarily greater than 0 (and smaller than 0.5) under Assumption 2. This result reflects the spurious effect in the t tests. Since both the dependent variable and the regressor are stationary and ergodic, the spurious effect is not really expected (see Phillips 1986, p.318). The surprising results we get here suggest that the cause for the spurious effect has more to do with the strong persistence than stationarity and ergodicity of the variables involved.

It is interesting to compare the divergence rates of the t ratios here with the $\sqrt{T}$ rate we observe in Model 1. We note that the divergence rates in the present model depend on the magnitudes of the two fractional differencing parameters d_1 and d_2 while those in Model 1 do not. Furthermore, the t ratios diverge more slowly in the present model than in Model 1. In particular, the divergence rate of t_β can become very slow when both d_1 and d_2 approach to their lower boundary 0.25.

Let's turn to the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$. We note that both of them converge in probability to zero as in the conventional case of no spurious effects. However, their convergence rates are much slower than the usual $T^{-1/2}$ rate and, if they are normalized appropriately, their limiting distributions are not standard normal either. In contrast to these irregular convergence rates of the OLS estimators, the estimated variances s_β^2 and s_α^2 nevertheless converge at the standard T^{-1} rate. It is such disparity in the convergence rates between the OLS estimators, which converge at rates slower than $T^{-1/2}$, and their standard errors, which converge at the standard $T^{-1/2}$ rates, that causes the resulting t ratios to diverge and hence the spurious effect.

R^2 in the present model converges to 0 as in the case of no spurious effects. It is different from what we observe in Model 1 where R^2 converges to a random variable. Consequently, as the sample size increases, the declining R^2 in the present model will correctly reflect the fact that the regressor does not help explain the variations in the dependent variable.

The DW statistic does not converge in probability to zero and this result is also different from that of Model 1. Its limit $2-2\rho_v(1)$ is similar to the one we find in the conventional AR(1) case. This limit depends on the fractional differencing parameter d_1 of the dependent variable v_t and can only take value in the range of (0, 4/3), which is to the left of the value 2.

We now consider a less restricted specification of Model 2 which is defined by the following assumption.

Assumption 2A. *The sum of the two fractional differencing parameters d_1 and d_2 is greater than 0.5.*

Since the Gaussian distribution is not assumed while one of the fractional differencing parameters d_1 and d_2 can be smaller than 0.25, Assumption 2A is thus considerably less stringent than Assumption 2. The price we pay for such generality is that we are not able to express the limiting distribution of some statistics in closed form as indicated in the following corollary.

Corollary 2. *If Assumption 2 is replaced by Assumption 2A in Theorem 2, then all the conclusions there remain true except that, while $T^2\widehat{\beta}/\sigma_y\sigma_x$, $T^{3/2}t_\beta/\sigma_y\sigma_x$, and $T^4R^2/\sigma_y^2\sigma_x^2$ still converge weakly, the exact specifications of their limits are unknown. In particular, the process $Z(1)$ in Theorem 2 will be replaced by a process of an unknown form.*

The main finding in this corollary is that, even though the limiting distributions of some statistics are not readily expressible, all the discussions of the spurious effects following Theorem 2 still apply to the more general specification of Model 2. The analysis of Model 2 can thus be summarized as follows. The OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$ (as well as R^2) do converge in probability to zero, correctly reflecting the lack of a relationship between the dependent variable and the regressor. But the convergence rates of $\widehat{\beta}$ and $\widehat{\alpha}$ are too slow in comparison with those of their standard errors. Consequently, the t ratios diverge and the t tests fail. The upshot is that the usual t tests can become invalid even when the dependent variable and the regressor are both stationary and ergodic (so long as they are sufficiently persistent).

A profound implication from Model 2 is as follows: If we begin with Model 1 where both the dependent variable and the regressor are nonstationary fractionally integrated processes with the orders of integration $1+d_1$ and $1+d_2$, respectively, where $d_1+d_2 > 0.5$, then first-differencing both variables cannot completely eliminate the spurious effects. While R^2 may be reduced and the DW statistic may be increased, the t ratios may still be so large that we cannot avoid making a spurious inference. This is a fairly serious problem with the regression for the fractionally integrated processes. It implies that even the popular first-differencing procedure might not prevent us from finding a spurious relationship among highly persistent data series. One lesson we learn from this discussion is that it is very important to check individual data series for possible long memory before regression can be applied.

2.2.3. Two Intermediate Cases: Model 3 and Model 4

Model 3 and Model 4 can be considered as two intermediate models between Model 1 and Model 2 in that one of the dependent variable and the regressor is stationary while the other is not. We expect the asymptotic results for these two new models to be hybrid of those of Model 1 and Model 2.

In Model 3 a nonstationary $I(1+d_1)$ process y_t is regressed on an independent and stationary $I(d_2)$ process w_t:

$$y_t = \alpha + \beta w_t + u_t, \qquad \text{where } d_2 > 0.$$

Note that the fractional differencing parameter d_2 for the regressor w_t here is assumed to be positive; i.e., w_t has long memory. The asymptotic properties of the OLS estimators for Model 3 are given in the following theorem:

Theorem 3. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results.*

1. $\frac{T}{\sigma_y \sigma_x} \widehat{\beta} \Rightarrow$

$$\frac{1}{\gamma_w(0)} \left\{ \int_0^1 B_{0.5+d_1}(s)\, dB_{0.5+d_2}(s) - \left[\int B_{0.5+d_1}(s) ds \right] \cdot B_{0.5+d_2}(1) \right\}$$

$$\equiv \beta_*.$$

Note that $\sigma_y \sigma_x / T = O(T^{d_1+d_2})$.

2. $\frac{1}{\sigma_y} \widehat{\alpha} \Rightarrow \int_0^1 B_{0.5+d_1}(s)\, ds \equiv \alpha_*.$

Note that $\sigma_y = O(T^{0.5+d_1})$.

3. $\frac{1}{\sigma_y^2} s^2 \Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_1}(s)\, ds \right]^2 \equiv \sigma_*^2.$

Note that $\sigma_y^2 = O(T^{1+2d_1})$.

4. $$\frac{T}{\sigma_y^2} s_\beta^2 \Rightarrow \frac{\sigma_*^2}{\gamma_w(0)},$$

where σ_*^2 *is defined in 3. Note that* $\sigma_y^2/T = O(T^{2d_1})$.

5. $$\frac{T}{\sigma_y^2} s_\alpha^2 \Rightarrow \sigma_*^2,$$

where σ_*^2 *is defined in 3. Note that* $\sigma_y^2/T = O(T^{2d_1})$.

6. $$\frac{\sqrt{T}}{\sigma_x} t_\beta \Rightarrow \frac{\sqrt{\gamma_w(0)} \cdot \beta_*}{\sigma_*},$$

where β_* *is defined in 1 and* σ_*^2 *is defined in 3. Note that* $\sigma_x/\sqrt{T} = O(T^{d_2})$.

7. $$\frac{1}{\sqrt{T}} t_\alpha \Rightarrow \frac{\alpha_*}{\sigma_*},$$

where α_* *is defined in 2 and* σ_*^2 *is defined in 3.*

8. $$\frac{T^2}{\sigma_x^2} R^2 \Rightarrow \frac{\gamma_w(0) \cdot \beta_*^2}{\sigma_*^2},$$

where β_* *is defined in 1 and* σ_*^2 *is defined in 3. Note that* $\sigma_x^2/T^2 =$

$O(T^{2d_2-1})$.

9. $DW \xrightarrow{p} 0$.

Here $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions.

Since both t ratios diverge, Model 3 also suffers from the spurious effect in terms of the t tests. Moreover, we find the results that the OLS estimator $\widehat{\alpha}$ diverge and that DW converges in probability to 0 are close to what we get in Model 1, while the result of converging R^2 is the same as that of Model 2. So Model 3 is indeed a mixture of Model 1 and 2.

It should be pointed out that in Theorem 3 the range of the fractional differencing parameter d_2 of the regressor w_t is restricted to the positive half of the original range $(-0.5, 0.5)$. For the case of a negative d_2, it is quite straightforward to show that the t ratios are convergent and there is no spurious effect.

In Model 4 a stationary $I(d_1)$ process is regressed on an independent and nonstationary $I(1+d_2)$ process

$$v_t = \alpha + \beta x_t + u_t, \qquad \text{where } d_1 > 0.$$

Similar to the restriction imposed on Model 3, the fractional differencing parameter d_1 of the dependent variable v_t is assumed to be positive so that v_t has long memory. The asymptotic theory for Model 4 is presented in the following theorem.

Theorem 4. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results.*

1. $\frac{T\sigma_x}{\sigma_y}\widehat{\beta} \Rightarrow$

$$\frac{\int_0^1 B_{0.5+d_2}(s)\, dB_{0.5+d_1}(s) - B_{0.5+d_1}(1)\cdot\left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2} \equiv \beta_*.$$

   Note that $\sigma_y/T\sigma_x = O(T^{d_1-d_2-1})$.

2. $\frac{T}{\sigma_y}\widehat{\alpha} \;\Rightarrow\; B_{0.5+d_1}(1) - \beta_*\int_0^1 B_{0.5+d_2}(s)\, ds \equiv \alpha_*,$

   where β_* *is defined in 1. Note that* $\sigma_y/T = O(T^{d_1-0.5})$.

3. $s^2 \xrightarrow{p} \gamma_v(0)$.

4. $T\sigma_x^2 \cdot s_\beta^2 \;\Rightarrow\; \frac{\gamma_v(0)}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2} \equiv \sigma_{*\beta}^2.$

   Note that $1/T\sigma_x^2 = O(T^{-2-2d_2})$.

5. $T \cdot s_\alpha^2 \;\Rightarrow$

$$\gamma_v(0)\left\{1+\frac{\left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2}{\int_0^1 \left[B_{0.5+d_2}(s)\right]^2\,ds-\left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2}\right\}\equiv\sigma_{*\alpha}^2.$$

6. $\dfrac{\sqrt{T}}{\sigma_y}t_\beta \;\Rightarrow\; \dfrac{\beta_*}{\sigma_{*\beta}}$,

   where β_* is defined in 1 and $\sigma_{*\beta}^2$ is defined in 4. Note that $\sigma_y/\sqrt{T}=O(T^{d_1})$.

7. $\dfrac{\sqrt{T}}{\sigma_y}t_\alpha \;\Rightarrow\; \dfrac{\alpha_*}{\sigma_{*\alpha}}$,

   where α_* is defined in 2 and $\sigma_{*\alpha}^2$ is defined in 5. Note that $\sigma_y/\sqrt{T}=O(T^{d_1})$.

8. $\dfrac{T^2}{\sigma_y^2}R^2 \;\Rightarrow\; \dfrac{\beta_*^2}{\sigma_{*\beta}^2}$,

   where β_* is defined in 1 and $\sigma_{*\beta}^2$ is defined in 4. Note that $\sigma_y^2/T^2=O(T^{2d_1-1})$.

9. $DW \;\xrightarrow{p}\; 2-2\rho_v(1)=\dfrac{2(1-2d_1)}{1-d_1}$.

Here $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions.

Since both t ratios diverge (at the same rate), the spurious effect in terms of failing t tests again exists in Model 4. But contrary to the results in Model 3, the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$, together with R^2, all converge in probability to zero, while the DW statistic converges to $2 - 2\rho_v(1)$. These findings obviously bring Model 4 closer to Model 2 than to Model 1.

2.2.4. The Relationship between the Orders of Integration and the Divergence Rates

The divergent t ratios in the above four models and the resulting failure of the t tests are referred to as the spurious effects. In this section we compare the divergence rates of t ratios across the four models and investigate how they are related to the respective model specifications.

First note that the divergence rates of the t ratio t_β are $T^{0.5}$, $T^{d_1+d_2-0.5}$, T^{d_2}, and T^{d_1}, respectively, for Models 1 to 4. Let's also compare the specifications of the four models using Model 1 as the benchmark: Model 3 differs from Model 1 in that the order of integration in the regressor is reduced from above 0.5 to below 0.5 (but above 0); Model 4 differs from Model 1 in that the order of integration in the dependent variable is reduced from above 0.5 to below 0.5 (but above 0); and, finally, Model 2 differs from Model 1 in that the orders of integration in both the dependent variable and the regressor are reduced from above 0.5 to below 0.5 (but their sum is assumed to be greater than 0.5). By associating these changes in the orders of integration with the changes in the divergence rates of t_β, we can conclude that reducing the order of integration in the dependent variable causes the divergence rate of t_β to decrease by the order of $T^{d_1-0.5}$ and reducing the order of integration in the regressor causes the divergence rate of t_β to decline by the order of $T^{d_2-0.5}$, while these two effects are cumulative as in Model 2.

Recall that in Models 2, 3 and 4 restrictions have been imposed on the usual ranges $(-0.5, 0.5)$ of the fractional differencing parameters d_1 and d_2. In Model 3 the range of d_2 is restricted to be $(0, 0.5)$, which is also the range of d_1 in Model 4, while the sum of d_1 and d_2 must be greater than 0.5 in Model 2. From the analysis in the previous paragraph, particularly the fact that the divergence rates are directly related to the magnitudes of d_1 and d_2, we come to realize that the restricted ranges of d_1 and d_2 in Models 2, 3, and 4 ensure the reduction in the divergence rates of t_β from the $T^{0.5}$ level is not too great so that t_β remains divergent (in which case the spurious effects occur). Although we did not explicitly consider the asymptotic theory for cases where the fractional differencing parameters lie outside their prescribed ranges, it is readily seen that the conditions we impose on the ranges are not only sufficient but also necessary for the existence of the spurious effect in terms of divergent t_β.

From a similar analysis for the divergence rates of the t ratio t_α we also find that reducing the order of integration in the dependent variable causes the divergence rate of t_α to decrease by the order of $T^{d_1-0.5}$, while reducing the order of integration in the regressor *does not* cause the divergence rate of t_α to change, as we probably should have expected.

It is also interesting to see how the changes in the orders of integration of the dependent variable and the regressor affect the large-sample property of R^2. Recall that in Model 1 R^2 converges to a random variable and such asymptotic behavior of R^2 is considered part of the spurious effect by Phillips (1986). But when we examine Models 2, 3, and 4, we note that reducing the order of integration in the dependent variable helps to increase its convergence rate by the order of T^{1-2d_1} while reducing the order of integration in the regressor helps to increase the convergence rate by the order of T^{1-2d_2}. As a result, in Models 2, 3, and 4, R^2 all converge to 0, correctly reflecting the fact that there is no relationship between the regressor and the dependent variable. This finding implies that the spurious effects in Models 2, 3, and 4 are confined to the two t ratios while the asymptotic tendency of R^2 toward zero is not affected by the spurious effects (though the convergence rates are).

The sharp difference in the asymptotic behavior between the t ratios and R^2 in Models 2, 3, and 4 actually offers us an opportunity to diagnose the spurious effect in these models. That is, when we find two highly significant t ratios coexisting with a completely contradictory near-zero R^2, we are effectively reminded of the possibilities that one of the Models 2, 3, and 4 may be at work and that the dependent variable and the regressor may possess strong long memory, while one of them may even be nonstationary. With the possibility of such an informal diagnosis, it seems that the spurious effects in Models 2, 3, and 4 are less damaging than those in Model 1 in the sense that in Model 1 there is no internal inconsistency among the OLS estimates to indicate the spurious effects.

Finally, let's briefly state a few more results about the asymptotic tendency of the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$ and the DW statistic. First, we note that $\widehat{\beta}$ will converge in probability to zero unless the dependent variable is nonstationary and its order of integration is sufficiently large. Secondly, whether $\widehat{\alpha}$ diverges or not and whether the DW statistic converges in probability to zero or not depends entirely on whether the dependent variable is nonstationary or not. Note that, as mentioned earlier, even though the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$ can converge in probability to zero in the four proposed models, the corresponding t ratios always diverge and it is these divergent t ratios that are referred to as the spurious effects.

2.2.5. Model 5 and Model 6: Detrending Fractionally Integrated Processes

As has been pointed out by Nelson and Kang (1981, 1984) and Durlauf and Phillips (1988), detrending integrated processes results in the spurious effect of finding a significant trend. In this section we extend their analysis by considering the potential problems in detrending fractionally integrated processes. It turns out that the spurious effect of divergent t ratios exists as long as the fractional differencing parameter is larger than zero. The implication is that whenever there is long memory in the process, the routine procedure of detrending can produce misleading results. It appears that the spurious effect in detrending occurs more often than we previously thought.

In our analysis of detrending fractionally integrated processes, we separate the nonstationary case from the stationary case. In Model

5 we examine the regression of a nonstationary $I(1+d_1)$ process y_t on a time trend t:

$$y_t = \alpha + \beta t + u_t.$$

The asymptotic theory for the OLS estimation is given in the following theory.

Theorem 5. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results.*

1. $$\frac{T}{\sigma_y}\widehat{\beta} \;\Rightarrow\; 12\int_0^1 s\cdot B_{0.5+d_1}(s)\,ds - 6\int_0^1 B_{0.5+d_1}(s)\,ds \equiv \beta_*.$$

   Note that $\sigma_y/T = O(T^{d_1-0.5})$.

2. $$\frac{1}{\sigma_y}\widehat{\alpha} \;\Rightarrow\; 4\int_0^1 B_{0.5+d_1}(s)\,ds - 6\int_0^1 s\cdot B_{0.5+d_1}(s)\,ds \equiv \alpha_*.$$

3. $$\frac{1}{\sigma_y^2}s^2 \;\Rightarrow\; \int_0^1 [B_{0.5+d_1}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_1}(s)\,ds\right]^2$$
   $$- 12\left[\int_0^1 s\,B_{0.5+d_1}(s)\,ds - \frac{1}{2}\int_0^1 B_{0.5+d_1}(s)\,ds\right]^2 \equiv \sigma_*^2.$$

4. $$\frac{T^3}{\sigma_y^2}s_\beta^2 \;\Rightarrow\; 12\sigma_*^2,$$

   where σ_*^2 is defined in 3. Note that $\sigma_y^2/T^3 = O(T^{2d_1-2})$.

5. $$\frac{T}{\sigma_y^2}s_\alpha^2 \;\Rightarrow\; 4\sigma_*^2,$$

where σ_*^2 is defined in 3. Note that $\sigma_y^2/T = O(T^{2d_1})$.

6. $$\frac{1}{\sqrt{T}} t_\beta \;\Rightarrow\; \frac{\beta_*}{\sqrt{12}\,\sigma_*},$$

   where β_* is defined in 1 and σ_*^2 is defined in 3.

7. $$\frac{1}{\sqrt{T}} t_\alpha \;\Rightarrow\; \frac{\alpha_*}{2\sigma_*},$$

   where α_* is defined in 2 and σ_*^2 is defined in 3.

8. $$R^2 \;\Rightarrow\; \frac{\beta_*^2}{12\int_0^1 [B_{0.5+d_1}(s)]^2\, ds - 12\left[\int_0^1 B_{0.5+d_1}(s)\, ds\right]^2},$$

   where β_* is defined in 1.

9. $$DW \xrightarrow{p} 0 \quad \text{and} \quad \sigma_y^2 \cdot DW \;\Rightarrow\; \frac{\gamma_v(0)}{\sigma_*^2}.$$

Here, $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions.

The results on detrending a stationary long memory $I(d_1)$ process v_t

$$v_t = \alpha + \beta t + u_t, \qquad \text{where } d_1 > 0,$$

which is our Model 6, are presented in the following theorem.

Theorem 6. *Given that Assumption 1 holds, then, as $T \to \infty$, we have the following results.*

1. $$\frac{T^2}{\sigma_y}\widehat{\beta} \;\Rightarrow\; 6\,B_{0.5+d_1}(1) - 12\int_0^1 B_{0.5+d_1}(s)\,ds \equiv \beta_*.$$

   Note that $\sigma_y/T^2 = O(T^{d_1-1.5})$.

2. $$\frac{T}{\sigma_y}\widehat{\alpha} \;\Rightarrow\; 6\int_0^1 B_{0.5+d_1}(s)\,ds - 2\,B_{0.5+d_1}(1) \equiv \alpha_*.$$

   Note that $\sigma_y/T = O(T^{d_1-0.5})$.

3. $s^2 \xrightarrow{p} \gamma_v(0)$.

4. $T^3 \cdot s_\beta^2 \xrightarrow{p} 12\gamma_v(0)$.

5. $T \cdot s_\alpha^2 \xrightarrow{p} 4\gamma_v(0)$.

6. $$\frac{\sqrt{T}}{\sigma_y}t_\beta \;\Rightarrow\; \frac{\beta_*}{\sqrt{12\,\gamma_v(0)}},$$

   where β_* is defined in 1. Note that $\sigma_y/\sqrt{T} = O(T^{d_1})$.

7. $$\frac{\sqrt{T}}{\sigma_y}t_\alpha \;\Rightarrow\; \frac{\alpha_*}{2\sqrt{\gamma_v(0)}},$$

where α_* is defined in 2. Note that $\sigma_y/\sqrt{T} = O(T^{d_1})$.

8. $$\frac{T^2}{\sigma_y^2} R^2 \;\Rightarrow\; \frac{\beta_*^2}{12\gamma_v(0)},$$

where β_* is defined in 1. Note that $\sigma_y^2/T^2 = O(T^{2d_1-1})$.

9. $$DW \;\xrightarrow{p}\; 2 - 2\rho_v(1) = \frac{2(1-2d_1)}{1-d_1}.$$

Here, $B_{0.5+d_1}(t)$ and $B_{0.5+d_2}(t)$ are two independent fractional Brownian motions.

In terms of the convergence (or divergence) rates of the various OLS estimators, Models 5 and 6 can be conveniently viewed as "special cases" of Models 1 and 4, respectively. More specifically, if we replace the term σ_x by T in those normalizing factors in Theorems 1 and 4, then we immediately get all the normalizing factors in Theorems 5 and 6. For example, while the normalizing factor for $\widehat{\beta}$ in Theorem 1 is σ_x/σ_y, the one in Theorem 5 is T/σ_y. Similarly, while the normalizing factor for $\widehat{\beta}$ in Theorem 4 is $T\sigma_x/\sigma_y$, the one in Theorem 6 is T^2/σ_y. Given this observation, we then conclude that all the analyses about Models 1 and 4 can be readily extended to Models 5 and 6. In particular, the divergence rates of the t ratios, which respectively are in the orders of $T^{0.5}$ and T^{d_1} in Models 1 and 4, are also the rates in Models 5 and 6. (Note that in both Model 4 and Model 6 the same condition $d_1 > 0$ is imposed on the stationary dependent variable v_t so that the resulting t ratios are divergent.) As a result, the type of spurious effects we observe in Models 1 and 4 occur again in Models 5 and 6. That is, detrending a fractionally integrated process with a positive fractional differencing parameter, certainly including the usual case of the $I(1)$ process, will result in the spurious finding of a significant trend. One important inference we draw from Models 5 and 6 is that the cause for the spurious effect in detrending a process is neither nonstationarity nor lack of ergodicity but long memory in the process.

From Models 5 and 6 we also note the following result: If the data series are nonstationary with the order of integration greater than 1, then the spurious effect can happen to the detrending procedure even after the series are first-differenced. What first-differencing does to the detrending procedure in such a case is simply reducing R^2, increasing the value of the DW statistic, and slowing down the divergence of the two t ratios from the $T^{0.5}$ rate to the T^{d_1} rate. Based on this observation, it seems that the spurious effects in detrending may occur more often than we previously thought.

2.3. Monte Carlo Experiments

After the theoretical analysis of the spurious effect, we now conduct an extensive Monte Carlo experiment to investigate the relevance of the theory in small sample applications. The design of the Monte Carlo study is standard. The Monte Carlo experiment for each model is based on 10,000 replications with different sample sizes (T). To construct T values of the stationary $I(d)$ process, we first generate T independent values from the standard normal distribution and form a $T \times 1$ column vector e. We then calculate the T analytic autocovariances of the $I(d)$ process, from which we construct the $T \times T$ variance-covariance matrix Σ and compute its Cholesky decomposition C (i.e., $\Sigma = CC'$). Finally, the vector p of the T realized values of the $I(d)$ process is defined by $p = Ce$. This algorithm was suggested by McLeod and Hipel (1978) and Hosking (1984).

To verify item 6 of Theorem 1 that the t ratio t_β for the slope coefficient β diverges at the $\sqrt{T}$ rate, we test the null hypothesis

$$H_o\colon \beta = 0$$

at different level of significance ($\aleph$) using the traditional two-tailed t test. Tables 2–1 contains the information about the rejection percentages and the averages of the absolute value of t_β under Model 1, where both dependent variable and regressor are nonstationary.

