# Mean Reversion and Stationarity in Financial Time Series Generated from Diffusion Models*

Jihyun Kim†
Toulouse School of Economics
University of Toulouse Capitole

Joon Y. Park‡
Department of Economics
Indiana University
and Sungkyunkwan University

October 15, 2016

**Abstract**

This paper analyzes the mean reversion and unit root properties of general diffusion models and their discrete samples. In particular, we find that the Dickey-Fuller unit root test applied to discrete samples from a diffusion model becomes a test of no mean reversion rather than a unit root, or more generally, nonstationarity in the underlying diffusion. The unit root test has a well defined limit distribution if and only if the underlying diffusion has no mean reversion, and diverges to minus infinity in probability if and only if the underlying diffusion has mean reversion. It is shown, on the other hand, that diffusions are mean-reverting as long as their drift terms play the dominant role, and therefore, nonstationary diffusions may well have mean reversion.

JEL Classification: C12, C22, C58

Key words and phrases: mean reversion, stationarity and nonstationarity, diffusion model, unit root test, stationarity test.

---

*We are grateful to Yacine Aït-Sahalia, Yoosoon Chang, Oliver Linton, Morten Ørregaard Nielsen, Nour Meddahi, Christian Gouriéroux, Olivier Scaillet and seminar participants at Indiana University, Toulouse School of Economics, University of Cambridge, University of Maastricht, Singapore Management University, Rutgers University, University of Illinois at Urbana-Champaign, University of York, University of Liverpool, CREST, Aarhus University, 2013 Midwest Econometrics Group Meetings, 2013 North American Econometric Society Meeting, 2013 SETA, 2014 Toulouse Financial Econometrics Conference, 7th French Econometrics Conference, and 9th International Conference on Computational and Financial Econometrics for useful comments and discussions. Various versions of the paper have been circulated under the titles "Mean Reversion and Stationarity: A New Perspective from Asymptotics of Diffusion Models" and "Mean Reversion and Unit Root Properties of Diffusion Models".

†31000 Toulouse, France. E-mail: jihyun.kim@tse-fr.eu.

‡Corresponding author. Bloomington, IN 47405-7104, USA. E-mail: joon@indiana.edu.

## 1 Introduction

The unit root test was originally developed to test for the unity of the largest root in an autoregressive time series model. However, the existence of a unit root entails some other important consequences as well on more general stochastic characteristics of the underlying time series. In particular, it implies that, every shock having a permanent effect, the effects of shocks are accumulated and build up a stochastic trend, which eventually makes any time series with a unit root nonstationary and non-mean-reverting. On the other hand, an autoregressive process becomes stationary and mean-reverting if it has the largest root less than unity in modulus, since every shock has only a transitory effect and is offset by a future shock with the opposite sign. Accordingly, the unit root test may also be regarded generally as a test for nonstationarity and no mean reversion of the underlying time series. In fact, the unit root test has commonly been used to test for nonstationarity and no mean reversion in a much broader class of time series models than autoregressive models.

Naturally, the unit root test has also been routinely applied to samples that are thought to be collected discretely at relatively high frequencies from diffusion type continuous time processes. As is well known, the laws of motions for economic and financial time series such as stock prices, exchange rates and interest rates, among others, are typically modeled as diffusions. However, the asymptotics of the unit root test applied to discrete samples from general diffusion models have long been unknown, except for the simple case that the underlying diffusion models are driven by Brownian motion and stationary Ornstein-Uhlenbeck processes respectively for the null and alternative hypotheses, as in Shiller and Perron (1985), Perron (1989, 1991) and Chambers (2004, 2008), among others. In particular, it has been completely unknown whether the unit root test has any discriminatory power in distinguishing general stationary and nonstationary, or mean-reverting and non-mean-reverting, diffusion models. Moreover, the effects of time span and sampling frequency on the power of the unit root test have never been theoretically analyzed for general stationary diffusion models.

In this paper, we investigate the longrun behaviors of general diffusion models including their unit root and mean reversion properties. Our investigation is extensive and complete, where mean reversion is defined more generally as reversion to the sample mean. We consider the entire class of recurrent diffusions covering all positive and null recurrent diffusions, which includes in particular stationary diffusions having no proper moments as well as general nonstationary diffusions. For the sample from such a general class of diffusions, we show that the Dickey-Fuller unit root test has a well defined limit distribution if and only if the underlying diffusion does not have mean reversion, and it diverges to minus infinity in probability if and only if the underlying diffusion has mean reversion. The unit root test therefore has perfect discriminatory power, if used to discriminate non-mean-reverting diffusions against mean-reverting diffusions. Nevertheless, the test cannot be used to test for nonstationarity of the underlying diffusion. All stationary diffusions, including those without finite mean, are mean-reverting. However, nonstationary diffusions may also well be mean-reverting if they have drift terms dominating their diffusion terms.

The existence of mean-reverting nonstationary diffusions has some important and far-reaching implications. First, it implies that nonstationary financial time series may not be necessarily non-mean-reverting. This opens up a possibility, for instance, that stock prices, which are widely believed to be nonstationary, are still mean-reverting. See, e.g., Fama and French (1988), Poterba and Summers (1988) and Kim et al. (1991). Second, in any cointegrating relationships, we may allow for the disequilibrium error process to be nonstationary as long as it is mean-reverting. As a result, we may extend the notion of cointegration in financial time series studied in, e.g., Baillie and Bollerslev (1989, 1994) and Diebold et al. (1994). The extended notion of cointegration also has some important consequences in actual financial investments. In fact, one popular short-term speculation strategy known as pairs trading utilizes co-movements in the prices of two or more stocks, where we may define co-movements more generally using the extended notion of cointegration. For more details on pairs trading, the reader is referred to Bossaerts (1988), Bossaerts and Green (1989) and Gatev et al. (2006).

As a test of no mean reversion in our general setup, the limit distribution of the Dickey-Fuller test becomes heavily model-dependent and reies on the underlying diffusion model in a complicated manner. Therefore, the usual Dickey-Fuller critical values cannot be used. The limit distribution of the test is generally represented as a functional of a skew Bessel process, and reduces to the Dickey-Fuller distribution only if the underlying diffusion becomes a Brownian motion in the limit. However, the asymptotic critical values of the test may be obtained using the standard sub-sampling procedure, and the usual sub-sample test relying on sub-sample critical values is asymptotically valid. As an illustration, we use the test to examine the presence of mean reversion in some important financial time series, and study the finite sample performance by simulation using their fitted models. The exchange rates of Japanese Yen and Swiss Franc against US Dollars are mean-reverting, while those of all other currencies are not. Moreover, VIX index is clearly mean-reverting, whereas interest rates are generally not.

The rest of the paper is organized as follows. Section 2 presents the background and preliminaries that are necessary to understand the subsequent development of asymptotic theory developed in the paper. The diffusion model and various notions to investigate its longrun behaviors, and regularly and rapidly varying functions with their useful properties are introduced. In Section 4, We consider the Dickey-Fuller unit root test and develop its asymptotics. We consider the mean-reverting behaviors of diffusion models, and show that the unit root test is indeed a test for the absence of mean reversion. Section 5 concludes the paper, and all mathematical proofs are in Appendix.

# 2 Background and Preliminaries

## 2.1 Diffusion Model

We consider the diffusion process $X$ given by the time-homogeneous stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

where $\mu$ and $\sigma$ are, respectively, the drift and diffusion functions, and $W$ is the standard Brownian motion. We denote by $\mathcal{D} = (\underline{x}, \overline{x})$ the domain of the diffusion process $X$, where we set $\underline{x} = -\infty$ or 0 with $\overline{x} = \infty$. This causes no loss in generality, since we may simply consider $X - \underline{x}$ or $-X$ to allow for a more general case. In what follows, we denote by $x_B = \underline{x}$ or $\overline{x}$ the boundary of $\mathcal{D}$. Throughout the paper, we assume

**Assumption 2.1.** *We assume that (a) $\sigma^2(x) > 0$ for all $x \in \mathcal{D}$, and (b) $\mu(x)/\sigma^2(x)$ and $1/\sigma^2(x)$ are locally integrable at every $x \in \mathcal{D}$.*

Assumption 2.1 provides a simple sufficient set of conditions to ensure that a weak solution to the stochastic differential equation (1) exists uniquely up to an explosion time. See, e.g., Theorem 5.5.15 in Karatzas and Shreve (1991).