If we treat the result of hypothesis testing in every replication as a binomial trial and define the null hypothesis being rejected as a success, then each hypothesis testing is a binomial trial with probability of success $\aleph$. Therefore, the 95% confidence interval for $\aleph = 0.05$ is

$$0.05 \pm 1.96\sqrt{\frac{0.05 \times 0.95}{10000}} \simeq (4.57\%, 5.43\%).$$

As shown in Table 2–1, the rejection percentages at every value of $\aleph$ are outside their corresponding 95% confidence intervals. Moreover, we find the average absolute value of t_β increases with the sample size increases. These results support Theorem 1 that t_β diverges with T increases.

To further verify the theory that the t ratio t_β for the slope coefficient β diverges at the $\sqrt{T}$ rate, we calculate the ratios of the averages of $|t_\beta|$ for two sample sizes T_1 and T_2. These ratios can be compared with $\sqrt{T_1/T_2}$. As shown in Table 2–2, these ratios are all very close to the corresponding $\sqrt{T_1/T_2}$.

Table 2–3 contains the simulation results for Model 2, where both dependent variable and regressor are stationary. Once again, we find the rejection percentages at every value of $\aleph$ increases as T increases and the average of the absolute value of t_β diverges with T increases. We also calculate the ratios the averages of $|t_\beta|$ for two sample sizes T_1 and T_2 and compare them with $(T_1/T_2)^{d_1+d_2-0.5}$ in Table 2–4. We find these ratios are very close to 1.

The simulation results in the lowest block of Table 2–3 and Table 2–4 are based on a chi-square distribution with degree of freedom 1 χ_1^2 instead of the standard normal distribution. More specifically, the innovations of the fractionally integrated processes are generated as independent $\chi_1^2 - 1$ random variables. The conclusions we draw from these simulations are the same, which implies that the spurious effects will occur to Model 2 irrespective of the distributional assumption.

Table 2–5 and Table 2–6 contain simulation results for Models 3 and 4. Again, the rejection percentages at every $\aleph$ value and the averages of $|t_\beta|$ diverge as sample sizes increase, which convincingly support Theorems 3 and 4.

2.4. Conclusion

In our analysis of spurious regressions for the long memory fractionally integrated processes, we find that no matter whether the dependent variable and the regressor are stationary or not, as long as their orders of integration sum up to a value greater than 0.5, the t ratios become divergent. So it is the long memory, instead of nonstationarity or lack of ergodicity, that causes the spurious effects in terms of failing t tests. Nonstationarity in one or both of the dependent variable and the regressor only helps to accelerate the divergence rates of the t ratios. We thus learn that spurious effects might occur more often than we previously believed as they can arise even among stationary series and the usual first-differencing procedure may not be able to completely eliminate spurious effects when data possess strong long memory.

In subsection 2.4 we have carefully examined the exact relationships between the orders of integration in the fractionally integrated processes and the divergence rates in the t ratios. From this analysis we gain many insights into the problem of spurious effects which are not available in Phillips' (1986) classical study of $I(1)$ processes. In short, it is found that the extents of spurious effects are directly related to the degrees of long memory in the data. Our results on detrending fractionally integrated processes also greatly broaden Durlauf and Phillips' (1988) theory of spurious detrending in which the relationship between the orders of integration and the divergence rates of the t ratios again plays a useful role in the analysis.

A fairly extensive Monte Carlo study has also been conducted to verify the theoretical results, especially those of convergence rates, we have established in this chapter. Our theoretical results are well supported by simulation.

A few generalizations of our study are worthy of further consideration. A natural extension is to consider the multiple regression where there are more than one non-constant regressor. Another one is to allow the fractionally integrated processes to have non-zero means. Based on Phillips' (1986) work, we expect most, if not all, of the asymptotic results we obtain from the simple regression case to hold in the multiple regression of fractionally integrated processes with drifts.

One aspect of our study that is slightly more restricted than Phillips' (1986) and Durlauf and Phillips' (1988) analysis is that the fractionally integrated processes we consider are built on white noises a_t and b_t that are required to satisfy the relatively stringent conditions as specified in Assumption 1. These conditions effectively rule out the possibility of allowing short-run dynamics such as the ARMA components in the fractionally integrated processes we have studied. Although relaxing Assumption 1 to incorporate the short-run dynamics does not seem to pose too great technical difficulty and we do not expect substantial changes in the analysis of spurious effects, some modification in the theorem proofs is nevertheless necessary and is beyond the scope of this paper.

Finally, our study of spurious regression can serve as the basis for the analyses of "fractional cointegration" where the dependent variable and regressors are related $I(d)$ processes. This line of the work appears to be quite important and has attracted a lot attention in the literature recently. One of the pioneer works in this area is by Cheung and Lai (1993). The research on this topic has also been conducted in Chapter 3 of this dissertation.

2.5. Mathematical Proof

A.1. Proof of Lemma 1

The proofs of items 1, 2, and 3 are straightforward applications of the continuous mapping theorem to the Davydov's results. They are omitted here. Item 4 follows directly from Davydov's result, while items 5 and 6 are due to ergodicity of the two stationary processes v_t and w_t.

To prove item 7, we note, since v_t and w_t are assumed to be independent and have zero means, the autocovariance of the product $v_t w_t$ at lag j is the product of their respective autocovariance at lag j: $\gamma_v(j)\gamma_w(j)$. Also, it is well-known that $\gamma_v(j) = O(j^{2d_1-1})$ and $\gamma_w(j) = O(j^{2d_2-1})$ if $d_1 \neq 0$ and $d_2 \neq 0$. Consequently, we have

$$\mathrm{Var}\left(\sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{v_t w_t}\right) = T \cdot \sum_{j=-(T-1)}^{T} \left(1 - \frac{|j|}{T}\right) \gamma_v(j)\gamma_w(j)$$

$$= T \cdot \sum_{j=-(T-1)}^{T} O(|j|^{2d_1+2d_2-2})$$

$$\begin{cases} = O(T^{2d_1+2d_2}), & \text{if } \ d_1 + d_2 \geq 0.5, \\ = O(T)\cdot\zeta(2 - 2d_1 - 2d_2), & \text{if } \ d_1 + d_2 < 0.5, \text{ and } \ d_1, d_2 > 0, \\ \leq O(T), & \text{otherwise,} \end{cases}$$

where $\zeta(\cdot)$ is Riemann's zeta function. Given this result and the fact that $\mathrm{E(v_t w_t)} = 0$, then Chebyshev's inequality implies that, for any $\epsilon > 0$,

$$P\left(\left|\frac{1}{T^{\epsilon+d_1+d_2}} \sum_{t=1}^{T} v_t w_t\right| > \delta\right) \ < \ \frac{1}{\delta^2} \mathrm{Var}\left(\frac{1}{\mathrm{T^{\epsilon+d_1+d_2}}} \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{v_t w_t}\right) = \mathrm{o}(1),$$

when $d_1 + d_2 \geq 0.5$, and

$$P\left(\left|\frac{1}{T^{\epsilon+0.5}}\sum_{t=1}^{T} v_t w_t\right| > \delta\right) < \frac{1}{\delta^2}\operatorname{Var}\left(\frac{1}{T^{\epsilon+0.5}}\sum_{t=1}^{T} v_t w_t\right) = o(1),$$

when $d_1 + d_2 < 0.5$. Consequently,

$$\sum_{t=1}^{T} v_t w_t = \begin{cases} o_p(T^{\epsilon+d_1+d_2}), & \text{if } d_1 + d_2 \geq 0.5, \\ o_p(T^{\epsilon+0.5}), & \text{otherwise.} \end{cases}$$

To prove item 8, we note

$$\frac{1}{T}\sum_{t=1}^{T}\frac{y_t}{\sigma_y}\cdot\frac{x_t}{\sigma_x} = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{t-1}}{\sigma_y}\cdot\frac{x_{t-1}}{\sigma_x} + \frac{1}{T\sigma_y\sigma_x}\sum_{t=1}^{T}(v_t x_{t-1} + w_t y_{t-1} + v_t w_t)$$

$$= \sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\frac{y_{[Ts]}}{\sigma_y}\cdot\frac{x_{[Ts]}}{\sigma_x}\,ds + o_p(1)$$

$$= \int_0^1 \frac{y_{[Ts]}}{\sigma_y}\cdot\frac{x_{[Ts]}}{\sigma_x}\,ds + o_p(1)$$

$$\Rightarrow \int_0^1 B_{0.5+d_1}(s)\cdot B_{0.5+d_2}(s)\,ds.$$

To show the second term at the end of the first line is $o_p(1)$, we note that

$$\sum_{t=1}^{T} v_t x_{t-1} \leq \left(\sum_{t=1}^{T} v_t^2\right)^{1/2}\left(\sum_{t=1}^{T} x_t^2\right)^{1/2}$$

$$= O_p(T^{0.5})\cdot O_p(T^{1+d_2}) = O_p(T^{1.5+d_2}),$$

$$\sum_{t=1}^{T} y_{t-1} w_t \leq \left(\sum_{t=1}^{T} y_t^2\right)^{1/2}\left(\sum_{t=1}^{T} w_t^2\right)^{1/2}$$

$$= O_p(T^{1+d_1})\cdot O_p(T^{0.5}) = O_p(T^{1.5+d_1}).$$

The orders of the four sums of squares are based on the results of items 2 and 5. We also note that $T\sigma_y\sigma_x = O(T^{2+d_1+d_2})$. These results, together with item 7, imply

$$\frac{1}{T\sigma_y\sigma_x}\sum_{t=1}^{T} v_t x_{t-1} \le O_p(T^{-0.5-d_1}), \qquad \frac{1}{T\sigma_y\sigma_x}\sum_{t=1}^{T} y_{t-1} w_t \le O_p(T^{-0.5-d_2}),$$

and

$$\frac{1}{T\sigma_y\sigma_x}\sum_{t=1}^{T} v_t w_t = \max\{o_p(T^{\epsilon-2}),\, o_p(T^{\epsilon-1.5-d_1-d_2})\},$$

for any $\epsilon > 0$. So the above three terms all approach to 0 and the proof of item 8 is completed. To prove item 9, we first note that we are considering the case $d_2 > 0$ only. Now, for a sufficiently large T, we have

$$\frac{w_t}{\sigma_x} = \frac{x_t - x_{t-1}}{\sigma_x} \approx \frac{dx_{t-1}}{\sigma_x} = \frac{dx_{[Ts]}}{\sigma_x}, \qquad \text{for } \frac{t-1}{T} \le s < \frac{t}{T}.$$

This observation, together with Davydov's result, the continuous mapping theorem, and item 7, imply

$$\begin{aligned}
\sum_{t=1}^{T} \frac{y_t}{\sigma_y}\cdot\frac{w_t}{\sigma_x} &= \sum_{t=1}^{T} \frac{y_{t-1}}{\sigma_y}\cdot\frac{w_t}{\sigma_x} + \sum_{t=1}^{T} \frac{v_t}{\sigma_y}\cdot\frac{w_t}{\sigma_x} \\
&= \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \frac{y_{[Ts]}}{\sigma_x}\cdot\frac{dx_{[Ts]}}{\sigma_y} + o_p(1) \\
&= \int_0^1 \frac{y_{[Ts]}}{\sigma_x}\cdot\frac{dx_{[Ts]}}{\sigma_y} + o_p(1) \;\Rightarrow\; \int_0^1 B_{0.5+d_1}(s)\, dB_{0.5+d_2}(s).
\end{aligned}$$

Note that the condition $d_2 > 0$ is needed to ensure that

$$\sum_{t=1}^{T} \left(\frac{v_t}{\sigma_y}\right)\left(\frac{w_t}{\sigma_x}\right) = o_p(1).$$

The result for $\sum_{t=1}^{T}(v_t/\sigma_y)(x_t/\sigma_x)$ given $d_1 > 0$ can be proved in a similar fashion. To prove item 10, we first note that

$$\sum_{t=1}^{T} y_{t-1} = \sum_{t=1}^{T}(T-t)v_t = T\sum_{t=1}^{T} v_t - \sum_{t=1}^{T} t\cdot v_t.$$

Hence,

$$\sum_{t=1}^{T}\frac{t}{T}\cdot\frac{v_t}{\sigma_y} = \sum_{t=1}^{T}\frac{v_t}{\sigma_y} - \frac{1}{T}\sum_{t=1}^{T}\frac{y_{t-1}}{\sigma_y} \quad\Rightarrow\quad B_{0.5+d_1}(1) - \int_0^1 B_{0.5+d_1}(s)\,ds,$$

by applying the results of items 4 and 1. The weak limit of

$$\sum_{t=1}^{T}(\frac{t}{T})(\frac{w_t}{\sigma_x})$$

can be derived using a similar argument. To prove item 11, we note

$$\frac{1}{T}\sum_{t=1}^{T}\frac{t}{T}\cdot\frac{y_t}{\sigma_y} = \frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\cdot\frac{y_{t-1}}{\sigma_y} + \frac{1}{T^2}\sum_{t=1}^{T}\frac{y_{t-1}}{\sigma_y} + \frac{1}{T}\sum_{t=1}^{T}\frac{t}{T}\cdot\frac{v_t}{\sigma_y}$$

$$= \sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\frac{[Ts]}{T}\cdot y_{[Ts]}\,ds + o_p(1)$$

$$= \int_0^1 \frac{[Ts]}{T}\cdot y_{[Ts]}\,ds + o_p(1)$$

$$\Rightarrow \int_0^1 s\cdot B_{0.5+d_1}(s)\,ds.$$

The orders of the last two terms at the end of the first line are based on the results of items 1 and 10. The result for $(1/T)\sum_{t=1}^{T}(t/T)(x_t/\sigma_x)$ can be proved by a similar argument.

A.2. Proof of Theorem 1

Let's first summarize the formulas for all relevant statistics in the simple linear regression model $y_t = \alpha + \beta x_t + u_t$:

$$\widehat{\beta} = \frac{\sum_{t=1}^{T} y_t x_t - \frac{1}{T}\sum_{t=1}^{T} y_t \sum_{t=1}^{T} x_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}, \qquad \widehat{\alpha} = \frac{1}{T}\sum_{t=1}^{T} y_t - \widehat{\beta}\cdot\frac{1}{T}\sum_{t=1}^{T} x_t,$$

$$\widehat{u}_t = y_t - \widehat{\alpha} - \widehat{\beta} x_t$$

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2 - \widehat{\beta}^2\cdot\frac{1}{T}\sum_{t=1}^{T}(x_t - \overline{x})^2,$$

$$s_{\beta}^2 = \frac{s^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2}, \qquad s_{\alpha}^2 = s^2\left[\frac{1}{T} + \frac{(\overline{x})^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2}\right],$$

$$R^2 = \frac{\widehat{\beta}^2 \sum_{t=1}^{T}(x_t - \overline{x})^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2}, \qquad DW = \frac{\sum_{t=2}^{T}(\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T}\widehat{u}_t^2}.$$

To prove item 1, we note

$$\frac{\sigma_x}{\sigma_y}\widehat{\beta} = \frac{\frac{1}{T\sigma_y\sigma_x}\sum_{t=1}^{T} y_t x_t - \left(\frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t\right)\left(\frac{1}{T\sigma_x}\sum_{t=1}^{T} x_t\right)}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \overline{x})^2} \Rightarrow \beta_*,$$

where the weak convergence is due to items 1, 3, and 8 of Lemma 1.

To prove item 2, we note

$$\frac{1}{\sigma_y}\widehat{\alpha} = \frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t - \frac{\sigma_x}{\sigma_y}\widehat{\beta}\cdot\frac{1}{T\sigma_x}\sum_{t=1}^{T} x_t \quad \Rightarrow \quad \alpha_*,$$

where the weak convergence is based on item 1 above and item 1 of Lemma 1. To prove item 3, we have

$$\frac{1}{\sigma_y^2}s^2 = \frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \overline{y})^2 - \left(\frac{\sigma_x}{\sigma_y}\widehat{\beta}\right)^2\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \quad \Rightarrow \quad \sigma_*^2,$$

where the weak convergence is based on item 1 above and item 3 of Lemma 1. To prove item 4, we note

$$\frac{T\sigma_x^2}{\sigma_y^2}s_\beta^2 = \frac{\dfrac{1}{\sigma_y^2}s^2}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \overline{x})^2} \quad \Rightarrow \quad \sigma_{*\beta}^2,$$

where the weak convergence is based on item 3 above and item 3 of Lemma 1. To prove item 5, we see

$$\frac{T}{\sigma_y^2}s_\alpha^2 = \frac{1}{\sigma_y^2}s^2\left[1 + \frac{\left(\dfrac{1}{T\sigma_x}\displaystyle\sum_{t=1}^{T}x_t\right)^2}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \overline{x})^2}\right] \quad \Rightarrow \quad \sigma_{*\alpha}^2,$$

where the weak convergence is based on item 3 above and items 1 and 3 of Lemma 1. Items 6 and 7 are direct results from items 1, 2, 4, and

5. To prove item 8, we note

$$R^2 = \frac{\left(\frac{\sigma_x}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \overline{x})^2}{\frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

$$\Rightarrow \frac{\beta_*^2\left\{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2\right\}}{\int_0^1 [B_{0.5+d_1}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_1}(s)\, ds\right]^2},$$

where the weak convergence is based on item 1 above and item 3 of Lemma 1. To prove item 9, we note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = [(y_t - \widehat{\alpha} - \widehat{\beta}x_t) - (y_{t-1} - \widehat{\alpha} - \widehat{\beta}x_{t-1})]^2$$

$$= (v_t - \widehat{\beta}w_t)^2 = v_t^2 - 2\widehat{\beta}v_t w_t + \widehat{\beta}^2 w_t^2,$$

and, based on items 5 and 7 of Lemma 1, we have

$$\frac{1}{T\sigma_y^2}\sum_{t=2}^{T} v_t^2 = o_p(1), \qquad \frac{1}{T\sigma_x^2}\sum_{t=2}^{T} w_t^2 = o_p(1),$$

and

$$\frac{1}{T\sigma_y\sigma_x}\sum_{t=2}^{T} v_t w_t = \max\{o_p(T^{\epsilon-2}),\, o_p(T^{\epsilon-1.5-d_1-d_2})\},$$

for any $\epsilon > 0$. Consequently,

$$DW = \frac{\frac{1}{T\sigma_y^2}\sum_{t=2}^{T} v_t^2 - 2\cdot\frac{\sigma_x}{\sigma_y}\widehat{\beta}\cdot\frac{1}{T\sigma_y\sigma_x}\sum_{t=2}^{T} v_t w_t + \left(\frac{\sigma_x}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T\sigma_x^2}\sum_{t=2}^{T} w_t^2}{\frac{1}{T\sigma_y^2}\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$= o_p(1).$$

Here we note that the denominator of DW is s^2/σ_y^2 which converges weakly to σ_*^2 by the result of item 3.

A.3. Proof of Theorem 2

To prove item 1, we note

$$\frac{T^2}{\sigma_y\sigma_x}\widehat{\beta} = \frac{\frac{T}{\sigma_y\sigma_x}\sum_{t=1}^{T} v_t w_t - \left(\frac{1}{\sigma_y}\sum_{t=1}^{T} v_t\right)\left(\frac{1}{\sigma_x}\sum_{t=1}^{T} w_t\right)}{\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2}$$

$$\Rightarrow \quad \frac{\sqrt{\gamma_v(0)\gamma_w(0)}\cdot Z(1) - B_{0.5+d_1}(1)\cdot B_{0.5+d_2}(1)}{\gamma_w(0)},$$

where the weak convergence is based on Corollary 1, and items 4 and 6 of Lemma 1. To prove item 2, we note that from item 1 above and item 4 of Lemma 1, we have

$$\frac{T}{\sigma_y}\widehat{\alpha} = \frac{1}{\sigma_y}\sum_{t=1}^{T} v_t - \frac{\sigma_x^2}{T^2}\cdot\frac{T^2}{\sigma_y\sigma_x}\widehat{\beta}\cdot\frac{1}{\sigma_x}\sum_{t=1}^{T} w_t \quad \Rightarrow \quad B_{0.5+d_1}(1),$$

where the second term converges in probability to zero since $\sigma_x^2/T^2 = O(T^{2d_2-1})$ which converges to zero for $0.25 < d_2 < 0.5$. To prove item 3, we note that from item 1 above and item 6 of Lemma 1, we have

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2 - \left(\frac{\sigma_y\sigma_x}{T^2}\right)^2\left(\frac{T^2}{\sigma_y\sigma_x}\widehat{\beta}\right)^2\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2 \quad \xrightarrow{p} \quad \gamma_v(0),$$

where the second term converges in probability to zero since $\sigma_y\sigma_x/T^2 = O(T^{d_1+d_2-1})$ which converges to zero for $0.25 < d_1,\ d_2 < 0.5$. To prove item 4, we have

$$T\cdot s_{\beta}^2 = \frac{s^2}{\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2} \quad \xrightarrow{p} \quad \frac{\gamma_v(0)}{\gamma_w(0)},$$

where the weak convergence is based on item 3 above and item 6 of Lemma 1. To prove item 5, we have

$$T \cdot s_\alpha^2 = s^2 \left[1 + \frac{\left(\frac{1}{T} \sum_{t=1}^{T} w_t \right)^2}{\frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2} \right] \xrightarrow{p} \gamma_v(0),$$

where the second term in the bracket converges in probability to zero since its numerator does. Items 6 and 7 are straightforward consequence of items 1, 2, 4, and 5. To prove item 8, we have

$$\frac{T^4}{\sigma_y^2 \sigma_x^2} R^2 = \frac{\left(\frac{T^2}{\sigma_y \sigma_x} \widehat{\beta} \right)^2 \frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2}{\frac{1}{T} \sum_{t=1}^{T} (v_t - \overline{v})^2}$$

$$\Rightarrow \left[\frac{\sqrt{\gamma_v(0)\gamma_w(0)} \cdot Z(1) - B_{0.5+d_1}(1) \cdot B_{0.5+d_2}(1)}{\gamma_w(0)} \right]^2 \frac{\gamma_w(0)}{\gamma_v(0)},$$

where the weak convergence is based on item 1 above and item 6 of Lemma 1. To prove item 9, we first note

$$\begin{aligned} (\widehat{u}_t - \widehat{u}_{t-1})^2 &= [(v_t - \widehat{\alpha} - \widehat{\beta} w_t) - (v_{t-1} - \widehat{\alpha} - \widehat{\beta} w_{t-1})]^2 \\ &= [(v_t - v_{t-1}) - \widehat{\beta}(w_t - w_{t-1})]^2 \\ &= (v_t - v_{t-1})^2 - 2\widehat{\beta}(v_t - v_{t-1})(w_t - w_{t-1}) \\ &\quad + \widehat{\beta}^2 (w_t - w_{t-1})^2. \end{aligned}$$

Also, from item 7 of Lemma 1 and the assumption that $d_1 + d_2 > 0.5$, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})(w_t - w_{t-1}) = \frac{1}{T}\sum_{t=2}^{T} v_t w_t + \frac{1}{T}\sum_{t=2}^{T} v_{t-1} w_{t-1}$$

$$- \frac{1}{T}\sum_{t=2}^{T} v_{t-1} w_t - \frac{1}{T}\sum_{t=2}^{T} v_t w_{t-1}$$

$$= o_p(T^{\epsilon - 1 + d_1 + d_2}),$$

for any $\epsilon > 0$. Moreover, we have $\widehat{\beta} = o_p(1)$ from item 1 above and

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} v_t v_{t-1}$$

$$\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\,\gamma_v(1).$$

All these results then imply DW statistic is

$$\frac{\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 - 2\widehat{\beta}\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})(w_t - w_{t-1}) + \widehat{\beta}^2\frac{1}{T}\sum_{t=2}^{T} w_t^2}{\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$\xrightarrow{p} 2 - 2\,\rho_v(1),$$

where the second and the third terms in the numerator converge in probability to zero while the denominator converges in probability to $\gamma_v(0)$.

A.4. Proof of Corollary 2

It suffices to show that $T^2\widehat{\beta}/\sigma_y\sigma_x = O_p(1)$. But from the proof for item 7 of Lemma 1 we know that if $d_1 + d_2 > 0.5$, then

$$\frac{T}{\sigma_y\sigma_x}\sum_{t=1}^{T} v_t w_t = O_p(1).$$

If we denote its weak limit by Z^* (of which the exact specification is unknown), then, by the same argument as in the proof of item 1 of Theorem 2, we have

$$\frac{T^2}{\sigma_y\sigma_x}\widehat{\beta} = \frac{\frac{T}{\sigma_y\sigma_x}\sum_{t=1}^{T} v_t w_t - \left(\frac{1}{\sigma_y}\sum_{t=1}^{T} v_t\right)\left(\frac{1}{\sigma_x}\sum_{t=1}^{T} w_t\right)}{\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2}$$

$$\Rightarrow \frac{Z^* - B_{0.5+d_1}(1)\cdot B_{0.5+d_2}(1)}{\gamma_w(0)}.$$

That is, $T^2\widehat{\beta}/\sigma_y\sigma_x$ is indeed $O_p(1)$. Given this result, then all other conclusions in Theorem 2 can be established by the same analysis as in the proof of Theorem 2. The only change required in the proof is that all $Z(1)$ in Theorem 2 be replaced by the Z^* process.