The scale function of the diffusion process $X$ in (1) is defined as

$$s(x) = \int_w^x \exp\left(-\int_w^y \frac{2\mu(z)}{\sigma^2(z)}dz\right) dy, \tag{2}$$

where the lower limits of the integrals can be arbitrarily chosen to be any point $w \in \mathcal{D}$. Defined as such, the scale function $s$ is uniquely identified up to any increasing affine transformation, i.e., if $s$ is a scale function, then so is $as + b$ for any constants $a > 0$ and $-\infty < b < \infty$. We also define the speed density

$$m(x) = \frac{1}{(\sigma^2 s')(x)} \tag{3}$$

on $\mathcal{D}$, where $s'$ is the derivative of $s$, often called the scale density, which is assumed to exist. The speed measure is defined to be the measure on $\mathcal{D}$ given by the speed density with respect to the Lebesgue measure. Note, under Assumption 2.1, that both the scale function and speed density are well defined, and that the scale function is strictly increasing, on $\mathcal{D}$.

Our asymptotic theory depends crucially on the recurrence property of the diffusion process $X$. To define the recurrence property, we let $\tau_y$ be the hitting time of a point $y$ in $\mathcal{D}$ that is given by $\tau_y = \inf\{t \geq 0 | X_t = y\}$. We say that the diffusion $X$ is recurrent if $\mathbb{P}\{\tau_y < \infty | X_0 = x\} = 1$ for all $x, y \in \mathcal{D}$. The recurrent diffusion $X$ is said to be positive recurrent if $\mathbb{E}[\tau_y < \infty | X_0 = x] < \infty$ for all $x, y \in \mathcal{D}$, and null recurrent if $\mathbb{E}[\tau_y < \infty | X_0 = x] = \infty$ for all $x, y \in \mathcal{D}$. Under some regularity conditions on $\mu$ and $\sigma$, the diffusion $X$ is recurrent if and only if the scale function $s$ in (2) is unbounded at both boundaries $\underline{x}$ and $\overline{x}$, i.e.,

$$s(\underline{x}) = -\infty \quad \text{and} \quad s(\overline{x}) = \infty.$$

Furthermore, the recurrent diffusion $X$ becomes positive recurrent or null recurrent, depending upon

$$m(\mathcal{D}) < \infty \quad \text{or} \quad m(\mathcal{D}) = \infty,$$

where $m$ is the speed measure defined in (3). A diffusion which is not recurrent is said to be transient.

Positive recurrent diffusions are stationary. More precisely, they have time invariant distributions, and if they are started from the time invariant distributions they become stationary. The time invariant density of the positive recurrent diffusion $X$ is given by

$$\pi(x) = \frac{m(x)}{m(\mathcal{D})}.$$

Null recurrent and transient diffusions are nonstationary. They do not have time invariant distributions, and their marginal distributions change over time. Out of these two different types of nonstationary processes, we mainly consider null recurrent diffusions in the paper. Brownian motion is the prime example of null recurrent diffusions. Typically, transient processes have upward or downward trends, in which case we may eliminate their trends using appropriate detrending methods so that they behave like recurrent processes. Like unit root processes in discrete time, null recurrent processes have stochastic trends and the standard law of large numbers and central limit theory in continuous time are not applicable. See, e.g., Jeong and Park (2013) and Kim and Park (2015) for more details on the statistical properties of null recurrent diffusions.

Let $X^s = s(X)$ be the scale transformation of $X$, which may be defined as $dX_t^s =$

$m_s^{-1/2}(X_t^s)dW_t$ with speed measure $m_s$ given by

$$m_s = \frac{1}{(s'\sigma)^2 \circ s^{-1}}.$$

Both recurrence and stationarity are preserved under scale transformation. First, $X$ is recurrent on $\mathcal{D}$ if and only if $X^s$ is recurrent on $\mathbb{R}$. Trivially, the scale function of $X^s$ is identity, since it is already in natural scale, and therefore, $X^s$ is recurrent if and only if its domain is given by the entire real line $\mathbb{R}$. However, the domain of $X^s$ becomes $\mathbb{R}$ if and only if $X$ is recurrent, i.e., $s(\underline{x}) = -\infty$ and $s(\overline{x}) = \infty$. Second, $X$ is stationary on $\mathcal{D}$ if and only if $X^s$ is stationary on $\mathbb{R}$, since $m_s(\mathbb{R}) = m(\mathcal{D})$.

**Example 2.1.** For an illustration, we consider the generalized Höpfner and Kutoyants (GHK) diffusion defined as

$$dX_t = \frac{aX_t}{(c + X_t^2)^{1-b}}dt + (c + X_t^2)^{b/2}dW_t \tag{4}$$

on $\mathbb{R}$ for $a, b \in \mathbb{R}$ and $c > 0$. The GHK model encompasses several diffusion models that are used earlier for illustrative purposes. If, for instance, $a = 0$ or $b = 0$, the GHK diffusion reduces to the diffusion considered by Chen et al. (2010) or Höpfner and Kutoyants (2003), respectively. Moreover, the speed density and speed measure of the GHK model are given respectively by

$$s'(x) = (x^2 + c)^{-a} \quad \text{and} \quad m(x) = (x^2 + c)^{a-b}.$$

The GHK process becomes recurrent if $a \leq 1/2$. Moreover, it becomes positive recurrent if $a - b < -1/2$.