A.5. Proof of Theorem 3

To prove item 1, we have

$$\frac{T}{\sigma_y\sigma_x}\widehat{\beta} = \frac{\frac{1}{\sigma_y\sigma_x}\sum_{t=1}^{T} y_t w_t - \left(\frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t\right)\left(\frac{1}{\sigma_x}\sum_{t=1}^{T} w_t\right)}{\frac{1}{T}\sum_{t=1}^{T}(w_t - \overline{w})^2} \Rightarrow \beta_*,$$

where the weak convergence is based on items 1, 4, 6, and 9 of Lemma 1. To prove item 2, we see

$$\frac{1}{\sigma_y}\widehat{\alpha} = \frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t - \frac{\sigma_x^2}{T^2}\cdot\frac{T}{\sigma_y\sigma_x}\widehat{\beta}\cdot\frac{1}{\sigma_x}\sum_{t=1}^{T} w_t \Rightarrow \alpha_*,$$

where the weak convergence is based on item 1 above and items 1 and 4 of Lemma 1, as well as the fact that $\sigma_x^2/T^2 = O(T^{2d_2-1})$ which converges to zero. To prove item 3, we see

$$\frac{1}{\sigma_y^2} s^2 = \frac{1}{T\sigma_y^2} \sum_{t=1}^{T} (y_t - \overline{y})^2 - \frac{\sigma_x^2}{T^2} \left(\frac{T}{\sigma_y \sigma_x} \widehat{\beta} \right)^2 \frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2 \;\Rightarrow\; \sigma_*^2,$$

where the weak convergence is based on item 1 above and items 3 and 6 of Lemma 1, as well as the fact that $\sigma_x^2/T^2 = O(T^{2d_2-1})$ which converges to zero. To prove item 4, we see

$$\frac{T}{\sigma_y^2} s_\beta^2 = \frac{\frac{1}{\sigma_y^2} s^2}{\frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2} \;\Rightarrow\; \frac{\sigma_*^2}{\gamma_w(0)},$$

where the weak convergence is based on item 3 above and item 6 of Lemma 1. To prove item 5, we have

$$\frac{T}{\sigma_y^2} s_\beta^2 = \frac{1}{\sigma_y^2} s^2 \left[1 + \frac{\frac{\sigma_x^2}{T^2} \left(\frac{1}{\sigma_x} \sum_{t=1}^{T} w_t \right)^2}{\frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2} \right] \;\Rightarrow\; \sigma_*^2,$$

where the weak convergence is based on item 3 above and items 4 and 6 of Lemma 1, as well as the fact that $\sigma_x^2/T^2 = O(T^{2d_2-1})$ which converges to zero so that the second term in the bracket does too. Items 6 and 7 are straightforward results of items 1, 2, 4, and 5. To prove item 8, we have

$$\frac{T^2}{\sigma_x^2} R^2 = \frac{\left(\frac{T}{\sigma_y \sigma_x} \widehat{\beta} \right)^2 \frac{1}{T} \sum_{t=1}^{T} (w_t - \overline{w})^2}{\frac{1}{T\sigma_y^2} \sum_{t=1}^{T} (y_t - \overline{y})^2} \;\Rightarrow\; \frac{\gamma_w(0) \cdot \beta_*^2}{\sigma_*^2},$$

where the weak convergence is based on items 1 and 3 above and items 3 and 6 of Lemma 1. To prove item 9, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = [(y_t - \widehat{\alpha} - \widehat{\beta} w_t) - (y_{t-1} - \widehat{\alpha} - \widehat{\beta} w_{t-1})]^2$$

$$= [v_t - \widehat{\beta}(w_t - w_{t-1})]^2$$

$$= v_t^2 - 2\widehat{\beta} v_t (w_t - w_{t-1}) + \widehat{\beta}^2 (w_t - w_{t-1})^2.$$

Also, from item 7 of Lemma 1, we have

$$\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t(w_t - w_{t-1}) = \frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_t - \frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_{t-1}$$

$$= \max\{o_p(T^{\epsilon-1.5+d_2}),\, o_p(T^{\epsilon-1-d_1})\},$$

for any $\epsilon > 0$. Moreover, we know that from item 1 above $\widehat{\beta}/\sigma_y = O_p(T^{d_2-0.5})$ converges in probability to zero. From item 5 of Lemma 1, we also know that $\sum_{t=2}^{T} v_t^2/T \xrightarrow{p} \gamma_v(0)$ and that

$$\frac{1}{T}\sum_{t=2}^{T}(w_t - w_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T} w_t^2 + \frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} w_t w_{t-1}$$

$$\xrightarrow{p} \gamma_w(0) + \gamma_w(0) - 2\,\gamma_w(1).$$

All these results imply DW statistic is

$$\frac{\dfrac{1}{\sigma_y^2}\dfrac{1}{T}\displaystyle\sum_{t=2}^{T} v_t^2 - 2\dfrac{\widehat{\beta}}{\sigma_y}\dfrac{1}{T\sigma_y}\displaystyle\sum_{t=2}^{T} v_t(w_t - w_{t-1}) + \left(\dfrac{\widehat{\beta}}{\sigma_y}\right)^2 \dfrac{1}{T}\displaystyle\sum_{t=2}^{T}(w_t - w_{t-1})^2}{\dfrac{1}{T\sigma_y^2}\displaystyle\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$\xrightarrow{p} 0,$$

where the three terms in the numerator all converge in probability to zero while the denominator converges weakly to σ_*^2 from the result of item 3 above.

A.6. Proof of Theorem 4

To prove item 1, we have

$$\frac{T\sigma_x}{\sigma_y}\widehat{\beta} = \frac{\dfrac{1}{\sigma_y\sigma_x}\displaystyle\sum_{t=1}^{T} v_t x_t - \left(\frac{1}{\sigma_y}\sum_{t=1}^{T} v_t\right)\left(\frac{1}{T\sigma_x}\sum_{t=1}^{T} x_t\right)}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \overline{x})^2} \Rightarrow \beta_*,$$

where the weak convergence is based on items 1, 3, 4, and 9 of Lemma 1. To prove item 2, we see

$$\frac{T}{\sigma_y}\widehat{\alpha} = \frac{1}{\sigma_y}\sum_{t=1}^{T} v_t - \frac{T\sigma_x}{\sigma_y}\widehat{\beta}\cdot\frac{1}{T\sigma_x}\sum_{t=1}^{T} x_t \Rightarrow \alpha_*,$$

where the weak convergence is based on item 1 above and items 1 and 4 of Lemma 1. To prove item 3, we see

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2 - \frac{\sigma_y^2}{T^2}\left(\frac{T\sigma_x}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{p} \gamma_v(0),$$

where the weak convergence is based on item 1 above and items 3 and 6 of Lemma 1, as well as the fact that $\sigma_y^2/T^2 = O(T^{2d_1-1})$ which converges to zero. To prove item 4, we see

$$T\sigma_x^2\cdot s_\beta^2 = \frac{s^2}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \overline{x})^2} \Rightarrow \sigma_{*\beta}^2,$$

where the weak convergence is based on item 3 above and item 3 of Lemma 1. To prove item 5, we have

$$T\cdot s_\alpha^2 = s^2\left[1 + \frac{\left(\dfrac{1}{T\sigma_x}\displaystyle\sum_{t=1}^{T} x_t\right)^2}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \overline{x})^2}\right] \Rightarrow \sigma_{*\alpha}^2,$$

where the weak convergence is based on item 3 above and items 1 and 3 of Lemma 1. Items 6 and 7 are straightforward results of items 1, 2, 4, and 5. To prove item 8, we have

$$\frac{T^2}{\sigma_y^2} R^2 = \frac{\left(\frac{T\sigma_x}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \overline{x})^2}{\frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2} \Rightarrow \frac{\beta_*^2}{\sigma_{*\beta}^2},$$

where the weak convergence is based on item 1 above and items 3 and 6 of Lemma 1. To prove item 9, we first note

$$\begin{aligned}(\widehat{u}_t - \widehat{u}_{t-1})^2 &= [(v_t - \widehat{\alpha} - \widehat{\beta}x_t) - (v_{t-1} - \widehat{\alpha} - \widehat{\beta}x_{t-1})]^2 \\ &= (v_t - v_{t-1} - \widehat{\beta}w_t)^2 \\ &= (v_t - v_{t-1})^2 - 2\widehat{\beta}(v_t - v_{t-1})w_t + \widehat{\beta}^2 w_t^2.\end{aligned}$$

Also, from item 7 of Lemma 1, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})w_t &= \frac{1}{T}\sum_{t=2}^{T} v_t w_t - \frac{1}{T}\sum_{t=2}^{T} v_{t-1} w_t \\ &= \max\{o_p(T^{\epsilon-1+d_1+d_2}),\ o_p(T^{\epsilon-0.5})\},\end{aligned}$$

for any $\epsilon > 0$. Moreover, we know from item 1 above that $\widehat{\beta} = O_p(T^{d_1-d_2-1})$ converges in probability to zero. From item 5 of Lemma 1, we also know that $\sum_{t=2}^{T} w_t^2/T \xrightarrow{p} \gamma_w(0)$ and that

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 &= \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} v_t v_{t-1} \\ &\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\,\gamma_v(1).\end{aligned}$$

Finally,

$$DW = \frac{\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 - 2\widehat{\beta}\cdot\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})w_t + \widehat{\beta}^2\cdot\frac{1}{T}\sum_{t=2}^{T} w_t^2}{\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$\xrightarrow{p} \quad 2\rho_v(0) - 2\rho_v(1),$$

where the second and the third terms in the numerator converge in probability to zero and the denominator converges in probability to $\gamma_v(0)$ from the result of item 3 above.

A.7. Proof of Theorem 5

To prove item 1, we note

$$\frac{T}{\sigma_y}\widehat{\beta} = \frac{\frac{1}{T^2\sigma_y}\sum_{t=1}^{T} t\cdot y_t - \left(\frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t\right)\left(\frac{1}{T^2}\sum_{t=1}^{T} t\right)}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2} \quad \Rightarrow \quad \beta_*,$$

where the weak convergence is due to items 1 and 11 of Lemma 1 and the facts that

$$\frac{1}{T^2}\sum_{t=1}^{T} t \;\rightarrow\; \frac{1}{2}, \qquad \frac{1}{T^3}\sum_{t=1}^{T} t^2 \;\rightarrow\; \frac{1}{3},$$

and

$$\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2 \;\rightarrow\; \frac{1}{12},$$

as T goes to infinity. To prove item 2, we note

$$\frac{1}{\sigma_y}\widehat{\alpha} = \frac{1}{T\sigma_y}\sum_{t=1}^{T} y_t - \frac{T}{\sigma_y}\widehat{\beta}\cdot\frac{1}{T^2}\sum_{t=1}^{T} t \;\Rightarrow\; \alpha_*,$$

where the weak convergence is based on item 1 above and item 1 of Lemma 1. To prove item 3, we have

$$\frac{1}{\sigma_y^2}s^2 = \frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \overline{y})^2 - \left(\frac{T}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2 \Rightarrow \sigma_*^2,$$

where the weak convergence is based on item 1 above and item 3 of Lemma 1. To prove item 4, we note

$$\frac{T^3}{\sigma_y^2}s_\beta^2 = \frac{\frac{1}{\sigma_y^2}s^2}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2} \Rightarrow 12\sigma_*^2,$$

where the weak convergence is based on item 3 above. To prove item 5, we see

$$\frac{T}{\sigma_y^2}s_\alpha^2 = \frac{1}{\sigma_y^2}s^2\left[1 + \frac{\left(\frac{1}{T^2}\sum_{t=1}^{T} t\right)^2}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2}\right] \Rightarrow 4\sigma_*^2,$$

where the weak convergence is based on item 3 above. Items 6 and 7 are direct results from items 1, 2, 4, and 5. To prove item 8, we note

$$R^2 = \frac{\left(\frac{T}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2}{\frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

$$\Rightarrow \frac{\beta_*^2}{12\int_0^1 [B_{0.5+d_1}(s)]^2\, ds - 12\left[\int_0^1 B_{0.5+d_1}(s)\, ds\right]^2},$$

where the weak convergence is based on item 1 above and item 3 of Lemma 1. To prove item 9, we note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = [(y_t - \widehat{\alpha} - \widehat{\beta}\cdot t) - (y_{t-1} - \widehat{\alpha} - \widehat{\beta}(t-1))]^2$$

$$= (v_t - \widehat{\beta})^2 = v_t^2 - 2\widehat{\beta}v_t + \widehat{\beta}^2.$$

Then, by the results of item 1 above and items 4 and 5 of Lemma 1, we have

$$\sigma_y^2 \cdot DW = \frac{\frac{1}{T}\sum_{t=2}^{T} v_t^2 - 2\cdot\frac{T}{\sigma_y}\widehat{\beta}\cdot\frac{\sigma_y^2}{T^2}\cdot\frac{1}{\sigma_y}\sum_{t=2}^{T} v_t + \left(\frac{T}{\sigma_y}\widehat{\beta}\right)^2\frac{\sigma_y^2}{T^2}}{\frac{1}{T\sigma_y^2}\sum_{t=1}^{T}\widehat{u}_t^2} \Rightarrow \frac{\gamma_v(0)}{\sigma_*^2},$$

where $\sigma_y^2/T^2 = O(T^{2d_1-1})$ which converges to zero so that the second and the third terms in the numerator also converge in probability to zero. The denominator is s^2/σ_y^2 which converges weakly to σ_*^2 by the result of item 3.

A.8. Proof of Theorem 6

To prove item 1, we have

$$\frac{T^2}{\sigma_y}\widehat{\beta} = \frac{\frac{1}{T\sigma_y}\sum_{t=1}^{T} t\cdot v_t - \left(\frac{1}{\sigma_y}\sum_{t=1}^{T} v_t\right)\left(\frac{1}{T^2}\sum_{t=1}^{T} t\right)}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2} \Rightarrow \beta_*,$$

where the weak convergence is based on items 4 and 10 of Lemma 1. To prove item 2, we see

$$\frac{T}{\sigma_y}\widehat{\alpha} = \frac{1}{\sigma_y}\sum_{t=1}^{T} v_t - \frac{T^2}{\sigma_y}\widehat{\beta}\cdot\frac{1}{T^2}\sum_{t=1}^{T} t \Rightarrow \alpha_*,$$

where the weak convergence is based on item 1 above and item 4 of Lemma 1. To prove item 3, we see

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2 - \frac{\sigma_y^2}{T^2}\left(\frac{T^2}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2 \xrightarrow{p} \gamma_v(0),$$

where the weak convergence is based on item 1 above and item 6 of Lemma 1, as well as the fact that $\sigma_y^2/T^2 = O(T^{2d_1-1})$ which converges to zero. To prove item 4, we see

$$T^3 \cdot s_\beta^2 = \frac{s^2}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2} \Rightarrow 12\gamma_v(0),$$

where the weak convergence is based on item 3 above. To prove item 5, we have

$$T \cdot s_\alpha^2 = s^2 \left[1 + \frac{\left(\frac{1}{T^2}\sum_{t=1}^{T} t\right)^2}{\frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2}\right] \Rightarrow 4\gamma_v(0),$$

where the weak convergence is based on item 3 above. Items 6 and 7 are straightforward results of items 1, 2, 4, and 5. To prove item 8, we have

$$\frac{T^2}{\sigma_y^2} R^2 = \frac{\left(\frac{T^2}{\sigma_y}\widehat{\beta}\right)^2 \frac{1}{T^3}\sum_{t=1}^{T}\left(t - \frac{1}{T}\sum_{t=1}^{T} t\right)^2}{\frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2} \Rightarrow \frac{\beta_*^2}{12\gamma_v(0)},$$

where the weak convergence is based on item 1 above and item 6 of Lemma 1. To prove item 9, we first note

$$\begin{aligned}(\widehat{u}_t - \widehat{u}_{t-1})^2 &= [(v_t - \widehat{\alpha} - \widehat{\beta} \cdot t) - (v_{t-1} - \widehat{\alpha} - \widehat{\beta}(t-1))]^2 \\ &= (v_t - v_{t-1} - \widehat{\beta})^2 \\ &= (v_t - v_{t-1})^2 - 2\widehat{\beta}(v_t - v_{t-1}) + \widehat{\beta}^2.\end{aligned}$$

Now, from item 4 of Lemma 1, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1}) = \frac{1}{T}\sum_{t=2}^{T} v_t - \frac{1}{T}\sum_{t=2}^{T} v_{t-1} = o_p(1).$$

Moreover, we know from item 1 above that $\widehat{\beta} = O_p(T^{d_1-1.5})$ converges in probability to zero. Also, from item 5 of Lemma 1, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} v_t v_{t-1}$$

$$\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\,\gamma_v(1).$$

Finally,

$$DW = \frac{\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 - 2\widehat{\beta}\cdot\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1}) + \widehat{\beta}^2}{\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$\xrightarrow{p} 2\rho_v(0) - 2\rho_v(1),$$

where the second and the third terms in the numerator converge in probability to zero while the denominator converges in probability to $\gamma_v(0)$ from the result of item 3 above.

TABLE 2–1

REJECTION PERCENTAGES AND MEAN $|t_\beta|$ UNDER MODEL 1

y_t	x_t	T	1%	5%	10%	20%	30%	$\|t_\beta\|$
		125	0.487	0.594	0.652	0.721	0.774	3.0171
$I(0.7)$	$I(0.7)$	250	0.615	0.703	0.747	0.798	0.838	4.2187
		500	0.710	0.776	0.811	0.853	0.882	5.7634
		125	0.678	0.749	0.788	0.831	0.864	5.0990
$I(0.7)$	$I(1.3)$	250	0.765	0.819	0.849	0.882	0.907	7.0356
		500	0.824	0.863	0.885	0.909	0.927	9.7107
		125	0.664	0.746	0.787	0.837	0.868	5.0258
$I(1.3)$	$I(0.7)$	250	0.762	0.822	0.850	0.882	0.905	7.0125
		500	0.825	0.866	0.887	0.912	0.930	9.7648
		125	0.860	0.892	0.910	0.930	0.946	13.5955
$I(1.3)$	$I(1.3)$	250	0.902	0.925	0.937	0.952	0.963	19.5147
		500	0.935	0.951	0.957	0.966	0.972	27.6573

Note: the critical values of the two-tailed t tests are ± 2.576 for $\aleph = 0.01$, $\pm$ 1.96 for $\aleph = 0.05$, $\pm$ 1.645 for $\aleph = 0.10$, $\pm$ 1.282 for $\aleph = 0.20$, $\pm$ 1.0326 for $\aleph = 0.30$. $|t_\beta|$ is the average absolute value of t_β of the simulation.

TABLE 2–2

THE DIVERGENCE RATE OF MEAN $|t_\beta|$ UNDER MODEL 1

For $y_t = I(0.7)$ and $x_t = I(0.7)$,

$$\frac{4.2187}{3.0171} = \underline{0.9887} \cdot 2^{0.5} \quad \frac{5.7634}{4.2187} = \underline{0.9660} \cdot 2^{0.5}.$$

For $y_t = I(0.7)$ and $x_t = I(1.3)$,

$$\frac{7.0356}{5.0990} = \underline{0.9757} \cdot 2^{0.5} \quad \frac{9.7107}{7.0356} = \underline{0.9760} \cdot 2^{0.5}.$$

For $y_t = I(1.3)$ and $x_t = I(0.7)$,

$$\frac{7.0125}{5.0258} = \underline{0.9866} \cdot 2^{0.5} \quad \frac{9.7648}{7.0125} = \underline{0.9846} \cdot 2^{0.5}.$$

For $y_t = I(1.3)$ and $x_t = I(1.3)$,

$$\frac{19.5147}{13.5955} = \underline{1.0150} \cdot 2^{0.5} \quad \frac{27.6573}{19.5147} = \underline{1.0022} \cdot 2^{0.5}.$$

Note: the above numbers are taken from the last column of Table 2–1.

TABLE 2–3

REJECTION PERCENTAGES AND MEAN $|t_\beta|$ UNDER MODEL 2

v_t	w_t	T	1%	5%	10%	20%	30%	$\lvert t_\beta \rvert$
		125	0.117	0.229	0.312	0.430	0.526	1.3110
$I(0.4)$	$I(0.4)$	250	0.164	0.289	0.378	0.492	0.577	1.4887
		500	0.234	0.361	0.445	0.550	0.629	1.7423
		1000	0.312	0.436	0.511	0.604	0.678	2.0569
		125	0.047	0.129	0.203	0.321	0.429	1.0363
$I(0.3)$	$I(0.3)$	250	0.061	0.150	0.230	0.346	0.447	1.0901
		500	0.082	0.177	0.260	0.385	0.484	1.1766
		1000	0.101	0.214	0.301	0.422	0.517	1.2670
		125	0.040	0.115	0.183	0.304	0.413	1.0005
$I(0.4)$	$I(0.2)$	250	0.051	0.137	0.213	0.324	0.423	1.0440
		500	0.068	0.161	0.238	0.356	0.456	1.1177
		1000	0.087	0.194	0.269	0.393	0.490	1.2040
		125	0.043	0.121	0.195	0.312	0.417	1.0134
$I(0.2)$	$I(0.4)$	250	0.052	0.140	0.218	0.330	0.441	1.0616
		500	0.069	0.166	0.247	0.368	0.467	1.1309
		1000	0.089	0.196	0.277	0.400	0.498	1.2141
		125	0.048	0.123	0.192	0.311	0.416	1.0192
$I(0.3)$	$I(0.3)$	250	0.060	0.151	0.229	0.348	0.450	1.0912
		500	0.078	0.180	0.253	0.374	0.473	1.1612
		1000	0.105	0.217	0.296	0.409	0.504	1.2627

Note: the critical values of the two-tailed t tests are ± 2.576 for $\aleph = 0.01$, $\pm$ 1.96 for $\aleph = 0.05$, $\pm$ 1.645 for $\aleph = 0.10$, $\pm$ 1.282 for $\aleph = 0.20$, $\pm$ 1.0326 for $\aleph = 0.30$. For the first four rows of data, a_t and b_t are independent $N(0,1)$. For the last row of data, a_t and b_t are independent $\chi_1^2 - 1$. $|t_\beta|$ is the average absolute value of t_β of the simulation.

TABLE 2–4

THE DIVERGENCE RATE OF MEAN $|t_\beta|$ UNDER MODEL 2

For $v_t = I(0.4)$, $w_t = I(0.4)$, a_t and b_t are independent $N(0,1)$,

$$\frac{1.4887}{1.3110} = \underline{0.9223} \cdot 2^{0.3} \quad \frac{1.7423}{1.4887} = \underline{0.9506} \cdot 2^{0.3} \quad \frac{2.0569}{1.7423} = \underline{0.9589} \cdot 2^{0.3}.$$

For $v_t = I(0.3)$, $w_t = I(0.3)$, a_t and b_t are independent $N(0,1)$,

$$\frac{1.0901}{1.0363} = \underline{0.9815} \cdot 2^{0.1} \quad \frac{1.1766}{1.0901} = \underline{1.0071} \cdot 2^{0.1} \quad \frac{1.2670}{1.1766} = \underline{1.0047} \cdot 2^{0.1}.$$

For $v_t = I(0.4)$, $w_t = I(0.2)$, a_t and b_t are independent $N(0,1)$,

$$\frac{1.0440}{1.0005} = \underline{0.9736} \cdot 2^{0.1} \quad \frac{1.1177}{1.0440} = \underline{0.9989} \cdot 2^{0.1} \quad \frac{1.2040}{1.1177} = \underline{1.0051} \cdot 2^{0.1}.$$

For $v_t = I(0.2)$, $w_t = I(0.4)$, a_t and b_t are independent $N(0,1)$,

$$\frac{1.0616}{1.0134} = \underline{0.9774} \cdot 2^{0.1} \quad \frac{1.1309}{1.0616} = \underline{0.9939} \cdot 2^{0.1} \quad \frac{1.2141}{1.1309} = \underline{1.0017} \cdot 2^{0.1}.$$

For $v_t = I(0.3)$, $w_t = I(0.3)$, a_t and b_t are independent $\chi_1^2 - 1$,

$$\frac{1.0912}{1.0192} = \underline{0.9989} \cdot 2^{0.1} \quad \frac{1.1612}{1.0912} = \underline{0.9929} \cdot 2^{0.1} \quad \frac{1.2627}{1.1612} = \underline{1.0146} \cdot 2^{0.1}.$$

Note: the above numbers are taken from the last column of Table 2–3.