## 2.2 Regular Variation and Integrability Condition

We say that $f : (0, \infty) \to \mathbb{R}$ is regularly varying at infinity with index $\kappa$, and write as $f \in RV_\kappa$, if $f(\lambda x)/f(\lambda) \to x^\kappa$ as $\lambda \to \infty$ for all $x > 0$ with some $\kappa \in (-\infty, \infty)$. In particular, if $\kappa = 0$ and $f \in RV_0$, then $f$ is said to be slowly varying at infinity. See Bingham et al. (1993) for more discussions on regularly varying functions, as well as their alternative concepts and definitions. For our asymptotics, it is necessary to deal with functions defined on $\mathbb{R}$ and consider both boundaries $x_B = \pm\infty$. The required extension is straightforward and may easily be done as shown in Kim and Park (2015). In particular, for $f \in RV_\kappa$ on $\mathbb{R}$ for some $\kappa \in (-\infty, \infty)$, we have $f(\lambda x)/f(\lambda) \to \overline{f}(x)$ as $\lambda \to \infty$ for all

$x \neq 0$, where $\overline{f}$, called the *limit homogeneous function*, is given by

$$\overline{f}(x) = |x|^\kappa \big(a1\{x > 0\} + b1\{x < 0\}\big)$$

for some constants $a$ and $b$ such that $|a| + |b| \neq 0$. On the other hand, $f : \mathcal{D} \to \mathbb{R}$ is said to be rapidly varying at boundary $x_B$ with index $\infty$ or $-\infty$ if $\overline{\kappa} = \underline{\kappa} = \infty$ or $-\infty$ with $\overline{\kappa}$ and $\underline{\kappa}$ defined as $\overline{\kappa} = \sup_\kappa\{x^{-\kappa}f(x) \sim f_\kappa(x)$ at $x_B$ for some nondecreasing $f_\kappa\}$ and $\underline{\kappa} = \inf_\kappa\{x^{-\kappa}f(x) \sim f_\kappa(x)$ at $x_B$ for some nonincreasing $f_\kappa\}$. We write $f \in RV_\infty$ or $f \in RV_{-\infty}$ at $x_B$ for the rapidly varying $f$ of index $\infty$ and $-\infty$ at $x_B$, respectively.

Throughout the paper, we assume

**Assumption 2.2.** *We assume that (a) $s'$ is regularly or rapidly varying with index $\kappa \neq -1$, (b) $\sigma^2$ is regularly varying and (c) $m$ is either integrable or regularly varying.*

Note that $s' \in RV_{-1}$ if and only if $s \in RV_0$. In this case, $X$ may either be recurrent or transient, since a slowly varying function may converge or diverge. We exclude this boundary case in our asymptotic analysis. We may easily see that this case arises if and only if $x\mu(x)/\sigma^2(x) \to 1/2$ as $x \to x_B$ at $x_B = \pm\infty$. Furthermore, in this case, $s^{-1}$ becomes rapidly varying.

To effectively present our asymptotics, we introduce

**Definition 2.1.** Let $f$ be a nonintegrable (or $m$-nonintegrable) regularly varying function on $\mathcal{D}$. We say that $f$ is *strongly nonintegrable* (or $m$-strongly nonintegrable) if $f\ell$ is not integrable (or not $m$-integrable) for any slowly varying function $\ell$. On the other hand, we say that $f$ is *barely nonintegrable* ($m$-barely nonintegrable) if there exists some slowly varying function $\ell$ such that $f\ell$ is integrable (or $m$-integrable).

Following Definition 2.1, we say that a null recurrent diffusion $X$ is *strongly nonstationary* if its speed density $m$ is strongly nonintegrable, and *barely nonstationary* if its speed density $m$ is barely nonintegrable. We assume that $m_s$ and $s^{-1}$ have $+\infty$ as their dominating boundary, i.e., for $f = m_s, s^{-1}$ we have $f(-x)/f(x) = O(1)$ as $x \to \infty$. This assumption is not restrictive and made just for the convenience of exposition.