TABLE 2–5

REJECTION PERCENTAGES AND MEAN $|t_\beta|$ UNDER MODELS 3 AND 4

		T	1%	5%	10%	20%	30%	$\|t_\beta\|$
		125	0.172	0.294	0.375	0.487	0.577	1.4985
$I(0.7)$	$I(0.3)$	250	0.251	0.378	0.458	0.560	0.638	1.7864
		500	0.320	0.450	0.525	0.621	0.688	2.0863
		125	0.274	0.401	0.477	0.580	0.653	1.8791
$I(1.3)$	$I(0.3)$	250	0.357	0.483	0.556	0.645	0.710	2.2475
		500	0.449	0.561	0.626	0.703	0.760	2.7299
		125	0.167	0.293	0.374	0.486	0.574	1.4936
$I(0.3)$	$I(0.7)$	250	0.245	0.372	0.454	0.556	0.633	1.7801
		500	0.329	0.459	0.535	0.633	0.703	2.1336
		125	0.273	0.406	0.486	0.590	0.662	1.9003
$I(0.3)$	$I(1.3)$	250	0.369	0.491	0.561	0.649	0.710	2.2892
		500	0.453	0.565	0.628	0.708	0.768	2.7747

Note: the critical values of the two-tailed t tests are ± 2.576 for $\aleph = 0.01$, $\pm$ 1.96 for $\aleph = 0.05$, $\pm$ 1.645 for $\aleph = 0.10$, $\pm$ 1.282 for $\aleph = 0.20$, $\pm$ 1.0326 for $\aleph = 0.30$. $|t_\beta|$ is the average absolute value of t_β of the simulation.

TABLE 2–6

THE DIVERGENCE RATE OF MEAN $|t_\beta|$ UNDER MODELS 3 AND 4

For $y_t = I(0.7)$ and $w_t = I(0.3)$,

$$\frac{1.7864}{1.4985} = \underline{0.9683} \cdot 2^{0.3} \quad \frac{2.0863}{1.7864} = \underline{0.9486} \cdot 2^{0.3}.$$

For $y_t = I(1.3)$ and $w_t = I(0.3)$,

$$\frac{2.2475}{1.8791} = \underline{0.9715} \cdot 2^{0.3} \quad \frac{2.7299}{2.2475} = \underline{0.9866} \cdot 2^{0.3}.$$

For $v_t = I(0.3)$ and $x_t = I(0.7)$,

$$\frac{1.7801}{1.4936} = \underline{0.9681} \cdot 2^{0.3} \quad \frac{2.1336}{1.7801} = \underline{0.9736} \cdot 2^{0.3}.$$

For $v_t = I(0.3)$ and $x_t = I(1.3)$,

$$\frac{2.2892}{1.9003} = \underline{0.9785} \cdot 2^{0.3} \quad \frac{2.7747}{2.2892} = \underline{0.9845} \cdot 2^{0.3}.$$

Note: the above numbers are taken from the last column of Table 2–5.

CHAPTER 3

THE SPURIOUS EFFECT WHEN REGRESSOR AND DISTURBANCE ARE FRACTIONALLY INTEGRATED PROCESSES

3.1. Introduction

This chapter derives the asymptotic distributions for the OLS estimators and corresponding test statistics in the following simple linear regression model:

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \qquad t = 1, 2, \cdots,$$

where the regressor X_t and the disturbance term ε_t are both fractionally integrated long memory processes and independent of each other. We further assume that X_t is always nonstationary while ε_t can be either stationary or nonstationary, and in the latter case the order of integration of ε_t is smaller than that of X_t. In other words, the order of integration of the disturbance term ε_t must be smaller than that of X_t. We also assume $\beta \neq 0$ to get rid of the possibility of spurious regression which we have discussed in Chapter 2.

Similar models have been analyzed by Robinson and Hidalgo (1995) where they assume X_t and ε_t are both stationary long memory processes and then prove central limit theorems for a number of estimators of the slope coefficient β. Kramer (1986) and Phillips and Park (1988) have studied the asymptotic properties of a model with nonstationary $I(1)$ regressors and $I(0)$ disturbance. Moreover, Park and Phillips (1988, 1989) study the multiple regression case where X_t is a $m-$ dimensional process and may be cointegrated.

The main finding of this chapter is that the t ratio for the slope coefficient β diverges as the sample size increases, as long as the order of integration of the disturbance term ε_t is positive; i.e., ε_t has long memory. Consequently, if the traditional critical values are adopted, the null hypothesis for testing any finite value of β tends to be overly rejected and this is what we call the spurious effect. Moreover, it is found that the inclusion of an intercept or a time trend in the regression model does not change the convergence rates of the t ratios even though the asymptotic distributions of the t ratios are different. The cases where the regressor X_t contains a drift are also considered and we find that the convergence rates of the OLS estimators largely the same as those from the cases with a driftless X_t.

The asymptotic behavior of R^2 and DW is as follows: $R^2 \xrightarrow{p} 1$ for all the cases while $DW \xrightarrow{p} 2-2\rho_v(1)$ when ε_t is stationary, and $DW \xrightarrow{p} 0$ when ε_t is nonstationary.

Monte Carlo study is also included to evaluate the small sample properties of the t ratios, R^2, and DW. The Monte Carlo results support the theory quite well.

3.2. The Four Classes of Models

Given the stationary $I(d_1)$ and $I(d_2)$ processes v_t and w_t, and their respective partial sums $y_t = \sum_{j=1}^{t} v_j$ and $x_t = \sum_{j=1}^{t} w_j$, defined in Chapter 2, we considered the following two classes of simple linear regression models:

Model A-0: $C_t = \beta x_t + v_t, \quad \text{where } d_1 > 0,$

Model A-1: $C_t = \alpha + \beta x_t + v_t, \quad \text{where } d_1 > 0,$

Model A-2: $C_t = \alpha + \gamma t + \beta x_t + v_t, \quad \text{where } d_1 > 0,$

Model B-0: $C_t = \beta x_t + y_t, \quad \text{where } d_1 < d_2,$

Model B-1: $C_t = \alpha + \beta x_t + y_t, \quad \text{where } d_1 < d_2,$

Model B-2: $C_t = \alpha + \gamma t + \beta x_t + y_t, \quad \text{where } d_1 < d_2.$

The disturbance terms v_t of the class A models are stationary long memory $I(d_1)$ processes, while in the class B models the disturbance terms y_t are nonstationary $I(1 + d_1)$ processes and their orders of integration are assumed to be smaller than those of the regressor x_t which is also nonstationary $I(1 + d_2)$ processes. The two models with the label "1" do not include the intercept term, while the two models with the label "3" contain a time trend.

To study the effect of a regressor with a drift, we consider

$$x_t^o = \gamma' + x_{t-1}^o + w_t,$$

and, without loss of generality, γ' can be set at 1. Based on such an x_t^o, we have two additional classes of models that correspond to the previous two classes of models:

Model C-0: $C_t = \beta x_t^o + v_t$, where $d_1 > 0$,

Model C-1: $C_t = \alpha + \beta x_t^o + v_t$, where $d_1 > 0$,

Model C-2: $C_t = \alpha + \gamma t + \beta x_t^o + v_t$, where $d_1 > 0$,

Model D-0: $C_t = \beta x_t^o + y_t$, where $d_1 < d_2$

Model D-1: $C_t = \alpha + \beta x_t^o + y_t$, where $d_1 < d_2$

Model D-2: $C_t = \alpha + \gamma t + \beta x_t^o + y_t$, where $d_1 < d_2$.

3.3. The Class A Models

In this section we derive the asymptotic distributions of the OLS estimators and the corresponding test statistics for the class A models where the regressor x_t is a driftless nonstationary $I(1 + d_2)$ processes and the disturbance term is a stationary long memory $I(d_1)$ process with $d_1 > 0$:

Model A-0: $C_t = \beta x_t + v_t$,

Model A-1: $C_t = \alpha + \beta x_t + v_t$,

Model A-2: $C_t = \alpha + \gamma t + \beta x_t + v_t$.

Let their OLS estimation be denoted as follows:

Model A-0: $C_t = \widehat{\beta}_{(0)} x_t + \widehat{u}_t,$

Model A-1: $C_t = \widehat{\alpha}_{(1)} + \widehat{\beta}_{(1)} x_t + \widehat{u}_t,$

Model A-2: $C_t = \widehat{\alpha}_{(2)} + \widehat{\gamma}_{(2)} t + \widehat{\beta}_{(2)} x_t + \widehat{u}_t.$

We also adopt the following notation for the various statistics from the OLS estimation. Let $s^2_{\beta_{(0)}}$, $s^2_{\beta_{(1)}}$ and $s^2_{\beta_{(2)}}$, respectively, the variances of the corresponding OLS estimators of β, from which we have the t ratios $t_{\beta_{(0)}} = (\widehat{\beta}_{(0)} - \beta)/s_{\beta_{(0)}}$, $t_{\beta_{(1)}} = (\widehat{\beta}_{(1)} - \beta)/s_{\beta_{(1)}}$, and $t_{\beta_{(2)}} = (\widehat{\beta}_{(2)} - \beta)/s_{\beta_{(2)}}$. Furthermore, the notation introduced in Chapter 2 will be used repeatedly throughout this chapter. In particular $B_{0.5+d_1}(s)$ and $B_{0.5+d_2}(s)$ denote, unless otherwise stated, two independent fractional Brownian motions. It will substantially simplify our subsequent formulae to write these as $B_1(s)$ and $B_2(s)$, respectively. Thus, we will frequently use $B_1(s)$ and $B_2(s)$ in place of $B_{0.5+d_1}(s)$ and $B_{0.5+d_2}(s)$.

Theorem 1. *Given that Assumption 1 of Chapter 2 holds, then as $T \to \infty$, we have the following results:*

1. $$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \Rightarrow \frac{\displaystyle\int_0^1 B_2(s)\, dB_1(s)}{\displaystyle\int_0^1 [B_2(s)]^2\, ds} \equiv \beta_{0*}.$$

   Note that $\sigma_y/T\sigma_x = O(T^{d_1-d_2-1})$.

2. $$\frac{T}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \Rightarrow$$

$$\frac{B_1(1)\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds\right]\left[\int_0^1 B_2(s)\, dB_1(s)\right]}{\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds\right]^2} \equiv \alpha_{1*}.$$

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \Rightarrow$$

$$\frac{-B_1(1)\left[\int_0^1 B_2(s)\, ds\right] + \int_0^1 B_2(s)\, dB_1(s)}{\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds\right]^2} \equiv \beta_{1*}.$$

Note that $\sigma_y/T = O(T^{d_1-0.5})$.

3. $$\frac{T}{\sigma_y}\left(\widehat{\alpha}_{(2)} - \alpha\right) \Rightarrow \frac{\hbar_1}{\Delta} \equiv \alpha_{2*},$$

$$\frac{T^2}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right) \Rightarrow \frac{\hbar_2}{\Delta} \equiv \gamma_{2*},$$

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \Rightarrow \frac{\hbar_3}{\Delta} \equiv \beta_{2*},$$

where

$$\Delta \equiv \frac{\int_0^1 [B_2(s)]^2\, ds}{12} - \frac{\left[\int_0^1 B_2(s)\, ds\right]^2}{3} - \left[\int_0^1 s\, B_2(s)\, ds\right]^2 + \left[\int_0^1 B_2(s)\, ds\right]\left[\int_0^1 sB_2(s)\, ds\right],$$

$$\hbar_1 \equiv B_1(1)\left\{\frac{\int_0^1 [B_2(s)]^2\, ds}{3} - \left[\int_0^1 s\, B_2(s)\, ds\right]^2\right\}$$

$$+ [B_1(1)]\left\{\left[\int_0^1 B_2(s)\, ds\right]\left[\int_0^1 s\, B_2(s)\, ds\right] - \frac{\int_0^1 [B_2(s)]^2\, ds}{2}\right\}$$

$$- \left[\int_0^1 B_1(s)\, ds\right]\left\{\left[\int_0^1 B_2(s)\, ds\right]\left[\int_0^1 s\, B_2(s)\, ds\right] - \frac{\int_0^1 [B_2(s)]^2\, ds}{2}\right\}$$

$$+ \int_0^1 B_2(s)\, dB_1(s)\left[-\frac{\int_0^1 B_2(s)\, ds}{3} + \frac{\int_0^1 s\, B_2(s)\, ds}{2}\right],$$

$$\hbar_2 \equiv B_1(1)\left\{\left[\int_0^1 B_2(s)\, ds\right]\left[\int_0^1 s\, B_2(s)\, ds\right] - \frac{\int_0^1 [B_2(s)]^2\, ds}{2}\right\}$$

$$+ \left[B_1(1) - \int_0^1 B_1(s)\, ds\right]\left\{\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds\right]^2\right\}$$

$$+ \int_0^1 B_2(s)\, dB_1(s)\left[\frac{\int_0^1 B_2(s)\, ds}{2} - \int_0^1 s\, B_2(s)\, ds\right],$$

$$\hbar_3 \equiv B_1(1)\left[-\frac{\int_0^1 B_2(s)\,ds}{3}+\frac{\int_0^1 s\,B_2(s)\,ds}{2}\right]$$

$$+\left[B_1(1)-\int_0^1 B_1(s)\,ds\right]\left[\frac{\int_0^1 B_2(s)\,ds}{2}-\int_0^1 s\,B_2(s)\,ds\right]$$

$$+\frac{\int_0^1 B_2(s)\,dB_1(s)}{12}.$$

Note that $\sigma_y/T^2 = O(T^{d_1-1.5})$.

Theorem 1 indicates that the OLS estimators of the regression coefficients are consistent in all three setups of the class A models. The convergence rates of the OLS estimators of α and β are independent of whether the time trend is included in the model or not while their asymptotic distributions are different among the three different setups. It is also interesting to note that the asymptotic distributions of $(\widehat{\alpha}_{(1)} - \alpha)$ and $(\widehat{\beta}_{(1)} - \beta)$ in Model A-1 are the same as those of $\widehat{\alpha}$ and $\widehat{\beta}$ in Model 4 of Chapter 2, since the latter are nothing but the special cases of the former with $\alpha = \beta = 0$.

The asymptotic distributions of the t ratios for the OLS estimators of β, R^2, and the Durbin-Watson statistic DW are presented in the next theorem. We refer to these nonstandard results as the spurious effects in class A models.

Theorem 2. *Given that Assumption 1 of Chapter 2 holds, then as $T \to \infty$, we have the following results:*

1. $$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(0)}} \Rightarrow \beta_{0*}\left\{\frac{1}{\gamma_v(0)}\int_0^1 [B_2(s)]^2\, ds\right\}^{1/2},$$

   where β_{0*} is defined in item 1 of Theorem 1.

   $R^2 \xrightarrow{p} 1.$

   $DW \xrightarrow{p} 2 - 2\rho_v(1).$

2. $$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(1)}} \Rightarrow \beta_{1*}\left\{\frac{1}{\gamma_v(0)}\left\{\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds\right]^2\right\}\right\}^{1/2},$$

   where β_{1*} is defined in item 2 of Theorem 1.

   $R^2 \xrightarrow{p} 1.$

   $DW \xrightarrow{p} 2 - 2\rho_v(1).$

3. $$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(2)}} \Rightarrow \beta_{2*}\sqrt{\frac{12\Delta}{\gamma_v(0)}},$$

   where Δ and β_{2*} are defined in item 3 of Theorem 1.

   $R^2 \xrightarrow{p} 1.$

   $DW \xrightarrow{p} 2 - 2\rho_v(1).$

The t ratios for the OLS estimators of β in all three setups diverge at the same T^{d_1} rate. This finding implies spurious effects in t tests in that the null hypothesis for testing any finite value of β tends to be overly rejected. Theorem 2 also shows that $R^2 \xrightarrow{p} 1$ and $DW \xrightarrow{p} 2 - 2\rho_v(1)$ for all three setups.

3.4. The Class B Models

In this section we derive the asymptotic distributions of the OLS estimators and the corresponding test statistics for the class B models where the regressor x_t is a driftless nonstationary $I(1+d_2)$ processes and the disturbance term is a nonstationary long memory $I(1+d_1)$ process with $d_1 < d_2$:

$$\text{Model B-0:} \qquad C_t = \beta x_t + y_t,$$

$$\text{Model B-1:} \qquad C_t = \alpha + \beta x_t + y_t,$$

$$\text{Model B-2:} \qquad C_t = \alpha + \gamma t + \beta x_t + y_t.$$

The notation for the OLS estimators and test statistics will be the same as that used for the class A models in the previous section.

Theorem 3. *Given that Assumption 1 of Chapter 2 holds, then, as $T \longrightarrow \infty$, we have the following results:*

1. $$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \Rightarrow \frac{\int_0^1 B_1(s)\, B_2(s)\, ds}{\int_0^1 [B_2(s)]^2\, ds} \equiv \beta_{0*}.$$

Note that $\sigma_y/\sigma_x = O(T^{d_1-d_2})$.

2. $$\frac{1}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \Rightarrow$$

$$\frac{\left[\int_0^1 B_1(s)\,ds\right]\int_0^1 [B_2(s)]^2\,ds - \left[\int_0^1 B_2(s)\,ds\right]\left[\int_0^1 B_1(s)\,B_2(s)\,ds\right]}{\int_0^1 [B_2(s)]^2\,ds - \left[\int_0^1 B_2(s)\,ds\right]^2}$$

$$\equiv \quad \alpha_{1*}.$$

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \quad \Rightarrow$$

$$\frac{\int_0^1 B_1(s)\,B_2(s)\,ds - \left[\int_0^1 B_1(s)\,ds\right]\left[\int_0^1 B_2(s)\,ds\right]}{\int_0^1 [B_2(s)]^2\,ds - \left[\int_0^1 B_2(s)\,ds\right]^2}$$

$$\equiv \beta_{1*}.$$

3. $\frac{1}{\sigma_y}\left(\widehat{\alpha}_{(2)} - \alpha\right) \quad \Rightarrow \quad \frac{\zeta_1}{\Delta} \equiv \alpha_{2*},$

$$\frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right) \quad \Rightarrow \quad \frac{\zeta_2}{\Delta} \equiv \gamma_{2*},$$

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \quad \Rightarrow \quad \frac{\zeta_3}{\Delta} \equiv \beta_{2*},$$

where Δ is defined in item 3 of Theorem 1.

$$\zeta_1 \equiv \int_0^1 B_1(s)\,ds \left\{ \frac{\int_0^1 [B_2(s)]^2\,ds}{3} - \left[\int_0^1 s\,B_2(s)\,ds\right]^2 \right\}$$

$$+\int_0^1 s\,B_1(s)\,ds\left\{\left[\int_0^1 B_2(s)\,ds\right]\left[\int_0^1 s\,B_2(s)\,ds\right]-\frac{\int_0^1 [B_2(s)]^2\,ds}{2}\right\}$$

$$+\int_0^1 B_2(s)\,B_1(s)\,ds\left[-\frac{\int_0^1 B_2(s)\,ds}{3}+\frac{\int_0^1 s\,B_2(s)\,ds}{2}\right],$$

$$\zeta_2\equiv\left[\int_0^1 B_1(s)\,ds\right]\left[\int_0^1 B_2(s)\,ds\right]\left[\int_0^1 s\,B_2(s)\,ds\right]$$

$$-\left[\int_0^1 B_1(s)ds\right]\frac{\int_0^1 [B_2(s)]^2\,ds}{2}$$

$$+\int_0^1 s\,B_1(s)\,ds\left\{\int_0^1 [B_2(s)]^2\,ds-\left[\int_0^1 B_2(s)\,ds\right]^2\right\}$$

$$+\int_0^1 B_2(s)\,B_1(s)\,ds\left[\frac{\int_0^1 B_2(s)\,ds}{2}-\int_0^1 s\,B_2(s)\,ds\right],$$

$$\zeta_3\equiv\int_0^1 B_1(s)\,ds\left[-\frac{\int_0^1 B_2(s)\,ds}{3}+\frac{\int_0^1 s\,B_2(s)\,ds}{2}\right]$$

$$+\int_0^1 s\,B_1(s)\,ds\left[\frac{\int_0^1 B_2(s)\,ds}{2}-\int_0^1 s\,B_2(s)\,ds\right]$$

$$+\frac{\int_0^1 B_2(s)\,B_1(s)\,ds}{12}.$$

In the class B models the OLS estimators of β and γ converge while those of the intercept α diverge. That is, the OLS estimators of β and γ are consistent but those of α are not. These results are different from what we have derived for the class A models, where all the OLS estimators are consistent. Also note that the asymptotic distributions of $(\widehat{\alpha}_{(1)} - \alpha)$ and $(\widehat{\beta}_{(1)} - \beta)$ in Model B-1 are identical to that of $\widehat{\alpha}$ and $\widehat{\beta}$ in Model 1 of Chapter 2 since Model 1 of Chapter 2 is simply a special case of Model B-1, just like Model 4 of Chapter 2 is a special case of Model A-1.

The following theorem gives the asymptotic distributions of t ratios, R^2 and DW for the class B models.

Theorem 4. *Given that Assumption 1 of Chapter 2 holds, then as $T \longrightarrow \infty$, we have the following results:*

1. $$\frac{1}{\sqrt{T}} t_{\beta_{(0)}} \Rightarrow \frac{\beta_{0*}}{\left\{ \dfrac{\int_0^1 [B_1(s)]^2\,ds}{\int_0^1 [B_2(s)]^2\,ds} - \left\{ \dfrac{\int_0^1 B_1(s)\,B_2(s)\,ds}{\int_0^1 [B_2(s)]^2\,ds} \right\}^2 \right\}^{1/2}},$$

where β_{0*} is defined in item 1 of Theorem 3.

$$R^2 \xrightarrow{p} 1.$$

$DW \xrightarrow{p} 0.$

2. $$\frac{1}{\sqrt{T}} t_{\beta_{(1)}} \Rightarrow \frac{\beta_{1*}}{\left\{ \dfrac{\sigma_{1*}^2}{\int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds \right]^2} \right\}^{1/2}},$$

where β_{1*} is defined in item 2 of Theorem 3.

$$\sigma_{1*}^2 \equiv \int_0^1 [B_1(s)]^2\, ds - \left[\int_0^1 B_1(s)\, ds \right]^2 - \beta_{1*}^2 \left\{ \int_0^1 [B_2(s)]^2\, ds - \left[\int_0^1 B_2(s)\, ds \right]^2 \right\}.$$

$R^2 \xrightarrow{p} 1.$

$DW \xrightarrow{p} 0.$

3. $$\frac{1}{\sqrt{T}} t_{\beta_{(2)}} \Rightarrow \beta_{2*} \sqrt{\frac{12\Delta}{\sigma_{2*}^2}},$$

where Δ is defined in item 3 of Theorem 1.

$$\sigma_{2*}^2 \equiv \int_0^1 [B_1(s)]^2\, ds - \left[\int_0^1 B_1(s)\, ds \right]^2 + \frac{\gamma_{2*}^2}{12} + \beta_{2*}^2 \int_0^1 [B_2(s)]^2\, ds$$

$$- \beta_{2*}^2 \left[\int_0^1 B_2(s)\, ds \right]^2 - 2\gamma_{2*} \int_0^1 s\, B_1(s)\, ds + \gamma_{2*} \int_0^1 B_1(s)\, ds$$

$$- 2\beta_{2*} \left\{ \int_0^1 B_1(s) B_2(s)\, ds - \left[\int_0^1 B_1(s)\, ds \right] \left[\int_0^1 B_2(s)\, ds \right] \right\}$$

$$+ 2\gamma_{2*}\beta_{2*} \left[\int_0^1 s\, B_2(s)\, ds - \frac{1}{2} \int_0^1 B_2(s)\, ds \right],$$

where γ_{2*} and β_{2*} are defined in item 3 of Theorem 3.

$R^2 \xrightarrow{p} 1.$

$DW \xrightarrow{p} 0.$

The t ratios for β diverge at the same $T^{1/2}$ rate, which indicates that reducing the order of integration in the disturbance term from above 0.5 as in the class A models to below 0.5 as in the class B models causes the divergence rate of t ratio to decrease by the order of $T^{d_1 - 1/2}$. It is the same as the conclusion we arrived in Chapter 2. Theorem 4 also shows that $R^2 \xrightarrow{p} 1$ and $DW \xrightarrow{p} 0$ for all three setups of the class B models. That is, while R^2 still grows to become 1 as T increases, DW reduces to 0, instead of $2 - 2\rho_v(1)$ as in the class A models.

From the results of Theorems 2 and 4 for both class A and class B models, we conclude that the t tests for the slope coefficients are affected by the spurious effects due to the long memory in the regressor and in the disturbance term.

3.5. The Class C Models

In this section we derive the asymptotic distributions of the OLS estimators and the corresponding test statistics for the class C models where the regressor x_t is a nonstationary $I(1+d_2)$ processes with a drift and the disturbance term is a stationary long memory $I(d_1)$ process with $d_1 > 0$:

$$\text{Model C-0:} \qquad C_t = \beta x_t^o + v_t,$$

$$\text{Model C-1:} \qquad C_t = \alpha + \beta x_t^o + v_t,$$

$$\text{Model C-2:} \qquad C_t = \alpha + \gamma t + \beta x_t^o + v_t.$$

Theorem 5 presents the asymptotic distributions of OLS estimators of α, γ, and β, while Theorem 6 shows the asymptotic distributions of t ratios, R^2, and DW.

Theorem 5. *Given that Assumption 1 of Chapter 2 holds, then as* $T \to \infty$, *we have the following results:*

1. $$\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \Rightarrow 3B_1(1) - 3\int_0^1 B_1(s)\,ds \equiv \beta_{0*}.$$

2. $$\frac{T}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \Rightarrow 6\int_0^1 B_1(s)\,ds - 2B_1(1) \equiv \alpha_{1*}.$$

$$\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \Rightarrow 6B_1(1) - 12\int_0^1 B_1(s)\,ds \equiv \beta_{1*}.$$

3. $$\frac{T}{\sigma_y}\left(\widehat{\alpha}_{(2)} - \alpha\right) \Rightarrow \alpha_{2*},$$

$$\frac{T^2}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right) \Rightarrow \gamma_{2*},$$

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \Rightarrow \beta_{2*},$$

where α_{2*}, γ_{2*} and β_{2*} are defined in item 3 of Theorem 1.

Theorem 6. *Given that Assumption 1 of Chapter 2 holds, then as $T \to \infty$, we have the following results:*

1. $$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(0)}} \Rightarrow \frac{\beta_{0*}}{\sqrt{3\gamma_v(0)}},$$

where β_{0*} is defined in item 1 of Theorem 5.

$$R^2 \xrightarrow{p} 1.$$

$$DW \xrightarrow{p} 2 - 2\rho_v(1).$$

2. $$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(1)}} \Rightarrow \frac{\beta_{1*}}{\sqrt{12\gamma_v(0)}},$$

where β_{1*} is defined in item 2 of Theorem 5.

$$R^2 \xrightarrow{p} 1.$$

$$DW \xrightarrow{p} 2 - 2\rho_v(1).$$

3. $\dfrac{\sqrt{T}}{\sigma_y} t_{\beta_{(2)}} \Rightarrow \beta_{2*}\sqrt{\dfrac{12\Delta}{\gamma_v(0)}},$

where β_{2*} and Δ is defined in item 3 of Theorem 1.

$R^2 \xrightarrow{p} 1.$

$DW \xrightarrow{p} 2 - 2\rho_v(1).$

By comparing the results of Theorems 5 and 6 for the class C models with those of Theorems 1 and 2 for the class A models, we learn how the drift in the regressor affects the asymptotic behavior of the OLS estimators and t test statistics. Two interesting findings are worth mentioning. First, the asymptotic distributions for the statistic $t_{\beta_{(2)}}$ are identical in both Models A-2 and C-2 where the time trend is included as a regressor. Secondly, the asymptotic distributions of $\widehat{\alpha}_{(1)}-\alpha$, $\widehat{\beta}_{(1)}-\beta$ and $t_{\beta_{(1)}}$ in Model C-1 are the same as those of Model 6 in Chapter 2 in which the regressor is the time trend. The intuition is that an nonstationary fractionally integrated process x_t with a nonzero drift in Model C-1 behaves asymptotically like a deterministic trend as in Model 6 of Chapter 2.

3.6. The Class D Models

In this section we derive the asymptotic distributions of the OLS estimators and the corresponding test statistics for the class D models where the regressor x_t is a nonstationary $I(1 + d_2)$ processes with a drift and the disturbance term is a nonstationary $I(1+d_1)$ process with $d_1 < d_2$:

$$\text{Model D-0:} \qquad C_t = \beta x_t^o + y_t,$$

$$\text{Model D-1:} \qquad C_t = \alpha + \beta x_t^o + y_t,$$

$$\text{Model D-2:} \qquad C_t = \alpha + \gamma t + \beta x_t^o + y_t.$$

Theorem 7. *Given that Assumption 1 of Chapter 2 holds, then $T \to \infty$, we have the following results:*

1. $$\frac{T}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \;\Rightarrow\; 3\int_0^1 s\, B_1(s)\, ds \;\equiv\; \beta_{0*}.$$

2. $$\frac{1}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \;\Rightarrow\; 4\int_0^1 B_1(s)\, ds - 6\int_0^1 s\, B_1(s)\, ds \;\equiv\; \alpha_{1*}.$$

$$\frac{T}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \;\Rightarrow\; 12\int_0^1 s\, B_1(s)\, ds - 6\int_0^1 B_1(s)\, ds \;\equiv\; \beta_{1*}.$$

3. $$\frac{1}{\sigma_y}\left(\widehat{\alpha}_{(2)} - \alpha\right) \;\Rightarrow\; \alpha_{2*},$$

$$\frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right) \;\Rightarrow\; \gamma_{2*},$$

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \;\Rightarrow\; \beta_{2*},$$

where α_{2*}, γ_{2*} and β_{2*} are defined in item 3 of Theorem 3.

Theorem 8. *Given that Assumption 1 of Chapter 2 holds, then as $T \to \infty$, we have the following results:*

1. $$\frac{1}{\sqrt{T}} t_{\beta_{(0)}} \Rightarrow \frac{\beta_{0*}}{\left\{3\int [B_1(s)]^2\,ds - 9\left[\int s\,B_1(s)\,ds\right]^2\right\}^{1/2}},$$

   where β_{0*} is defined in item 1 of Theorem 7.

   $R^2 \xrightarrow{p} 1.$

   $DW \xrightarrow{p} 0.$

2. $$\frac{1}{\sqrt{T}} t_{\beta_{(1)}} \Rightarrow \frac{\beta_{1*}}{\sqrt{12\sigma_{1*}^2}},$$

   where β_{1*} is defined in item 2 of Theorem 7.

   $$\sigma_{1*}^2 \equiv \int_0^1 [B_1(s)]^2\,ds - \left[\int_0^1 B_1(s)\,ds\right]^2 - 12\left[\int_0^1 s\,B_1(s)\,ds - \frac{\int_0^1 B_1(s)\,ds}{2}\right]^2.$$

   $R^2 \xrightarrow{p} 1.$

   $DW \xrightarrow{p} 0.$

3. $$\frac{1}{\sqrt{T}} t_{\beta_{(2)}} \Rightarrow \beta_{2*}\sqrt{\frac{12\Delta}{\sigma_{2*}^2}},$$

where β_{2*} is defined in item 3 of Theorem 3, Δ is defined in item 3 of Theorem 1, and σ_{2*}^2 is defined in item 3 of Theorem 4.

$$R^2 \xrightarrow{p} 1.$$

$$DW \xrightarrow{p} 0.$$

By comparing the results of Theorems 7 and 8 for the class D models with those of Theorems 3 and 4 for the class B models, we learn how the drift in the regressor affects the asymptotic behavior of the OLS estimators and t test statistics. Three interesting findings are worth mentioning. First, the asymptotic distributions for the statistic $t_{\beta_{(2)}}$ are identical in both Models B-2 and D-2 where the time trend is included. Secondly, the asymptotic distributions of $\widehat{\alpha}_{(1)} - \alpha$, $\widehat{\beta}_{(1)} - \beta$ and $t_{\beta_{(1)}}$ in Model D-1 are the same as those of Model 5 in Chapter 2 in which the regressor is the time trend. The intuition is that an nonstationary fractionally integrated process x_t with a nonzero drift in Model D-1 behaves asymptotically like a deterministic trend as in Model 5 of Chapter 2. Thirdly, if $d_2 < d_1$, then the OLS estimator $\widehat{\beta}_{(2)}$ in Model D-2 diverges so that $\widehat{\beta}_{(2)}$ can be an inconsistent estimator of β, in which case R^2 will not converge in probability to 1.

3.7. Monte Carlo Experiments

Monte Carlo experiments are conducted to investigate the relevance of the theory in small sample applications. The Monte Carlo experiment for each model is based on 10,000 replications with three different sample sizes (T). The algorithm for simulating the stationary fractionally integrated processes v_t and w_t is the same as in Chapter 2, while the nonstationary series y_t and w_t are constructed as their partial sums. Given the four series v_t, w_t, y_t, and w_t, the series for the dependent variable C_t can be easily computed based on various model specifications. In our Monte Carlo study we focus on the specification that $\alpha = \beta = 1$ and $\gamma = 0$.

For two model specifications: $x_t = I(1)$, $\varepsilon_t = I(0.3)$; and $x_t = I(1)$, $\varepsilon_t = I(0.7)$, Table 3–1 contains the results on the rejection percentages of the two-tailed t test for the null hypothesis

$$H_o\colon \beta = 1$$

at various levels of significance ($\aleph$). Table 3–1 also contains the average R^2 and the average DW and the average of the absolute value of t_β. The Monte Carlo results on another pair of specifications: $x_t^o = x_t + t$ $\varepsilon_t = I(0.3)$; and $x_t^o = x_t + t$, $\varepsilon_t = I(0.7)$, are in Table 3–3.

Theorems 2 and 6 indicate that t ratios diverge at the T^{d_1} rate irrespective of whether there is a drift in the regressor x_t or not. Furthermore, Theorems 4 and 8 show that t ratios diverge at the $T^{0.5}$ rate irrespective of whether there is a drift in the regressor x_t or not. So the probability of rejecting null hypothesis of $\beta = 1$ should increase as T increases. All of the rejection percentages at every value of $\aleph$ in Table 3–1 and Table 3–3 support these theoretical results. Moreover, based on the simulated results on $|t_\beta|$ we estimate their divergence rates in Table 3–2 and Table 3–4. It is found that the divergence rates are quite close to the theoretical rates $T^{0.3}$ and $T^{0.5}$, respectively.

Let us now consider the asymptotic behavior of R^2 and Durbin-Watson statistic DW. Our theory suggests that $R^2 \xrightarrow{p} 1$ for all twelve models. As shown in Tables 3–1 and 3–3, the average R^2 increases with T increases and they are all very close to 1. Our theory also says that when the disturbance term is stationary, then $DW \xrightarrow{p} 2 - 2\rho_v(1)$, which is 1.1429 when $v_t = I(0.3)$. The results in Tables 3–1 and 3–3 indicate that DW approaches the theoretical value as T increases. Finally, when the disturbance term is nonstationary, Theorems 4 and 8 suggest that $DW \xrightarrow{p} 0$. Our simulation results show that DW does decrease as T increases.

3.8. Conclusion

In this chapter we derive the asymptotic distributions of the OLS estimators and the corresponding test statistics for four classes of simple linear regression models where the regressor and the disturbance term are both fractionally integrated processes. The main finding is that the t ratios for the slope coefficients β are divergent and therefore the null hypothesis for testing any finite value of β tends to be overly rejected. This latter result is referred to as the spurious effect.

3.9. Mathematical Proof

Most of the proofs of our theorems are based on the functional central limit theorem (CLT) and the continuous mapping theorem (CMT).

A.1. Proof of Theorem 1

To prove item 1, we see

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)}-\beta\right)=\frac{\dfrac{1}{\sigma_y\sigma_x}\displaystyle\sum_{t=1}^{T}x_tv_t}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}x_t^2}\Rightarrow\beta_{0*},$$

where the weak convergence is based on items 2 and 9 of Lemma 1 of Chapter 2 and the CMT.

To prove item 2, we see

$$\begin{bmatrix}\widehat{\alpha}_{(1)}-\alpha\\ \widehat{\beta}_{(1)}-\beta\end{bmatrix}=\begin{bmatrix}T & \displaystyle\sum_{t=1}^{T}x_t\\ \displaystyle\sum_{t=1}^{T}x_t & \displaystyle\sum_{t=1}^{T}x_t^2\end{bmatrix}^{-1}\begin{bmatrix}\displaystyle\sum_{t=1}^{T}v_t\\ \displaystyle\sum_{t=1}^{T}x_tv_t\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\left(\displaystyle\sum_{t=1}^{T}x_t^2\right)\left(\displaystyle\sum_{t=1}^{T}v_t\right)-\left(\displaystyle\sum_{t=1}^{T}x_t\right)\left(\displaystyle\sum_{t=1}^{T}x_tv_t\right)}{T\displaystyle\sum_{t=1}^{T}(x_t-\bar{x})^2}\\ \dfrac{-\left(\displaystyle\sum_{t=1}^{T}x_t\right)\left(\displaystyle\sum_{t=1}^{T}v_t\right)+T\left(\displaystyle\sum_{t=1}^{T}x_tv_t\right)}{T\displaystyle\sum_{t=1}^{T}(x_t-\bar{x})^2}\end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} \frac{T}{\sigma_y}\left(\widehat{\alpha}_{(1)}-\alpha\right) \\ \frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)}-\beta\right) \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\frac{1}{T\sigma_y\sigma_x^2}\left[\left(\sum_{t=1}^{T} x_t^2\right)\left(\sum_{t=1}^{T} v_t\right)-\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} v_t x_t\right)\right]}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t-\bar{x})^2} \\ \dfrac{\frac{1}{T\sigma_y\sigma_x}\left[-\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} v_t\right)+T\left(\sum_{t=1}^{T} x_t v_t\right)\right]}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t-\bar{x})^2} \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{1*} \\ \beta_{1*} \end{bmatrix},$$

where the weak convergence is based on items 1, 2, 3, 4 and 9 of Lemma 1 of Chapter 2 and the CMT.

To prove item 3, we see

$$\begin{bmatrix} \widehat{\alpha}_{(2)}-\alpha \\ \widehat{\gamma}_{(2)}-\gamma \\ \widehat{\beta}_{(2)}-\beta \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} v_t \\ \sum_{t=1}^{T} t v_t \\ \sum_{t=1}^{T} x_t v_t \end{bmatrix} = \frac{1}{D}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix},$$

where

$$k_1 = \left(\sum_{t=1}^{T} v_t\right) \left\{ \left(\sum_{t=1}^{T} t^2\right) \left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=1}^{T} t x_t\right)^2 \right\}$$

$$+ \left(\sum_{t=1}^{T} t v_t\right) \left(\sum_{t=1}^{T} x_t\right) \left(\sum_{t=1}^{T} t x_t\right) - \left(\sum_{t=1}^{T} t v_t\right) \left(\sum_{t=1}^{T} t\right) \left(\sum_{t=1}^{T} x_t^2\right)$$

$$+ \left(\sum_{t=1}^{T} x_t v_t\right) \left\{ - \left(\sum_{t=1}^{T} t^2\right) \left(\sum_{t=1}^{T} x_t\right) + \left(\sum_{t=1}^{T} t\right) \left(\sum_{t=1}^{T} t x_t\right) \right\},$$

$$k_2 = \left(\sum_{t=1}^{T} v_t\right) \left\{ \left(\sum_{t=1}^{T} t x_t\right) \left(\sum_{t=1}^{T} x_t\right) - \left(\sum_{t=1}^{T} t\right) \left(\sum_{t=1}^{T} x_t^2\right) \right\}$$

$$+ \left(\sum_{t=1}^{T} t v_t\right) \cdot T \left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=1}^{T} t v_t\right) \left(\sum_{t=1}^{T} x_t\right)^2$$

$$+ \left(\sum_{t=1}^{T} x_t v_t\right) \left\{ \left(\sum_{t=1}^{T} t\right) \left(\sum_{t=1}^{T} x_t\right) - T \left(\sum_{t=1}^{T} t x_t\right) \right\},$$

$$k_3 = \left(\sum_{t=1}^{T} v_t\right) \left\{ - \left(\sum_{t=1}^{T} t^2\right) \left(\sum_{t=1}^{T} x_t\right) + \left(\sum_{t=1}^{T} t x_t\right) \left(\sum_{t=1}^{T} t\right) \right\}$$

$$+ \left(\sum_{t=1}^{T} t v_t\right) \left(\sum_{t=1}^{T} t\right) \left(\sum_{t=1}^{T} x_t\right) - T \left(\sum_{t=1}^{T} t x_t\right) \left(\sum_{t=1}^{T} t v_t\right)$$

$$+ \left(\sum_{t=1}^{T} x_t v_t\right) \left\{ T \left(\sum_{t=1}^{T} t^2\right) - \left(\sum_{t=1}^{T} t\right)^2 \right\},$$

$$D = T\left(\sum_{t=1}^{T} t^2\right)\left(\sum_{t=1}^{T} x_t^2\right) + 2\left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} tx_t\right)\left(\sum_{t=1}^{T} x_t\right)$$

$$-\left(\sum_{t=1}^{T} x_t\right)^2\left(\sum_{t=1}^{T} t^2\right) - T\left(\sum_{t=1}^{T} tx_t\right)^2 - \left(\sum_{t=1}^{T} x_t^2\right)\left(\sum_{t=1}^{T} t\right)^2.$$

Therefore,

$$\begin{bmatrix} \dfrac{T}{\sigma_y}\left(\widehat{\alpha}_{(2)} - \alpha\right) \\ \dfrac{T^2}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right) \\ \dfrac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \end{bmatrix} = \frac{1}{\dfrac{1}{T^5\sigma_x^2}D} \begin{bmatrix} \dfrac{k_1}{T^4\sigma_x^2\sigma_y} \\ \dfrac{k_2}{T^3\sigma_x^2\sigma_y} \\ \dfrac{k_3}{T^4\sigma_x\sigma_y} \end{bmatrix} \Rightarrow \frac{1}{\Delta}\begin{bmatrix} \hbar_1 \\ \hbar_2 \\ \hbar_3 \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{2*} \\ \gamma_{2*} \\ \beta_{2*} \end{bmatrix},$$

where the weak convergence is based on items 1, 2, 4, 9, 10 and 11 of Lemma 1 of Chapter 2 and the CMT.

A.2. Proof of Theorem 2

To prove item 1.1, we see

$$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(0)}} = \frac{\dfrac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right)}{\left(T\sigma_x^2\, s_{\beta_{(0)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$T\sigma_x^2\, s_{\beta_{(0)}}^2 = \frac{s_0^2}{\dfrac{1}{T\sigma_x^2}\sum_{t=1}^{T} x_t^2},$$

where

$$s_0^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\beta}_{(0)}x_t\right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\beta}_{(0)}x_t &= \beta x_t + v_t - \widehat{\beta}_{(0)}x_t \\ &= v_t - \left(\widehat{\beta}_{(0)} - \beta\right) x_t. \end{aligned}$$

And from item 1 of Theorem 1, we have $\widehat{\beta}_{(0)} - \beta = O_p(T^{d_1-d_2-1})$. Moreover, we know from items 2, 5 and 9 of Lemma 1 of Chapter 2 that

$$\sum_{t=1}^{T} x_t^2 = O_p(T^{2+2d_2}), \quad \frac{1}{T}\sum_{t=1}^{T} v_t^2 \xrightarrow{p} \gamma_v(0), \quad \sum_{t=1}^{T} x_t v_t = O_p(T^{1+d_1+d_2}).$$

Therefore, we have

$$\begin{aligned} s_0^2 = &\ \frac{1}{T}\sum_{t=1}^{T} v_t^2 - \frac{1}{T}\left(\widehat{\beta}_{(0)} - \beta\right)^2 \sum_{t=1}^{T} x_t^2 \\ &= \frac{1}{T}\sum_{t=1}^{T} v_t^2 - O(T^{-1})\cdot O_p(T^{2d_1-2d_2-2})\cdot O_p(T^{2+2d_2}) \\ &= \frac{1}{T}\sum_{t=1}^{T} v_t^2 - O_p(T^{2d_1-1}) = \frac{1}{T}\sum_{t=1}^{T} v_t^2 + o_p(1) \xrightarrow{p} \gamma_v(0), \end{aligned}$$

because $2d_1 - 1 < 0$. Moreover, we have

$$\frac{1}{T\sigma_x^2}\sum_{t=1}^{T} x_t^2 \ \Rightarrow \ \int_0^1 [B_{0.5+d_2}(s)]^2\, ds,$$

where the weak convergence is based on item 2 of Lemma 1 of Chapter 2. So we prove

$$T\sigma_x^2\, s_{\beta_{(0)}}^2 \ \Rightarrow \ \frac{\gamma_v(0)}{\displaystyle\int_0^1 [B_{0.5+d_2}(s)]^2\, ds}.$$

For the numerator, we note

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \;\Rightarrow\; \beta_{0*}$$

by using item 1 of Theorem 1. Combining the asymptotic distributions of the numerator and denominator, item 1.1 is proved.

To prove item 1.2, we see

$$R^2 = \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t - \bar{x})^2}{\sum_{t=1}^{T} (C_t - \bar{C})}$$

$$= \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t - \bar{x})^2}{\beta^2 \sum_{t=1}^{T} (x_t - \bar{x})^2 + 2\beta \sum_{t=1}^{T} (x_t - \bar{x})(v_t - \bar{v}) + \sum_{t=1}^{T} (v_t - \bar{v})^2}.$$

From items 1, 3, 4, 6 and 9 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T}(x_t - \bar{x})^2 = O_p(T^{2+2d_2}), \qquad \sum_{t=1}^{T}(v_t - \bar{v})^2 = O_p(T),$$

and

$$\sum_{t=1}^{T}(x_t - \bar{x})(v_t - \bar{v}) = \sum_{t=1}^{T} x_t v_t - \frac{1}{T}\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} v_t\right) = O_p(T^{1+d_1+d_2}).$$

Consequently,

$$R^2 = \frac{\widehat{\beta}^2_{(0)}}{\beta^2 + O_p(T^{-1+d_1-d_2}) + O_p(T^{-1-2d_2})} = \frac{\left[\beta + O_p(T^{d_1-d_2-1})\right]^2}{\beta^2 + o_p(1)} \xrightarrow{p} 1,$$

where the convergence in probability is based on $\widehat{\beta}_{(0)} = \beta + o_p(1)$ and the second and the third terms in the denominator coverge in probability to zero because $-1 + d_1 - d_2 < 0$ and $-1 - 2d_2 < 0$. To prove item 1.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left(C_t - \widehat{\beta}_{(0)} x_t - C_{t-1} + \widehat{\beta}_{(0)} x_{t-1}\right)^2$$

$$= \left[v_t - v_{t-1} - \left(\widehat{\beta}_{(0)} - \beta\right) w_t\right]^2$$

$$= (v_t - v_{t-1})^2 - 2\left(\widehat{\beta}_{(0)} - \beta\right)(v_t - v_{t-1}) w_t$$

$$+ \left(\widehat{\beta}_{(0)} - \beta\right)^2 w_t^2.$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1}) w_t = \frac{1}{T}\sum_{t=2}^{T} v_t w_t - \frac{1}{T}\sum_{t=2}^{T} v_{t-1} w_t$$

$$= \max\left\{o_p(T^{\epsilon-1+d_1+d_2}), o_p(T^{\epsilon-0.5})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 1 of Theorem 1 that $\widehat{\beta}_{(0)} - \beta = O_p(T^{d_1-d_2-1})$ converges in probability to zero. From item 5 of Lemma 1 of Chapter 2, we also know that $\sum_{t=2}^{T} w_t^2/T \xrightarrow{p} \gamma_w(0)$ and that

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} v_t v_{t-1}$$

$$\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\gamma_v(1).$$

Finally,

$$DW \xrightarrow{p} 2 - 2\rho_v(1),$$

because the second and third terms in the numerator of DW all converge in probability to zero and the denominator $\sum_{t=1}^{T} \widehat{u}_t^2/T = s_0^2$ converges in probability to $\gamma_v(0)$ from the above results.

To prove item 2.1, we see

$$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(1)}} = \frac{\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right)}{\left(T\sigma_x^2 s_{\beta_{(1)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$T\sigma_x^2 s_{\beta_{(1)}}^2 = \frac{s_1^2}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \bar{x})^2},$$

where

$$s_1^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t\right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t &= \alpha + \beta x_t + v_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t \\ &= (v_t - \bar{v}) - \left(\widehat{\beta}_{(1)} - \beta\right)(x_t - \bar{x}). \end{aligned}$$

And from item 2 of Theorem 1, we have $\widehat{\beta}_{(1)} - \beta = O_p(T^{d_1-d_2-1})$.

Moreover, we know from items 1, 3, 4, 6 and 9 of Lemma 1 of Chapter 2 that

$$\sum_{t=1}^{T}(x_t - \bar{x})^2 = O_p(T^{2+2d_2}), \quad \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 \xrightarrow{p} \gamma_v(0),$$

and

$$\sum_{t=1}^{T}(x_t - \bar{x})(v_t - \bar{v}) = \sum_{t=1}^{T}x_t v_t - \frac{1}{T}\left(\sum_{t=1}^{T}x_t\right)\left(\sum_{t=1}^{T}v_t\right) = O_p(T^{1+d_1+d_2}).$$

Therefore, we have

$$
\begin{aligned}
s_1^2 = & \frac{1}{T}\sum_{t=1}^{T}(v_t-\bar{v})^2 - \frac{1}{T}\left(\widehat{\beta}_{(1)}-\beta\right)^2\sum_{t=1}^{T}(x_t-\bar{x})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(v_t-\bar{v})^2 - O(T^{-1})\cdot O_p(T^{2d_1-2d_2-2})\cdot O_p(T^{2+2d_2}) \\
&= \frac{1}{T}\sum_{t=1}^{T}(v_t-\bar{v})^2 - O_p(T^{2d_1-1}) = \frac{1}{T}\sum_{t=1}^{T}(v_t-\bar{v})^2 + o_p(1) \\
&\xrightarrow{p} \gamma_v(0),
\end{aligned}
$$

where the convergence in probability to $\gamma_v(0)$ is based on item 6 of Lemma 1 of Chapter 2. Moreover, we have

$$
\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t-\bar{x})^2 \Rightarrow \int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2,
$$

where the weak convergence is based on item 3 of Lemma 1 of Chapter 2. So we prove

$$
T\sigma_x^2\, s_{\beta_{(0)}}^2 \Rightarrow \frac{\gamma_v(0)}{\int_0^1 [B_{0.5+d_2}(s)]^2\,ds - \left[\int_0^1 B_{0.5+d_2}(s)\,ds\right]^2}.
$$

For the numerator, we note

$$
\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)}-\beta\right) \Rightarrow \beta_{1*}
$$

by using item 2 of Theorem 1. Combining the asymptotic distributions of the numerator and denominator, item 2.1 is proved.

To prove item 2.2, we see

$$R^2 = 1 - \frac{Ts_1^2}{\sum_{t=1}^{T}(C_t - \bar{C})^2}$$

$$= 1 - \frac{Ts_1^2}{\beta^2 \sum_{t=1}^{T}(x_t - \bar{x})^2 + 2\beta \sum_{t=1}^{T}(v_t - \bar{v})(x - \bar{x}) + \sum_{t=1}^{T}(v_t - \bar{v})^2},$$

given there is a constant included in the regression. From items 1, 3, 4, 6 and 9 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T}(x_t - \bar{x})^2 = O_p(T^{2+2d_2}), \quad \sum_{t=1}^{T}(v_t - \bar{v})^2 = O_p(T),$$

and

$$\sum_{t=1}^{T}(x_t - \bar{x})(v_t - \bar{v}) = O_p(T^{1+d_1+d_2}).$$

We also note $s_1^2 \xrightarrow{p} \gamma_v(0)$, consequently,

$$R^2 = 1 - \frac{O_p(T)}{O_p(T^{2+2d_2}) + O_p(T^{1+d_1+d_2}) + O_p(T)}$$

$$= 1 - \frac{O_p(1)}{O_p(T^{1+2d_2}) + O_p(T^{d_1+d_2}) + O_p(1)} = 1 + o_p(1) \xrightarrow{p} 1,$$

where the first term in the denominator of the second term diverge because $1 + 2d_2 > 0$.

To prove item 2.3, we first note

$$(\hat{u}_t - \hat{u}_{t-1})^2 = \left(C_t - \hat{\alpha}_{(1)} - \hat{\beta}_{(1)}x_t - C_{t-1} + \hat{\alpha}_{(1)} + \hat{\beta}_{(1)}x_{t-1}\right)^2$$

$$= \left[v_t - v_{t-1} - \left(\hat{\beta}_{(1)} - \beta\right) w_t\right]^2$$

$$= (v_t - v_{t-1})^2 - 2\left(\hat{\beta}_{(1)} - \beta\right)(v_t - v_{t-1})\, w_t$$

$$+ \left(\hat{\beta}_{(1)} - \beta\right)^2 w_t^2.$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})\,w_t = \frac{1}{T}\sum_{t=2}^{T} v_t w_t - \frac{1}{T}\sum_{t=2}^{T} v_{t-1} w_t$$

$$= \max\left\{o_p(T^{\epsilon-1+d_1+d_2}),\, o_p(T^{\epsilon-0.5})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 2 of Theorem 1 that $\widehat{\beta}_{(1)} - \beta = O_p(T^{d_1-d_2-1})$ converges in probability to zero. From item 5 of Lemma 1 of Chapter 2, we also know that $\sum_{t=2}^{T} w_t^2/T \xrightarrow{p} \gamma_w(0)$ and that

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T} v_t v_{t-1}$$

$$\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\gamma_v(1).$$

Finally,

$$DW \xrightarrow{p} 2 - 2\rho_v(1),$$

because the second and third terms in the numerator of DW all converge in probability to zero and the denominator $\sum_{t=1}^{T} \widehat{u}_t^2/T = s_1^2$ converges in probability to $\gamma_v(0)$ from the above results.

To prove item 3.1, we see

$$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(2)}} = \frac{\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right)}{\left(T\sigma_x^2\, s_{\beta_{(2)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$T\sigma_x^2\, s_{\beta_{(2)}}^2 = s_2^2 \frac{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2}{T^4} \frac{1}{\frac{1}{T^5\sigma_x^2} D},$$

where D is defined in the proof of item 3 of Theorem 1 and

$$s_2^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t\right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t &= \alpha + \gamma t + \beta x_t + v_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t \\ &= (v_t - \bar{v}) - \left(\widehat{\gamma}_{(2)} - \gamma\right)(t - \bar{t}) \\ &\quad - \left(\widehat{\beta}_{(2)} - \beta\right)(x_t - \bar{x}). \end{aligned}$$

And from item 3 of Theorem 3, we have

$$\widehat{\beta}_{(2)} - \beta = O_p(T^{d_1 - d_2 - 1}) \qquad \text{and} \qquad \widehat{\gamma}_{(2)} - \gamma = O_p(T^{d_1 - 1.5}).$$

From items 3 and 6 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T}(x_t - \bar{x})^2 = O_p(T^{2+2d_2}) \qquad \text{and} \qquad \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 \xrightarrow{p} \gamma_v(0).$$

From items 4 and 10 of Lemma 1 of Chapter 2, we have

$$\begin{aligned} \sum_{t=1}^{T}(v_t - \bar{v})(t - \bar{t}) &= \sum_{t=1}^{T} v_t t - \frac{1}{T}\left(\sum_{t=1}^{T} v_t\right)\left(\sum_{t=1}^{T} t\right) \\ &= O_p(T^{1.5+d_1}) - O_p(T^{-1})\cdot O_p(T^{0.5+d_1})\cdot O(T^2) \\ &= O_p(T^{1.5+d_1}). \end{aligned}$$

From items 1, 4, 9 and 11 of Lemma 1 of Chapter 2 and the same arguments above, we have

$$\sum_{t=1}^{T}(v_t - \bar{v})(x_t - \bar{x}) = O_p(T^{1+d_1+d_2}), \sum_{t=1}^{T}(t - \bar{t})(x_t - \bar{x}) = O_p(T^{d_2+2.5}).$$

Therefore, we have

$$s_2^2 = \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 + \frac{1}{T}\left(\widehat{\gamma}_{(2)} - \gamma\right)^2 \sum_{t=1}^{T}(t - \bar{t})^2$$

$$+ \frac{1}{T}\left(\widehat{\beta}_{(2)} - \beta\right)^2 \sum_{t=1}^{T}(x_t - \bar{x})^2 - 2\cdot\frac{1}{T}\left(\widehat{\gamma}_{(2)} - \gamma\right)\cdot\sum_{t=1}^{T}(v_t - \bar{v})(t - \bar{t})$$

$$- 2\cdot\frac{1}{T}\left(\widehat{\beta}_{(2)} - \beta\right)\sum_{t=1}^{T}(x_t - \bar{x})(v_t - \bar{v})$$

$$+ 2\cdot\frac{1}{T}\left(\widehat{\gamma}_{(2)} - \gamma\right)\left(\widehat{\beta}_{(2)} - \beta\right)\sum_{t=1}^{T}(x_t - \bar{x})(t - \bar{t})$$

$$= \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 + O(T^{-1})\cdot O_p(T^{2d_1-3})\cdot O(T^3)$$

$$+ O(T^{-1})\cdot O_p(T^{2d_1-2d_2-2})\cdot O_p(T^{2+2d_2})$$

$$- O(T^{-1})\cdot O_p(T^{d_1-1.5})\cdot O_p(T^{1.5+d_1})$$

$$- O(T^{-1})\cdot O_p(T^{d_1-d_2-1})\cdot O_p(T^{1+d_1+d_2})$$

$$+ O(T^{-1})\cdot O_p(T^{d_1-1.5})\cdot O_p(T^{d_1-d_2-1})\cdot O_p(T^{2.5+d_2})$$

$$= \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 + O_p(T^{2d_1-1}) + O_p(T^{2d_1-1})$$

$$- O_p(T^{2d_1-1}) - O_p(T^{2d_1-1}) + O_p(T^{2d_1-1}) \xrightarrow{p} \gamma_v(0),$$

where convergence in probability is based on item 6 of Lemma 1 of Chapter 2. Therefore, we prove

$$T\sigma_x^2 s_{\beta_{(2)}}^2 \Rightarrow \frac{\gamma_v(0)}{12\Delta},$$

where the weak convergence is based on item 3 of Theorem 1 and the above results. For the numerator, we note

$$\frac{T\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)}-\beta\right) \Rightarrow \beta_{2*}$$

by using item 3 of Theorem 3. Combining the asymptotic distributions of the numerator and denominator, item 3.1 is proved.

To prove item 3.2, we see

$$R^2 = 1 - \frac{Ts_2^2}{\sum_{t=1}^{T}\left(C_t-\bar{C}\right)^2} = 1 - \frac{Ts_2^2}{\sum_{t=1}^{T}\left[\beta\left(x_t-\bar{x}\right)+\gamma\left(t-\bar{t}\right)+\left(v_t-\bar{v}\right)\right]^2},$$

given there is a constant included in the regression. The order of the denominator of the second term is $O(T^3)$ since the term $\sum_{t=1}^{T}\left(t-\bar{t}\right)^2$ has higher order than the other terms and $\sum_{t=1}^{T}\left(t-\bar{t}\right)^2 = O(T^3)$. We also note $s_2^2 \xrightarrow{p} \gamma_v(0)$, consequently,

$$R^2 = 1 - \frac{O_p(T)}{O(T^3)} = 1 - O_p(T^{-2}) \xrightarrow{p} 1.$$

To prove item 3.3, we first note

$$\begin{aligned}
\left(\widehat{u}_t-\widehat{u}_{t-1}\right)^2 &= \left[\left(v_t-v_{t-1}\right)-\left(\widehat{\beta}_{(2)}-\beta\right)w_t-\left(\widehat{\gamma}_{(2)}-\gamma\right)\right]^2 \\
&= \left(v_t-v_{t-1}\right)^2+\left(\widehat{\beta}_{(2)}-\beta\right)^2 w_t^2+\left(\widehat{\gamma}_{(2)}-\gamma\right)^2 \\
&\quad -2\left(\widehat{\beta}_{(2)}-\beta\right)\left(v_t-v_{t-1}\right)w_t \\
&\quad -2\left(\widehat{\gamma}_{(2)}-\gamma\right)\left(v_t-v_{t-1}\right)+2\left(\widehat{\beta}_{(2)}-\beta\right)\left(\widehat{\gamma}-\gamma\right)w_t.
\end{aligned}$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\sum_{t=2}^{T}\left(v_t-v_{t-1}\right)w_t = \max\left\{O_p(T^{d_1+d_2}), O_p(T^{0.5})\right\}.$$

Moreover, we know from item 3 of Theorem 3 that

$$\widehat{\beta}_{(2)} - \beta = O_p(T^{d_1-d_2-1}) \qquad \text{and} \qquad \widehat{\gamma}_{(2)} - \gamma = O_p(T^{d_1-1.5}).$$

From item 5 of Lemma 1 of Chapter 2, we know that

$$\sum_{t=2}^{T} v_t^2 = O_p(T) \qquad \text{and} \qquad \sum_{t=2}^{T} w_t^2 = O_p(T).$$

From item 4 of Lemma 1 of Chapter 2, we know that

$$\sum_{t=2}^{T} v_t = O_p(T^{0.5+d_1}) \quad \text{and} \qquad \sum_{t=2}^{T} w_t = O_p(T^{0.5+d_2}).$$

All these results imply

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}(\widehat{u}_t - \widehat{u}_{t-1})^2 = & \frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 + O(T^{-1})O_p(T^{2d_1-2d_2-2})O_p(T) \\
& + O(T^{-1})O_p(T^{2d_1-3})\cdot O(T) \\
& - O(T^{-1})\cdot O_p(T^{d_1-d_2-1})\cdot \max\left\{O_p(T^{d_1+d_2}), O_p(T^{0.5})\right\} \\
& - O(T^{-1})\cdot O_p(T^{d_1-1.5})\cdot O_p(T^{0.5+d_1}) \\
& + O(T^{-1})O_p(T^{d_1-d_2-1})\cdot O_p(T^{d_1-1.5})\cdot O_p(T^{0.5+d_2}) \\
& = \frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 + O_p(T^{2d_1-2d_2-2}) + O_p(T^{2d_1-4}) \\
& - \max\left\{O_p(T^{2d_1-2}), O_p(T^{d_1-d_2-1.5})\right\} - O_p(T^{2d_1-2}) \\
& + O_p(T^{2d_1-3}) \\
& = \frac{1}{T}\sum_{t=2}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_{t-1}^2 - 2\sum_{t=2}^{T} v_t v_{t-1} + o_p(1) \\
& \xrightarrow{p} 2\gamma_v(0) - 2\gamma_v(1).
\end{aligned}$$

Therefore,

$$DW \xrightarrow{p} 2 - 2\rho_v(1),$$

because $\sum_{t=1}^{T} \widehat{u}_t^2/T = s_2^2 \xrightarrow{p} \gamma_v(0)$.

A.3. Proof of Theorem 3

To prove item 1, we see

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) = \frac{\dfrac{1}{T\sigma_y\sigma_x}\displaystyle\sum_{t=1}^{T} x_t y_t}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T} x_t^2} \Rightarrow \beta_{0*},$$

where the weak convergence is based on items 2 and 8 of Lemma 1 of Chapter 2 and the CMT.

To prove item 2, we see

$$\begin{bmatrix} \widehat{\alpha}_{(1)} - \alpha \\ \widehat{\beta}_{(1)} - \beta \end{bmatrix} = \begin{bmatrix} T & \displaystyle\sum_{t=1}^{T} x_t \\ \displaystyle\sum_{t=1}^{T} x_t & \displaystyle\sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \displaystyle\sum_{t=1}^{T} y_t \\ \displaystyle\sum_{t=1}^{T} x_t y_t \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\left(\displaystyle\sum_{t=1}^{T} x_t^2\right)\left(\displaystyle\sum_{t=1}^{T} y_t\right) - \left(\displaystyle\sum_{t=1}^{T} x_t\right)\left(\displaystyle\sum_{t=1}^{T} x_t y_t\right)}{T\displaystyle\sum_{t=1}^{T}(x_t - \bar{x})^2} \\ \dfrac{-\left(\displaystyle\sum_{t=1}^{T} x_t\right)\left(\displaystyle\sum_{t=1}^{T} y_t\right) + T\left(\displaystyle\sum_{t=1}^{T} x_t y_t\right)}{T\displaystyle\sum_{t=1}^{T}(x_t - \bar{x})^2} \end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} \frac{1}{\sigma_y}\left(\widehat{\alpha}_{(1)}-\alpha\right) \\ \frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)}-\beta\right) \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\frac{1}{T^2\sigma_y\sigma_x^2}\left[\left(\sum_{t=1}^{T} x_t^2\right)\left(\sum_{t=1}^{T} y_t\right) - \left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} y_t x_t\right)\right]}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t-\bar{x})^2} \\ \dfrac{\frac{1}{T^2\sigma_y\sigma_x}\left[-\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} y_t\right) + T\left(\sum_{t=1}^{T} x_t y_t\right)\right]}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t-\bar{x})^2} \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{1*} \\ \beta_{1*} \end{bmatrix},$$

where the weak convergence is based on items 1, 2 and 8 of Lemma 1 of Chapter 2 and the CMT.

To prove item 3, we see

$$\begin{bmatrix} \widehat{\alpha}_{(2)}-\alpha \\ \widehat{\gamma}_{(2)}-\gamma \\ \widehat{\beta}_{(2)}-\beta \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} t y_t \\ \sum_{t=1}^{T} x_t y_t \end{bmatrix} = \frac{1}{D}\begin{bmatrix} k_1' \\ k_2' \\ k_3' \end{bmatrix},$$

where

$$k'_1 = \left(\sum_{t=1}^{T} y_t\right)\left\{\left(\sum_{t=1}^{T} t^2\right)\left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=1}^{T} tx_t\right)^2\right\}$$

$$+ \left(\sum_{t=1}^{T} ty_t\right)\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} tx_t\right) - \left(\sum_{t=1}^{T} ty_t\right)\left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} x_t^2\right)$$

$$+ \left(\sum_{t=1}^{T} x_t y_t\right)\left\{-\left(\sum_{t=1}^{T} t^2\right)\left(\sum_{t=1}^{T} x_t\right) + \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} tx_t\right)\right\},$$

$$k'_2 = \left(\sum_{t=1}^{T} y_t\right)\left\{\left(\sum_{t=1}^{T} tx_t\right)\left(\sum_{t=1}^{T} x_t\right) - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} x_t^2\right)\right\}$$

$$+ \left(\sum_{t=1}^{T} ty_t\right)\cdot T\left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=1}^{T} ty_t\right)\left(\sum_{t=1}^{T} x_t\right)^2$$

$$+ \left(\sum_{t=1}^{T} x_t y_t\right)\left\{\left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} x_t\right) - T\left(\sum_{t=1}^{T} tx_t\right)\right\},$$

$$k'_3 = \left(\sum_{t=1}^{T} y_t\right)\left\{-\left(\sum_{t=1}^{T} t^2\right)\left(\sum_{t=1}^{T} x_t\right) + \left(\sum_{t=1}^{T} tx_t\right)\left(\sum_{t=1}^{T} t\right)\right\}$$

$$+ \left(\sum_{t=1}^{T} ty_t\right)\left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} x_t\right) - T\left(\sum_{t=1}^{T} tx_t\right)\left(\sum_{t=1}^{T} ty_t\right)$$

$$+ \left(\sum_{t=1}^{T} x_t y_t\right)\left\{T\left(\sum_{t=1}^{T} t^2\right) - \left(\sum_{t=1}^{T} t\right)^2\right\}.$$

Therefore,

$$\begin{bmatrix} \frac{1}{\sigma_y}\left(\widehat{\alpha}_{(2)}-\alpha\right) \\ \frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)}-\gamma\right) \\ \frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)}-\beta\right) \end{bmatrix} = \frac{1}{\frac{1}{T^5\sigma_x^2}D} \begin{bmatrix} \frac{k_1'}{T^5\sigma_x^2\sigma_y} \\ \frac{k_2'}{T^4\sigma_x^2\sigma_y} \\ \frac{k_3'}{T^5\sigma_x\sigma_y} \end{bmatrix} \Rightarrow \frac{1}{\Delta}\begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{2*} \\ \gamma_{2*} \\ \beta_{2*} \end{bmatrix},$$

where the weak convergence is based on items 1, 2, 8 and 11 of Lemma 1 of Chapter 2 and the CMT.

A.4. Proof of Theorem 4

To prove item 1.1, we see

$$\frac{1}{\sqrt{T}}t_{\beta_{(0)}} = \frac{\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)}-\beta\right)}{\left(\frac{T\sigma_x^2}{\sigma_y^2}s^2_{\beta_{(0)}}\right)^{1/2}}.$$

For the denominator, we note

$$\frac{T\sigma_x^2}{\sigma_y^2}s^2_{\beta_{(0)}} = \frac{s_0^2}{\sigma_y^2}\frac{1}{\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}x_t^2},$$

where

$$s_0^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t-\widehat{\beta}_{(0)}x_t\right)^2,$$

and

$$\begin{aligned} C_t-\widehat{\beta}_{(0)}x_t &= \beta x_t + y_t - \widehat{\beta}_{(0)}x_t \\ &= y_t - \left(\widehat{\beta}_{(0)}-\beta\right)x_t. \end{aligned}$$

Therefore, we have

$$\frac{s_0^2}{\sigma_y^2} = \frac{1}{T\sigma_y^2}\sum_{t=1}^{T} y_t^2 - \left[\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right)\right]^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T} x_t^2$$

$$\Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\, ds - \frac{\left[\int_0^1 B_{0.5+d_1}(s)\, B_{0.5+d_2}(s)\, ds\right]^2}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds}$$

$$\Rightarrow \sigma_{0*}^2,$$

where the weak convergence is based on item 1 of Theorem 3 and items 2 and 8 of Lemma 1 of Chapter 2. Moreover, we have

$$\frac{1}{T\sigma_x^2}\sum_{t=1}^{T} x_t^2 \Rightarrow \int_0^1 [B_{0.5+d_2}(s)]^2\, ds,$$

where the weak convergence is based on item 2 of Lemma 1 of Chapter 2. So we prove

$$\frac{T\sigma_x^2}{\sigma_y^2} s_{\beta_{(0)}}^2 \Rightarrow \frac{\int_0^1 [B_{0.5+d_1}(s)]^2\, ds}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds} - \left\{\frac{\int_0^1 B_{0.5+d_1}(s)\, B_{0.5+d_2}(s)\, ds}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds}\right\}^2.$$

For the numerator, we note

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \Rightarrow \beta_{0*}$$

by using item 1 of Theorem 3. Combining the asymptotic distributions of the numerator and denominator, item 1.1 is proved.

To prove item 1.2, we see

$$R^2 = \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t - \bar{x})^2}{\sum_{t=1}^{T} (C_t - \bar{C})}$$

$$= \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t - \bar{x})^2}{\beta^2 \sum_{t=1}^{T} (x_t - \bar{x})^2 + 2\beta \sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y}) + \sum_{t=1}^{T} (y_t - \bar{y})^2}.$$

From items 1, 3 and 8 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T} (x_t - \bar{x})^2 = O_p(T^{2+2d_2}), \quad \sum_{t=1}^{T} (y_t - \bar{y})^2 = O_p(T^{2+2d_1}),$$

and

$$\sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y}) = O_p(T^{2+d_1+d_2}).$$

Consequently,

$$R^2 = \frac{\widehat{\beta}^2_{(0)}}{\beta^2 + O_p(T^{d_1-d_2}) + O_p(T^{2d_1-2d_2})} = \frac{\left[\beta + O_p(T^{d_1-d_2})\right]^2}{\beta^2 + o_p(1)} \xrightarrow{p} 1,$$

where the convergence in probability is based on $\widehat{\beta}_{(0)} = \beta + o_p(1)$ and the second and the third terms in the denominator coverge in probability to zero because $d_1 - d_2 < 0$.

To prove item 1.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left(C_t - \widehat{\beta}_{(0)} x_t - C_{t-1} + \widehat{\beta}_{(0)} x_{t-1}\right)^2$$

$$= v_t^2 - 2\left(\widehat{\beta}_{(0)} - \beta\right) v_t w_t + \left(\widehat{\beta}_{(0)} - \beta\right)^2 w_t^2.$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_t = \max\left\{o_p(T^{\epsilon-1.5+d_2}), o_p(T^{\epsilon-1-d_1})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 1 of Theorem 3 that $\left(\widehat{\beta}_{(0)} - \beta\right)/\sigma_y = O_p(T^{-d_2-0.5})$ converges in probability to zero. From item 5 of Lemma 1 of Chapter 2, we also know that $\sum_{t=2}^{T} v_t^2/T \xrightarrow{p} \gamma_v(0)$ and that $\sum_{t=2}^{T} w_t^2/T \xrightarrow{p} \gamma_w(0)$. All these results imply DW statistic is

$$\frac{\dfrac{1}{\sigma_y^2}\cdot\dfrac{1}{T}\displaystyle\sum_{t=2}^{T} v_t^2 - 2\frac{\widehat{\beta}_{(0)} - \beta}{\sigma_y}\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_t + \left(\frac{\widehat{\beta}_{(0)} - \beta}{\sigma_y}\right)^2 \frac{1}{T}\sum_{t=2}^{T} w_t^2}{\dfrac{1}{T\sigma_y^2}\displaystyle\sum_{t=1}^{T}\widehat{u}_t^2} \xrightarrow{p} 0,$$

where the three terms in the numerator all converge in probability to zero while the denominator $\sum_{t=1}^{T}\widehat{u}_t^2/T\sigma_y^2 = s_0^2/\sigma_y^2 \Rightarrow \sigma_{0*}^2$ from the above results.

To prove item 2.1, we see

$$\frac{1}{\sqrt{T}} t_{\beta_{(1)}} = \frac{\dfrac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right)}{\left(\dfrac{T\sigma_x^2}{\sigma_y^2} s_{\beta_{(1)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$\frac{T\sigma_x^2}{\sigma_y^2} s_{\beta_{(1)}}^2 = \frac{s_1^2}{\sigma_y^2}\frac{1}{\dfrac{1}{T\sigma_x^2}\displaystyle\sum_{t=1}^{T}(x_t - \bar{x})^2},$$

where

$$s_1^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t\right)^2,$$

and

$$C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t = \alpha + \beta x_t + y_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t$$
$$= (y_t - \bar{y}) - \left(\widehat{\beta}_{(1)} - \beta\right)(x_t - \bar{x}).$$

Therefore, we have

$$\frac{s_1^2}{\sigma_y^2} = \frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \bar{y})^2 - \left[\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right)\right]^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \bar{x})^2$$
$$\Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_1}(s)\, ds\right]^2$$
$$- \beta_{1*}^2 \left\{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2\right\}$$
$$\Rightarrow \sigma_{1*}^2,$$

where the weak convergence is based on item 2 of Theorem 3 and items 1, 3 and 8 of Lemma 1 of Chapter 2. Moreover, we have

$$\frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \bar{x})^2 \Rightarrow \int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2,$$

where the weak convergence is based on item 3 of Lemma 1 of Chapter 2. So we prove

$$\frac{T\sigma_x^2}{\sigma_y^2} s_{\beta_{(1)}}^2 \Rightarrow \frac{\sigma_{1*}^2}{\int_0^1 [B_{0.5+d_2}(s)]^2\, ds - \left[\int_0^1 B_{0.5+d_2}(s)\, ds\right]^2}.$$

For the numerator, we note

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \Rightarrow \beta_{1*}$$

by using item 2 of Theorem 3. Combining the asymptotic distributions of the numerator and denominator, item 2.1 is proved.

To prove item 2.2, we see

$$R^2 = 1 - \frac{Ts_1^2}{\sum_{t=1}^{T} (C_t - \bar{C})^2}$$

$$= 1 - \frac{Ts_1^2}{\beta^2 \sum_{t=1}^{T} (x_t - \bar{x})^2 + 2\beta \sum_{t=1}^{T} (y_t - y_t)(x - \bar{x}) + \sum_{t=1}^{T} (y_t - \bar{y})^2},$$

given there is a constant included in the regression. From items 1, 3 and 8 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T} (x_t - \bar{x})^2 = O_p(T^{2+2d_2}), \quad \sum_{t=1}^{T} (y_t - \bar{y})^2 = O_p(T^{2+2d_1}),$$

and

$$\sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y}) = O_p(T^{2+d_1+d_2}).$$

We also note $s_1^2/\sigma_y^2 \Rightarrow \sigma_{1*}^2$, consequently,

$$R^2 = 1 - \frac{O_p(T^{2+2d_1})}{O_p(T^{2+2d_2}) + O_p(T^{2+d_1+d_2}) + O_p(T^{2+2d_1})}$$

$$= 1 - \frac{O_p(1)}{O_p(T^{2d_2-2d_1}) + O_p(T^{d_2-d_1}) + O_p(1)}$$

$$= 1 + o_p(1) \xrightarrow{p} 1,$$

where the first and the second terms in the denominator of the second term diverge because $d_2 - d_1 > 0$.

To prove item 2.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left(C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t - C_{t-1} + \widehat{\alpha}_{(1)} + \widehat{\beta}_{(1)} x_{t-1}\right)^2$$

$$= v_t^2 - 2\left(\widehat{\beta}_{(1)} - \beta\right) v_t w_t + \left(\widehat{\beta}_{(1)} - \beta\right)^2 w_t^2.$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_t = \max\left\{o_p(T^{\epsilon-1.5+d_2}), o_p(T^{\epsilon-1-d_1})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 2 of Theorem 3 that $\left(\widehat{\beta}_{(1)} - \beta\right)/\sigma_y = O_p(T^{-d_2-0.5})$ converges in probability to zero. From item 5 of Lemma 1 of Chapter 2, we also know that $\sum_{t=2}^{T} v_t^2/T \xrightarrow{p} \gamma_v(0)$ and that $\sum_{t=2}^{T} w_t^2/T \xrightarrow{p} \gamma_w(0)$. All these results imply DW statistic is

$$\frac{\dfrac{1}{\sigma_y^2}\cdot\dfrac{1}{T}\displaystyle\sum_{t=2}^{T} v_t^2 - 2\frac{\widehat{\beta}_{(1)}-\beta}{\sigma_y}\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t w_t + \left(\frac{\widehat{\beta}_{(1)}-\beta}{\sigma_y}\right)^2\frac{1}{T}\sum_{t=2}^{T} w_t^2}{\dfrac{1}{T\sigma_y^2}\displaystyle\sum_{t=1}^{T}\widehat{u}_t^2} \xrightarrow{p} 0,$$

where the three terms in the numerator all converge in probability to zero while the denominator $\sum_{t=1}^{T}\widehat{u}_t^2/T\sigma_y^2 = s_1^2/\sigma_y^2 \Rightarrow \sigma_{1*}^2$ from the above results.

To prove item 3.1, we see

$$\frac{1}{\sqrt{T}}t_{\beta_{(2)}} = \frac{\dfrac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right)}{\left(\dfrac{T\sigma_x^2}{\sigma_y^2}s_{\beta_{(2)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$\frac{T\sigma_x^2}{\sigma_y^2}s_{\beta_{(2)}}^2 = \frac{s_2^2}{\sigma_y^2}\frac{T\displaystyle\sum_{t=1}^{T}t^2 - \left(\sum_{t=1}^{T}t\right)^2}{T^4}\frac{1}{\dfrac{1}{T^5\sigma_x^2}D},$$

where D is defined in the proof of item 3 of Theorem 1 and

$$s_2^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t\right)^2,$$

and

$$C_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t = \alpha + \gamma t + \beta x_t + y_t - \widehat{\alpha}_{(2)} - \widehat{\gamma}_{(2)}t - \widehat{\beta}_{(2)}x_t$$
$$= (y_t - \bar{y}) - \left(\widehat{\gamma}_{(2)} - \gamma\right)(t - \bar{t}) - \left(\widehat{\beta}_{(2)} - \beta\right)(x_t - \bar{x}).$$

Therefore, we have

$$\frac{s_2^2}{\sigma_y^2} = \frac{1}{T\sigma_y^2}\sum_{t=1}^{T}(y_t - \bar{y})^2 + \left[\frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right)\right]^2 \frac{1}{T^3}\sum_{t=1}^{T}(t - \bar{t})^2$$
$$+ \left[\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right)\right]^2 \frac{1}{T\sigma_x^2}\sum_{t=1}^{T}(x_t - \bar{x})^2$$
$$- 2\cdot\frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right)\cdot\frac{1}{T^2\sigma_y}\sum_{t=1}^{T}(y_t - \bar{y})(t - \bar{t})$$
$$- 2\cdot\left[\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right)\right]\frac{1}{T\sigma_x\sigma_y}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})$$
$$+ 2\cdot\left[\frac{T}{\sigma_y}\left(\widehat{\gamma}_{(2)} - \gamma\right)\right]\left[\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right)\right]\frac{1}{T^2\sigma_x}\sum_{t=1}^{T}(x_t - \bar{x})(t - \bar{t})$$
$$\Rightarrow \sigma_{2*}^2,$$

where the weak convergence is based on item 3 of Theorem 3 above and items 1, 3, 4, 8 and 11 of Lemma 1 of Chapter 2. Therefore, we prove

$$\frac{T\sigma_x^2}{\sigma_y^2} s_{\beta(2)}^2 \Rightarrow \frac{\sigma_{2*}^2}{12\Delta},$$

where the weak convergence is based on item 3 of Theorem 1 and the above results. For the numerator, we note

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(2)} - \beta\right) \Rightarrow \beta_{2*}$$

by using item 3 of Theorem 3. Combining the asymptotic distributions of the numerator and denominator, item 3.1 is proved.

To prove item 3.2, we see

$$R^2 = 1 - \frac{Ts_2^2}{\sum_{t=1}^{T}(C_t - \bar{C})^2} = 1 - \frac{Ts_2^2}{\sum_{t=1}^{T}[\beta(x_t - \bar{x}) + \gamma(t - \bar{t}) + (y_t - \bar{y})]^2},$$

given there is a constant included in the regression. The order of the denominator of the second term is $O(T^3)$ since the term $\sum_{t=1}^{T}(t-\bar{t})^2$ has higher order than the other terms and $\sum_{t=1}^{T}(t-\bar{t})^2 = O(T^3)$. We also note $s_2^2/\sigma_y^2 \Rightarrow \sigma_{2*}^2$, consequently,

$$R^2 = 1 - \frac{O_p(T^{2+2d_1})}{O(T^3)} = 1 - O_p(T^{2d_1-1}) \xrightarrow{p} 1,$$

because $2d_1 - 1 < 0$.

To prove item 3.3, we first note

$$\begin{aligned}(\widehat{u}_t - \widehat{u}_{t-1})^2 = & \left[v_t - \left(\widehat{\beta}_{(2)} - \beta\right) w_t - \left(\widehat{\gamma}_{(2)} - \gamma\right)\right]^2 \\ = & \ v_t^2 + \left(\widehat{\beta}_{(2)} - \beta\right)^2 w_t^2 + \left(\widehat{\gamma}_{(2)} - \gamma\right)^2 - 2\left(\widehat{\beta}_{(2)} - \beta\right) v_t w_t \\ & - 2\left(\widehat{\gamma}_{(2)} - \gamma\right) v_t + 2\left(\widehat{\beta}_{(2)} - \beta\right)\left(\widehat{\gamma} - \gamma\right) w_t.\end{aligned}$$

And from item 7 of Lemma 1 of Chapter 2, we have

$$\sum_{t=2}^{T} v_t w_t = \max\left\{O_p(T^{d_1+d_2}), O_p(T^{0.5})\right\}.$$

Moreover, we know from item 3 of Theorem 3 that

$$\widehat{\beta}_{(2)} - \beta = O_p(T^{d_1-d_2}) \qquad \text{and} \qquad \widehat{\gamma}_{(2)} - \gamma = O_p(T^{d_1-0.5}).$$

From item 5 of Lemma 1 of Chapter 2, we know that

$$\sum_{t=2}^{T} v_t^2 = O_p(T) \qquad \text{and} \qquad \sum_{t=2}^{T} w_t^2 = O_p(T).$$

From item 4 of Lemma 1 of Chapter 2, we know that

$$\sum_{t=2}^{T} v_t = O_p(T^{0.5+d_1}) \quad \text{and} \qquad \sum_{t=2}^{T} w_t = O_p(T^{0.5+d_2}).$$

All these results imply

$$\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2$$

$$= O_p(T) + O_p(T^{2d_1-2d_2}) \cdot O_p(T) + O_p(T^{2d_1-1}) \cdot O(T)$$

$$- O_p(T^{d_1-d_2}) \cdot \max\left\{O_p(T^{d_1+d_2}), O_p(T^{0.5})\right\} - O_p(T^{d_1-0.5}) \cdot O_p(T^{0.5+d_1})$$

$$+ O_p(T^{d_1-d_2}) \cdot O_p(T^{d_1-0.5}) \cdot O_p(T^{0.5+d_2})$$

$$= O_p(T) + O_p(T^{2d_1-2d_2+1}) + O_p(T^{2d_1})$$

$$- \max\left\{O_p(T^{2d_1}), O_p(T^{d_1-d_2+0.5})\right\} - O_p(T^{2d_1}) + O_p(T^{2d_1}).$$

We also know $\sum_{t=1}^{T} \widehat{u}_t^2 = Ts_2^2 = O_p(T^{2+2d_1})$. Therefore, we note

$$\frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2} = O_p(T^{-1-2d_1}) + O_p(T^{-1-2d_2}) + O_p(T^{-2})$$

$$- \max\left\{O_p(T^{-2}), O_p(T^{-1.5-d_1-d_2})\right\}$$

$$- O_p(T^{-2}) + O_p(T^{-2}).$$

Consequently, $DW \xrightarrow{p} 0$ because $-1-2d_1<0, -1-2d_2<0$ and $-1.5-d_1-d_2<0$.

To prove Theorem 5, let us first present the following lemma which will shorten our presentation considerably.

Lemma A.1. *Given that Assumption 1 of Chapter 2 holds, then, as $T\to\infty$, we have the following results:*

1. $$\frac{1}{T\sigma_y}\sum_{t=1}^{T} x_t^o v_t \;\Rightarrow\; B_{0.5+d_1}(1) - \int_0^1 B_{0.5+d_1}(s)\,ds.$$

2. $$\frac{1}{T^2\sigma_y}\sum_{t=1}^{T} x_t^o y_t \;\Rightarrow\; \int_0^1 s\cdot B_{0.5+d_1}(s)\,ds.$$

3. $$\frac{1}{T^2}\sum_{t=1}^{T} x_t^o \xrightarrow{p} \frac{1}{2}, \qquad \frac{1}{T^3}\sum_{t=1}^{T} x_t^{o2} \xrightarrow{p} \frac{1}{3},$$

$$\frac{1}{T^3}\sum_{t=1}^{T} (x_t^o-\bar{x}^o)^2 \xrightarrow{p} \frac{1}{12}$$

and $$\frac{1}{T^3}\sum_{t=1}^{T} x_t^o t \xrightarrow{p} \frac{1}{3}.$$

A.5. Proof of Lemma A.1

To prove item 1, we see

$$\frac{1}{T\sigma_y}\sum_{t=1}^{T} x_t^o v_t = \frac{1}{T\sigma_y}\sum_{t=1}^{T}(t + x_t)\, v_t = \frac{1}{T\sigma_y}\sum_{t=1}^{T} t v_t + \frac{1}{T\sigma_y}\sum_{t=1}^{T} x_t v_t$$

$$= \frac{1}{T\sigma_y}\sum_{t=1}^{T} t v_t + o_p(1)$$

$$\Rightarrow \quad B_{0.5+d_1}(1) - \int_0^1 B_{0.5+d_1}(s)\, ds,$$

where the weak convergence is based on item 10 of Lemma 1 of Chapter 2 and the fact that $T\sigma_y = O(T^{1.5+d_1})$ and $\sum_{t=1}^{T} x_t v_t = O_p(T^{1+d_1+d_2})$ by using item 9 of Lemma 1 of Chapter 2. The remaining items in Lemma A.1 can be proved by the same arguments.

A.6. Proof of Theorem 5

To prove item 1, we see

$$\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) = \frac{\frac{1}{T\sigma_y}\sum_{t=1}^{T} x_t^o v_t}{\frac{1}{T^3}\sum_{t=1}^{T} x_t^{o2}} \quad \Rightarrow \quad \beta_{0*},$$

where the weak convergence is based on items 1 and 3 of Lemma A.1 and the CMT.

To prove item 2, we see

$$\begin{bmatrix} \widehat{\alpha}_{(1)} - \alpha \\ \widehat{\beta}_{(1)} - \beta \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} x_t^o \\ \sum_{t=1}^{T} x_t^o & \sum_{t=1}^{T} x_t^{o2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} v_t \\ \sum_{t=1}^{T} x_t^o v_t \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\left(\sum_{t=1}^{T} x_t^{o2}\right)\left(\sum_{t=1}^{T} v_t\right) - \left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} x_t^o v_t\right)}{T\sum_{t=1}^{T}\left(x_t^o - \bar{x}^o\right)^2} \\ \dfrac{-\left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} v_t\right) + T\left(\sum_{t=1}^{T} x_t^o v_t\right)}{T\sum_{t=1}^{T}\left(x_t^o - \bar{x}^o\right)^2} \end{bmatrix} .$$

Therefore,

$$\begin{bmatrix} \dfrac{T}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \\ \dfrac{T^2}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\dfrac{1}{T^3\sigma_y}\left[\left(\sum_{t=1}^{T} x_t^{o2}\right)\left(\sum_{t=1}^{T} v_t\right) - \left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} v_t x_t^o\right)\right]}{\dfrac{1}{T^3}\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2} \\ \dfrac{\dfrac{1}{T^2\sigma_y}\left[-\left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} v_t\right) + T\left(\sum_{t=1}^{T} x_t^o v_t\right)\right]}{\dfrac{1}{T^3}\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2} \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{1*} \\ \beta_{1*} \end{bmatrix},$$

where the weak convergence is based on items 1 and 3 of Lemma A.1 and item 4 of Lemma 1 of Chapter 2 and the CMT.

To prove item 3, we see that the introduction of a nonzero drift in the trending variable when a time trend is included in the regression causes the covariance matrix to be singular. Therefore, we transform $C_t = \alpha + \gamma t + \beta x_t^o + v_t$ to be $C_t = \alpha + (\gamma + \beta)t + \beta x_t + v_t$. The remaining proof can follow the same arguments in the proof of item 3 of Theorem 1.

A.7. Proof of Theorem 6

To prove item 1.1, we see

$$\frac{\sqrt{T}}{\sigma_y} t_{\beta_{(0)}} = \frac{\dfrac{T^2}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right)}{\left(T^3\, s^2_{\beta_{(0)}}\right)^{1/2}}.$$

For the denominator, we note

$$T^3\, s^2_{\beta_{(0)}} = \frac{s_0^2}{\dfrac{1}{T^3}\sum_{t=1}^{T} x_t^{o2}},$$

where

$$s_0^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\beta}_{(0)}x_t^o\right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\beta}_{(0)}x_t^o &= \beta x_t^o + v_t - \widehat{\beta}_{(0)}x_t^o \\ &= v_t - \left(\widehat{\beta}_{(0)} - \beta\right)x_t^o. \end{aligned}$$

And from item 1 of Theorem 5, we have $\widehat{\beta}_{(0)} - \beta = O_p(T^{d_1-1.5})$. Moreover, we know from items 1 and 3 of Lemma A.1 and item 5 of Lemma 1 of Chapter 2 that

$$\sum_{t=1}^{T}x_t^{o2} = O(T^3), \qquad \frac{1}{T}\sum_{t=1}^{T}v_t^2 \xrightarrow{p} \gamma_v(0), \qquad \sum_{t=1}^{T}x_t^o v_t = O_p(T^{1.5+d_1}).$$

Therefore, we have

$$\begin{aligned} s_0^2 = & \frac{1}{T}\sum_{t=1}^{T}v_t^2 - \frac{1}{T}\left(\widehat{\beta}_{(0)} - \beta\right)^2\sum_{t=1}^{T}x_t^{o2} \\ = & \frac{1}{T}\sum_{t=1}^{T}v_t^2 - O(T^{-1})\cdot O_p(T^{2d_1-3})\cdot O(T^3) \\ = & \frac{1}{T}\sum_{t=1}^{T}v_t^2 - O_p(T^{2d_1-1}) = \frac{1}{T}\sum_{t=1}^{T}v_t^2 + o_p(1) \xrightarrow{p} \gamma_v(0), \end{aligned}$$

because $2d_1 - 1 < 0$. Moreover, we have $\sum_{t=1}^{T}x_t^{o2}/T^3 \xrightarrow{p} 1/3$ by using item 3 of Lemma A.1. So we prove $T^3 s_{\widehat{\beta}_{(0)}}^2 \xrightarrow{p} 3\gamma_v(0)$. For the numerator, we note

$$\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right) \Rightarrow \beta_{0*}$$

by using item 1 of Theorem 5. Combining the asymptotic distributions of the numerator and denominator, item 1.1 is proved.

To prove item 1.2, we see

$$R^2 = \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2}{\sum_{t=1}^{T} (C_t - \bar{C})}$$

$$= \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2}{\beta^2 \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 + 2\beta \sum_{t=1}^{T} (x_t^o - \bar{x}^o)(v_t - \bar{v}) + \sum_{t=1}^{T} (v_t - \bar{v})^2}.$$

From items 4 and 6 of Lemma 1 of Chapter 2 and items 1 and 3 of Lemma A.1, we have

$$\sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 = O(T^3), \qquad \sum_{t=1}^{T} (v_t - \bar{v})^2 = O_p(T),$$

and

$$\sum_{t=1}^{T} (x_t^o - \bar{x}^o)(v_t - \bar{v}) = O_p(T^{1.5+d_1}).$$

Consequently,

$$R^2 = \frac{\widehat{\beta}^2_{(0)}}{\beta^2 + O_p(T^{d_1-1.5}) + O_p(T^{-2})} = \frac{\left[\beta + O_p(T^{d_1-1.5})\right]^2}{\beta^2 + o_p(1)} \xrightarrow{p} 1,$$

where the convergence in probability is based on $\widehat{\beta}_{(0)} = \beta + o_p(1)$ and the second and the third terms in the denominator coverge in probability to zero because $d_1 - 1.5 < 0$.

To prove item 1.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left[v_t - v_{t-1} - \left(\widehat{\beta}_{(0)} - \beta\right)(1 + w_t)\right]^2$$

$$= (v_t - v_{t-1})^2 - 2\left(\widehat{\beta}_{(0)} - \beta\right)(v_t - v_{t-1})(1 + w_t)$$

$$+ \left(\widehat{\beta}_{(0)} - \beta\right)^2 (1 + w_t)^2.$$

And from items 4 and 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})(1+w_t)$$

$$= \max\left\{o_p(T^{\epsilon-0.5+d_1}, T^{\epsilon-1+d_1+d_2}), o_p(T^{\epsilon-0.5})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 1 of Theorem 5 that $\widehat{\beta}_{(0)} - \beta = O_p(T^{d_1-1.5})$ converges in probability to zero. From items 4 and 5 of Lemma 1 of Chapter 2, we also know that

$$\frac{1}{T}\sum_{t=2}^{T}(1+w_t)^2 = \frac{1}{T}\cdot T + 2\cdot\frac{1}{T}\sum_{t=2}^{T}w_t + \frac{1}{T}\sum_{t=2}^{T}w_t^2 \xrightarrow{p} 1 + \gamma_w(0)$$

and that

$$\frac{1}{T}\sum_{t=2}^{T}(v_t - v_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T}v_t^2 + \frac{1}{T}\sum_{t=2}^{T}v_{t-1}^2 - \frac{2}{T}\sum_{t=2}^{T}v_t v_{t-1}$$

$$\xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\gamma_v(1).$$

Finally, DW statistic is

$$DW \xrightarrow{p} 2 - 2\rho_v(1)$$

because the second and third terms in the numerator all converge in probability to zero and the denominator $\sum_{t=1}^{T}\widehat{u}_t^2/T = s_0^2$ converges in probability to $\gamma_v(0)$ from the above results.

To prove item 2.1, we see

$$\frac{\sqrt{T}}{\sigma_y}t_{\beta_{(1)}} = \frac{\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right)}{\left(T^3\, s^2_{\beta_{(1)}}\right)^{1/2}}.$$

For the denominator, we note

$$T^3\, s^2_{\beta_{(1)}} = \frac{s_1^2}{\frac{1}{T^3}\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2},$$

where

$$s_1^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t^o\right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t^o &= \alpha + \beta x_t^o + v_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)}x_t^o \\ &= (v_t - \bar{v}) - \left(\widehat{\beta}_{(1)} - \beta\right)(x_t^o - \bar{x}^o). \end{aligned}$$

And from item 2 of Theorem 5, we have $\widehat{\beta}_{(1)} - \beta = O_p(T^{d_1 - 1.5})$. Moreover, we know from items 1 and 3 of Lemma A.1 and items 4 and 6 of Lemma 1 of Chapter 2 that

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2 = O(T^3), \quad \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 \xrightarrow{p} \gamma_v(0),$$

and

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)(v_t - \bar{v}) = O_p(T^{1.5+d_1}).$$

Therefore, we have

$$\begin{aligned} s_1^2 = {} & \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 - \frac{1}{T}\left(\widehat{\beta}_{(1)} - \beta\right)^2\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 - O(T^{-1})\cdot O_p(T^{2d_1-3})\cdot O_p(T^3) \\ &= \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 - O_p(T^{2d_1-1}) = \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2 + o_p(1) \\ &\xrightarrow{p} \gamma_v(0), \end{aligned}$$

where the convergence in probability to $\gamma_v(0)$ is based on item 6 of Lemma 1 of Chapter 2. Moreover, we have $\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2/T^3 \xrightarrow{p}$

1/12 by using item 3 of Lemma A.1. So we prove $T^3 s^2_{\beta_{(0)}} \xrightarrow{p} 12\gamma_v(0)$. For the numerator, we note

$$\frac{T^2}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \;\Rightarrow\; \beta_{1*}$$

by using item 2 of Theorem 5. Combining the asymptotic distributions of the numerator and denominator, item 2.1 is proved.

To prove item 2.2, we see

$$R^2 = 1 - \frac{Ts_1^2}{\sum_{t=1}^{T}\left(C_t - \bar{C}\right)^2}$$

$$= 1 - \frac{Ts_1^2}{\beta^2 \sum_{t=1}^{T}\left(x_t^o - \bar{x}^o\right)^2 + 2\beta \sum_{t=1}^{T}\left(v_t - \bar{v}\right)\left(x^o - \bar{x}^o\right) + \sum_{t=1}^{T}\left(v_t - \bar{v}\right)^2},$$

given there is a constant included in the regression. From items 1 and 3 of Lemma A.1 and items 4 and 6 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2 = O(T^3), \quad \sum_{t=1}^{T}(v_t - \bar{v})^2 = O_p(T),$$

and

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)(v_t - \bar{v}) = O_p(T^{1.5+d_1}).$$

We also note $s_1^2 \xrightarrow{p} \gamma_v(0)$, consequently,

$$R^2 = 1 - \frac{O_p(T)}{O(T^3) + O_p(T^{1.5+d_1}) + O_p(T)}$$

$$= 1 - \frac{O_p(1)}{O(T^2) + O_p(T^{d_1+0.5}) + O_p(1)}$$

$$= 1 + o_p(1) \xrightarrow{p} 1,$$

where the first two terms in the denominator of the second term diverge because $d_1 + 0.5 > 0$.

To prove item 2.3, we first note

$$\left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2 = \left[v_t - v_{t-1} - \left(\widehat{\beta}_{(1)} - \beta\right)(1 + w_t)\right]^2$$

$$= \left(v_t - v_{t-1}\right)^2 - 2\left(\widehat{\beta}_{(1)} - \beta\right)\left(v_t - v_{t-1}\right)(1 + w_t) +$$

$$\left(\widehat{\beta}_{(1)} - \beta\right)^2 (1 + w_t)^2.$$

And from items 4 and 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T}\sum_{t=2}^{T}\left(v_t - v_{t-1}\right)(1 + w_t)$$

$$= \max\left\{o_p(T^{\epsilon-0.5+d_1}, T^{\epsilon-1+d_1+d_2}), o_p(T^{\epsilon-0.5})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 2 of Theorem 5 that $\widehat{\beta}_{(1)} - \beta = O_p(T^{d_1-1.5})$ converges in probability to zero. From items 4 and 5 of Lemma 1 of Chapter 2, we also know that

$$\frac{1}{T^2}\sum_{t=2}^{T}(1 + w_t)^2 = \frac{1}{T}\cdot T + \frac{1}{T}\sum_{t=2}^{T} w_t + \frac{1}{T}\sum_{t=2}^{T} w_t^2 \xrightarrow{p} 1 + \gamma_w(0),$$

and that

$$\frac{1}{T}\sum_{t=2}^{T}\left(v_t - v_{t-1}\right)^2 \xrightarrow{p} \gamma_v(0) + \gamma_v(0) - 2\gamma_v(1).$$

Finally, DW statistic is

$$DW \xrightarrow{p} 2 - 2\rho_v(1)$$

because the second and third terms in the numerator all converge in probability to zero and the denominator $\sum_{t=1}^{T}\widehat{u}_t^2/T = s_1^2$ converges in probability to $\gamma_v(0)$ from the above results.

The proof of item 3 of this theorem is identical to that of item 3 of Theorem 2, we won't repeat here.

A.8. Proof of Theorem 7

To prove item 1, we see

$$\frac{T}{\sigma_y}\left(\widehat{\beta}_{(0)}-\beta\right)=\frac{\dfrac{1}{T^2\sigma_y}\sum_{t=1}^{T}x_t^o y_t}{\dfrac{1}{T^3}\sum_{t=1}^{T}x_t^{o2}}\Rightarrow\beta_{0*},$$

where the weak convergence is based on items 2 and 3 of Lemma A.1 and the CMT.

To prove item 2, we see

$$\begin{bmatrix}\widehat{\alpha}_{(1)}-\alpha\\ \widehat{\beta}_{(1)}-\beta\end{bmatrix}=\begin{bmatrix}T & \sum_{t=1}^{T}x_t^o\\ \sum_{t=1}^{T}x_t^o & \sum_{t=1}^{T}x_t^{o2}\end{bmatrix}^{-1}\begin{bmatrix}\sum_{t=1}^{T}y_t\\ \sum_{t=1}^{T}x_t^o y_t\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\left(\sum_{t=1}^{T}x_t^{o2}\right)\left(\sum_{t=1}^{T}y_t\right)-\left(\sum_{t=1}^{T}x_t^o\right)\left(\sum_{t=1}^{T}x_t^o y_t\right)}{T\sum_{t=1}^{T}\left(x_t^o-\bar{x}^o\right)^2}\\ \dfrac{-\left(\sum_{t=1}^{T}x_t^o\right)\left(\sum_{t=1}^{T}y_t\right)+T\left(\sum_{t=1}^{T}x_t^o y_t\right)}{T\sum_{t=1}^{T}\left(x_t^o-\bar{x}^o\right)^2}\end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} \frac{1}{\sigma_y}\left(\widehat{\alpha}_{(1)} - \alpha\right) \\ \frac{T}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\frac{1}{T^4\sigma_y}\left[\left(\sum_{t=1}^{T} x_t^{o2}\right)\left(\sum_{t=1}^{T} y_t\right) - \left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} y_t x_t^o\right)\right]}{\frac{1}{T^3}\sum_{t=1}^{T}\left(x_t^o - \bar{x}^o\right)^2} \\ \dfrac{\frac{1}{T^3\sigma_y}\left[-\left(\sum_{t=1}^{T} x_t^o\right)\left(\sum_{t=1}^{T} y_t\right) + T\left(\sum_{t=1}^{T} x_t^o y_t\right)\right]}{\frac{1}{T^3}\sum_{t=1}^{T}\left(x_t^o - \bar{x}^o\right)^2} \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{1*} \\ \beta_{1*} \end{bmatrix},$$

where the weak convergence is based on items 2 and 3 of Lemma A.1 and item 1 of Lemma 1 of Chapter 2 and the CMT.

To prove item 3, we have to transform $C_t = \alpha + \gamma t + \beta x_t^o + y_t$ to be $C_t = \alpha + (\gamma + \beta)\, t + \beta x_t + v_t$ by the same arguments in item 3 of Theorem 5. The remaining proof can follow the same steps in the proof of item 3 of Theorem 3.

A.9. Proof of Theorem 8

To prove item 1.1, we see

$$\frac{1}{\sqrt{T}}\, t_{\beta_{(0)}} = \frac{\frac{T}{\sigma_y}\left(\widehat{\beta}_{(0)} - \beta\right)}{\left(\frac{T^3}{\sigma_y^2}\, s^2_{\beta_{(0)}}\right)^{1/2}}.$$

For the denominator, we note

$$\frac{T^3}{\sigma_y^2} s_{\beta_{(0)}}^2 = \frac{s_0^2}{\sigma_y^2} \frac{1}{\frac{1}{T^3} \sum_{t=1}^{T} x_t^{o2}},$$

where

$$s_0^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t^2 = \frac{1}{T} \sum_{t=1}^{T} \left(C_t - \widehat{\beta}_{(0)} x_t^o \right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\beta}_{(0)} x_t^o &= \beta x_t^o + y_t - \widehat{\beta}_{(0)} x_t^o \\ &= y_t - \left(\widehat{\beta}_{(0)} - \beta \right) x_t^o. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \frac{s_0^2}{\sigma_y^2} &= \frac{1}{T\sigma_y^2} \sum_{t=1}^{T} y_t^2 - \left[\frac{T}{\sigma_y} \left(\widehat{\beta}_{(0)} - \beta \right) \right]^2 \frac{1}{T^3} \sum_{t=1}^{T} x_t^{o2} \\ &\Rightarrow \int_0^1 [B_{0.5+d_1}(s)]^2 \, ds - 3 \left[\int_0^1 s B_{0.5+d_1}(s) \, ds \right]^2 \\ &= \sigma_{0*}^2, \end{aligned}$$

where the weak convergence is based on item 2 of Lemma 1 of Chapter 2 and items 2 and 3 of Lemma A.1 and item 1 of Theorem 7. Moreover, we have $\sum_{t=1}^{T} x_t^{o2}/T^3 \xrightarrow{p} 1/3$ by using item 3 of Lemma A.1. So we prove

$$\frac{T^3}{\sigma_y^2} s_{\beta_{(0)}}^2 \Rightarrow 3 \int_0^1 [B_{0.5+d_1}(s)]^2 \, ds - 9 \left[\int_0^1 s B_{0.5+d_1}(s) \, ds \right]^2.$$

For the numerator, we note

$$\frac{T}{\sigma_y} \left(\widehat{\beta}_{(0)} - \beta \right) \Rightarrow \beta_{0*}$$

by using item 1 of Theorem 7. Combining the asymptotic distributions of the numerator and denominator, item 1.1 is proved.

To prove item 1.2, we see

$$R^2 = \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2}{\sum_{t=1}^{T} (C_t - \bar{C})}$$

$$= \frac{\widehat{\beta}^2_{(0)} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2}{\beta^2 \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 + 2\beta \sum_{t=1}^{T} (x_t^o - \bar{x}^o)(y_t - \bar{y}) + \sum_{t=1}^{T} (y_t - \bar{y})^2}.$$

From items 2 and 3 of Lemma A.1 and items 1 and 3 of Lemma 1 of Chapter 2, we have

$$\sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 = O(T^3), \quad \sum_{t=1}^{T} (y_t - \bar{y})^2 = O_p(T^{2+2d_1}),$$

and

$$\sum_{t=1}^{T} (x_t^o - \bar{x}^o)(y_t - \bar{y}) == O_p(T^{2.5+d_1}).$$

Consequently,

$$R^2 = \frac{\widehat{\beta}^2_{(0)}}{\beta^2 + O_p(T^{d_1-0.5}) + O_p(T^{2d_1-1})} = \frac{\left[\beta + O_p(T^{d_1-0.5})\right]^2}{\beta^2 + o_p(1)} \xrightarrow{p} 1,$$

where the convergence in probability is based on $\widehat{\beta}_{(0)} = \beta + o_p(1)$ and the second and the third terms in the denominator coverge in probability to zero because $d_1 - 0.5 < 0$.

To prove item 1.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left[v_t - \left(\widehat{\beta}_{(0)} - \beta\right)(1 + w_t)\right]^2$$

$$= v_t^2 - 2\left(\widehat{\beta}_{(0)} - \beta\right) v_t (1 + w_t) + \left(\widehat{\beta}_{(0)} - \beta\right)^2 (1 + w_t)^2 .$$

And from items 4 and 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T\sigma_y}\sum_{t=2}^{T} v_t (1 + w_t)$$

$$= \max\left\{o_p(T^{\epsilon-1}), o_p(T^{\epsilon-1.5+d_2}), o_p(T^{\epsilon-1-d_1})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 1 of Theorem 7 that $\left(\widehat{\beta}_{(0)} - \beta\right)/\sigma_y = O_p(T^{-1})$ converges in probability to zero. From items 4 and 5 of Lemma 1 of Chapter 2, we also know that

$$\sum_{t=2}^{T} \frac{v_t^2}{T} \xrightarrow{p} \gamma_v(0)$$

and that

$$\frac{1}{T}\sum_{t=2}^{T} (1 + w_t)^2 = \frac{1}{T} \cdot T + 2 \cdot \frac{1}{T}\sum_{t=2}^{T} w_t + \frac{1}{T}\sum_{t=2}^{T} w_t^2 \xrightarrow{p} 1 + \gamma_w(0).$$

All these results imply

$$DW \xrightarrow{p} 0,$$

becuse the three terms in the numerator all converge in probability to zero while the denominator $\sum_{t=1}^{T} \widehat{u}_t^2/T\sigma_y^2 = s_0^2/\sigma_y^2 \Rightarrow \sigma_{0*}^2$ from the above results.

To prove item 2.1, we see

$$\frac{1}{\sqrt{T}} t_{\beta_{(1)}} = \frac{\dfrac{T}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right)}{\left(\dfrac{T^3}{\sigma_y^2} s_{\beta_{(1)}}^2\right)^{1/2}}.$$

For the denominator, we note

$$\frac{T^3}{\sigma_y^2} s_{\beta_{(1)}}^2 = \frac{s_1^2}{\sigma_y^2} \frac{1}{\frac{1}{T^3} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2},$$

where

$$s_1^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t^2 = \frac{1}{T} \sum_{t=1}^{T} \left(C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t^o \right)^2,$$

and

$$\begin{aligned} C_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t^o &= \alpha + \beta x_t^o + y_t - \widehat{\alpha}_{(1)} - \widehat{\beta}_{(1)} x_t^o \\ &= (y_t - \bar{y}) - \left(\widehat{\beta}_{(1)} - \beta \right) (x_t^o - \bar{x}^o). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \frac{s_1^2}{\sigma_y^2} &= \frac{1}{T\sigma_y^2} \sum_{t=1}^{T} (y_t - \bar{y})^2 - \left[\frac{T}{\sigma_y} \left(\widehat{\beta}_{(1)} - \beta \right) \right]^2 \frac{1}{T^3} \sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 \\ &\Rightarrow \int_0^1 \left[B_{0.5+d_1}(s) \right]^2 ds - \left[\int_0^1 B_{0.5+d_1}(s)\, ds \right]^2 \\ &\quad - 12 \left[\int_0^1 s B_{0.5+d_1}(s)\, ds - \frac{\int_0^1 B_{0.5+d_1}(s)\, ds}{2} \right]^2 \\ &\Rightarrow \sigma_{1*}^2, \end{aligned}$$

where the weak convergence is based on item 2 of Theorem 7 and items 1 and 3 of Lemma 1 of Chapter 2 and items 2 and 3 of Lemma A.1. Moreover, we have $\sum_{t=1}^{T} (x_t^o - \bar{x}^o)^2 / T^3 \xrightarrow{p} 1/12$ by using item 3 of Lemma A.1. So we prove

$$\frac{T^3}{\sigma_y^2} s_{\beta_{(1)}}^2 \Rightarrow 12\sigma_{1*}^2.$$

For the numerator, we note

$$\frac{\sigma_x}{\sigma_y}\left(\widehat{\beta}_{(1)} - \beta\right) \;\Rightarrow\; \beta_{1*}$$

by using item 2 of Theorem 7. Combining the asymptotic distributions of the numerator and denominator, item 2.1 is proved.

To prove item 2.2, we see

$$R^2 = 1 - \frac{Ts_1^2}{\sum_{t=1}^{T}(C_t - \bar{C})^2}$$

$$= 1 - \frac{Ts_1^2}{\beta^2\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2 + 2\beta\sum_{t=1}^{T}(y_t - y_t)(x^o - \bar{x}^o) + \sum_{t=1}^{T}(y_t - \bar{y})^2},$$

given there is a constant included in the regression. From items 1 and 3 of Lemma 1 of Chapter 2 and item 3 of Lemma A.1 we have

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)^2 = O(T^3), \quad \sum_{t=1}^{T}(y_t - \bar{y})^2 = O_p(T^{2+2d_1}),$$

and

$$\sum_{t=1}^{T}(x_t^o - \bar{x}^o)(y_t - \bar{y}) = O_p(T^{2.5+d_1}).$$

We also note $s_1^2/\sigma_y^2 \;\Rightarrow\; \sigma_{1*}^2$, consequently,

$$R^2 = 1 - \frac{O_p(T^{2+2d_1})}{O(T^3) + O_p(T^{2.5+d_1}) + O_p(T^{2+2d_1})}$$

$$= 1 - \frac{O_p(1)}{O(T^{1-2d_1}) + O_p(T^{0.5-d_1}) + O_p(1)}$$

$$= 1 + o_p(1) \xrightarrow{p} 1,$$

where the first and the second terms in the denominator of the second term diverge because $0.5 - d_1 > 0$.

To prove item 2.3, we first note

$$(\widehat{u}_t - \widehat{u}_{t-1})^2 = \left[v_t - \left(\widehat{\beta}_{(1)} - \beta\right)(1 + w_t)\right]^2$$

$$= v_t^2 - 2\left(\widehat{\beta}_{(1)} - \beta\right) v_t (1 + w_t) + \left(\widehat{\beta}_{(1)} - \beta\right)^2 (1 + w_t)^2 .$$

And from items 4 and 7 of Lemma 1 of Chapter 2, we have

$$\frac{1}{T\sigma_y} \sum_{t=2}^{T} v_t (1 + w_t) = \max\left\{o_p(T^{\epsilon-1}), o_p(T^{\epsilon-1.5+d_2}), o_p(T^{\epsilon-1-d_1})\right\},$$

for any $\epsilon > 0$. Moreover, we know from item 2 of Theorem 7 that $\left(\widehat{\beta}_{(1)} - \beta\right)/\sigma_y = O_p(T^{-1})$ converges in probability to zero. From items 4 and 5 of Lemma 1 of Chapter 2, we also know that

$$\sum_{t=2}^{T} \frac{v_t^2}{T} \xrightarrow{p} \gamma_v(0)$$

and that

$$\frac{1}{T}\sum_{t=2}^{T} (1 + w_t)^2 = \frac{1}{T}\cdot T + 2\cdot\frac{1}{T}\sum_{t=2}^{T} w_t + \frac{1}{T}\sum_{t=2}^{T} w_t^2 \xrightarrow{p} 1 + \gamma_w(0).$$

All these results imply

$$DW \xrightarrow{p} 0,$$

because the three terms in the numerator all converge in probability to zero while the denominator $\sum_{t=1}^{T} \widehat{u}_t^2 / T\sigma_y^2 = s_1^2/\sigma_y^2 \Rightarrow \sigma_{1*}^2$ from the above results.

The proof of item 3 of this theorem is identical to that of item 3 of Theorem 4, we won't repeat here.

TABLE 3–1

REGRESSION OF $C_t = 1 + x_t + \varepsilon_t$ ON 1 AND x_t, WHERE $x_t = I(1)$ AND $\varepsilon_t = I(d)$

d	T	1%	5%	10%	20%	30%	R^2	DW	$\mid t_\beta \mid$
	125	0.237	0.364	0.446	0.557	0.635	0.9158	1.375	1.7553
0.3	250	0.327	0.454	0.529	0.626	0.692	0.9542	1.312	2.1111
	500	0.416	0.532	0.602	0.684	0.744	0.9753	1.266	2.5547
	125	0.604	0.696	0.742	0.794	0.830	0.8062	0.458	4.0987
0.7	250	0.708	0.773	0.808	0.851	0.881	0.8551	0.333	5.7237
	500	0.782	0.839	0.864	0.894	0.915	0.8986	0.243	8.1469

Note: the critical values of the two-tailed t tests are ± 2.576 for $\aleph = 0.01$, $\pm$ 1.96 for $\aleph = 0.05$, $\pm$ 1.645 for $\aleph = 0.10$, $\pm$ 1.282 for $\aleph = 0.20$, $\pm$ 1.0326 for $\aleph = 0.30$. R^2 denotes the average R^2 of the simulation. DW denotes the average DW of the simulation. $\mid t_\beta \mid$ is the average absolute value of t_β of the simulation.

TABLE 3–2

THE DIVERGENCE RATE OF MEAN $|t_\beta|$ UNDER MODELS A–1 AND B–1

For $\varepsilon_t = I(0.3)$,

$$\frac{2.1111}{1.7553} = \underline{0.9769} \cdot 2^{0.3} \qquad \frac{2.5547}{2.1111} = \underline{0.9829} \cdot 2^{0.3}.$$

For $\varepsilon_t = I(0.7)$,

$$\frac{5.7237}{4.0987} = \underline{0.9875} \cdot 2^{0.5} \qquad \frac{8.1469}{5.7237} = \underline{1.0065} \cdot 2^{0.5}.$$

Note: the above numbers are taken from the last column of Table 3–1.

TABLE 3–3

REGRESSION OF $C_t = 1 + x_t^o + \varepsilon_t$ ON 1 AND x_t^o, WHERE $x_t^o = t + I(1)$ AND $\varepsilon_t = I(d)$

d	T	1%	5%	10%	20%	30%	R^2	DW	$\lvert t_\beta \rvert$
	125	0.317	0.448	0.519	0.614	0.682	0.9991	1.392	2.0598
0.3	250	0.407	0.526	0.589	0.672	0.735	0.9998	1.322	2.4628
	500	0.483	0.594	0.653	0.724	0.776	0.9999	1.275	2.9763
	125	0.735	0.798	0.829	0.867	0.889	0.9981	0.503	6.1141
0.7	250	0.802	0.850	0.874	0.901	0.920	0.9993	0.366	8.4465
	500	0.863	0.896	0.914	0.932	0.945	0.9998	0.269	11.8042

Note: the critical values of the two-tailed t tests are ± 2.576 for $\aleph = 0.01$, $\pm$ 1.96 for $\aleph = 0.05$, $\pm$ 1.645 for $\aleph = 0.10$, $\pm$ 1.282 for $\aleph = 0.20$, $\pm$ 1.0326 for $\aleph = 0.30$. R^2 denotes the average R^2 of the simulation. DW denotes the average DW of the simulation. $\lvert t_\beta \rvert$ is the average absolute value of t_β of the simulation.

TABLE 3–4

THE DIVERGENCE RATE OF MEAN $|t_\beta|$

UNDER MODELS C–1 AND D–1

For $\varepsilon_t = I(0.3)$,

$$\frac{2.4628}{2.0598} = \underline{0.9712} \cdot 2^{0.3} \qquad \frac{2.9763}{2.4628} = \underline{0.9816} \cdot 2^{0.3}.$$

For $\varepsilon_t = I(0.7)$,

$$\frac{8.4465}{6.1141} = \underline{0.9769} \cdot 2^{0.5} \qquad \frac{11.8042}{8.4465} = \underline{0.9882} \cdot 2^{0.5}.$$

Note: the above numbers are taken from the last column of Table 3–3.

CHAPTER 4

CONCLUSION

In this dissertation we consider spurious effects in a simple linear regression model of $I(d)$ processes. In Chapter 2 we find that spurious effects could occur when we regress a fractionally integrated process on a constant and another independent fractionally integrated process. The most interesting finding is that spurious effects could occur even when both the dependent variable and regressor are stationary. This implies the usual differencing procedure may not be sufficient for a complete avoidance of the spurious effect. The recent findings of the existence of long memory in many macroeconomic and financial time series remind us of the possibility that spurious effects may present in some previous empirical work.

In Chapter 3 we consider the asymptotic theory for the OLS estimators and the conventional test statistics when the regressor and disturbance term are independent fractionally integrated processes. The main finding is that nonstationarity in the regressor and long memory in the disturbance term may result in over-rejection of the null hypothesis.

From the analysis in Chapter 2 and Chapter 3, we conclude that the possible presence of fractionally integrated processes in the regression model may render the usual asymptotic theory for the OLS estimation useless. Before estimating the regression model, any suspicion of long memory in the variables should be investigated.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Baillie, R.T. (1995), "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, forthcoming.

[2] Cheung, Y.W. and K.S. Lai (1993), "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business and Economic Statistics*, 11, 103-112.

[3] Davydov, Y.A. (1970), "The Invariance Principle for Stationary Processes," *Theory of Probability and Its Applications*, 15, 487–489.

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

[5] Fox, R. and M.S. Taqqu (1987), "Multiple Stochastic Integrals with Dependent Integrators," *Journal of Multivariate Analysis*, 21, 105–127.

[6] Granger, C.W.J. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

[7] Granger, C.W.J. (1981), "Some Properties of Time Series Data and their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

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

[9] Granger, C.W.J. and R. Joyeux (1980), "An Introduction to Long-Memory Time Series Models and Fractionally Differencing," *Journal of Time Series Analysis*, 1, 15–29.

[10] Hosking, J.R.M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

[11] Hosking, J.R.M. (1984), "Modeling Persistence in Hydrological Time Series Using Fractional Differencing," *Water Resources Research*, 20, 1898–1908.

[12] Hurst, H. (1951), "Long-Term Storage Capacity of Reservoirs," *Trans. Amer. Soc. Civil Engrs.*, 116, 770-799.

[13] Kramer, W. (1986), "Least Squares Regression when the Independent Variable Follows an ARIMA Process," *Journal of the American Statistical Association*, 81, 150–154.

[14] Mandelbrot, B.B. and J.W. Van Ness (1968), "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, 10, 422–437.

[15] McLeod, A.I. and K.W. Hipel (1978), "Preservation of the Rescaled Adjusted Range 1: A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 14, 491–508.

[16] Nelson, C.R. and H. Kang (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica*, 49, 741–751.

[17] Nelson, C.R. and H. Kang (1984), "Pitfalls in the Use of Time as an Explanatory Variable in Regression," *Journal of Business and Economic Statistics*, 2, 73–82.

[18] Nelson, C.R. and C. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

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

[20] Park, J.Y. and P.C.B. Phillips (1989), "Statistical Inference in Regression with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–131.

[21] Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

[22] Phillips, P.C.B. and J.Y. Park (1988), "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111–115.

[23] Robinson, P.M. and F.J. Hidalgo (1995), "Time Series Regression with Long Range Dependence," unpublished manuscript.

[24] Sowell, F (1990), "The Fractional Unit Root Distribution," *Econometrica*, 58, 495–505.