In the development of our asymptotics, we consider the following three conditions. Here and elsewhere in the paper, we denote by $\iota$ the identity function on $\mathcal{D}$, i.e., $\iota(x) = x$ for all $x \in \mathcal{D}$.

(ST) : $m$ is either integrable or barely nonintegrable,

(DD) : $1/s'$ is either integrable or barely nonintegrable, and

(SI) : $\iota^2$ is either $m$-integrable or $m$-barely nonintegrable.

Clearly, ST is a condition related to the stationarity of $X$, and it holds if and only if $X$ is stationary or barely nonstationary. It is easy to see that ST holds if and only if $m \in RV_\kappa$ with $\kappa \leq -1$ at $x_B = \pm\infty$ and $\kappa \geq -1$ at $x_B = 0$. On the other hand, DD provides a condition that $X$ has a drift dominating its volatility. In fact, if we set $X_t dX_t = dN_t + dM_t$ with $dN_t = X_t\mu(X_t)dt$ and $dM_t = X_t\sigma(X_t)dW_t$, then we may show that

**Lemma 2.1.** *Let Assumptions 2.1 and 2.2 hold. Then DD holds if and only if* $M_T = o_p(N_T)$ *as* $T \to \infty$.

Note that DD holds if and only if $\sigma^2$ is $m$-integrable or $m$-barely nonintegrable, and therefore, DD holds if and only if $m\sigma^2 \in RV_\kappa$ with $\kappa \leq -1$ at $x_B = \pm\infty$. Due to the recurrence condition, $\sigma^2$ is always $m$-integrable at $x_B = 0$. Finally, SI requires the $m$-square integrability of identity function.

**Example 2.2.** For the GHK diffusion introduced in Example 2.1, DD and ST hold if and only if $a \leq -1/2$ and $a - b \leq -1/2$ respectively, and SI is satisfied if $a - b \leq -3/2$.

Throughout the paper, we let $f_s = f \circ s^{-1}$ for any function $f$ other than $m$. We may easily show by a change of variables in integrals that $m_s(f_s) = m(f)$ for any $f$ defined on $\mathcal{D}$. Moreover, for locally integrable $f$ on $\mathbb{R}$, we define $[f]$ as

$$[f](\lambda) = \int_{|x|<\lambda} f(x)dx.$$

This notation will be used without reference in what follows.

# 3 Asymptotic Theory of Unit Root Test

In the section, we develop the asymptotics for Dickey-Fuller test for unit root, which is based on the discrete samples $(X_{i\delta})$, $i = 1, \ldots, n$, collected from the diffusion $X = (X_t)$ over the sample span $T = n\delta$. In what follows, we will simply write $x_i = X_{i\delta}$, $i = 1, \ldots, n$, with $x_0 = X_0$.

## 3.1 Primary Asymptotics of Unit Root Test

To test for a unit root in $(x_i)$, we may use the first-order autoregression without any augmented lags, since $X$ is a markov process. In particular, we consider the regression

$$\Delta x_i = \alpha + \beta x_{i-1} + \varepsilon_i, \tag{5}$$

where $\Delta$ is the usual difference operator, and test the null hypothesis $\beta = 0$ against the alternative hypothesis $\beta < 0$ using the least squares regression. The least square estimator and the $t$-statistic for $\beta$ in (5), denoted respectively as $\hat{\beta}$ and $t(\hat{\beta})$, are given by

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_{i-1} - \overline{x}_n)\Delta x_i}{\sum_{i=1}^n (x_{i-1} - \overline{x}_n)^2}$$

and

$$t(\hat{\beta}) = \frac{\hat{\beta}}{\hat{\sigma}\left(\sum_{i=1}^n (x_{i-1} - \overline{x}_n)^2\right)^{-1/2}},$$

where $\overline{x}_n$ is the sample mean of $(x_i)$ and $\hat{\sigma}^2$ is the usual estimator for the variance of regression errors $(\varepsilon_i)$.

In the usual discrete time setup, the Dickey-Fuller test based on the $t$-statistic $t(\hat{\beta})$ from regression (5) is widely used to test for the null hypothesis of a unit root, i.e., $\beta = 0$, against the alternative hypothesis of stationarity, i.e., $\beta < 0$. Under the null hypothesis of a unit root, it has nonstandard, yet well-defined, limit distribution, as long as some mild regularity conditions are satisfied for the innovations $(\varepsilon_i)$. The limit distribution, which we call the Dickey-Fuller distribution, is usually represented as a functional of Brownian motion. On the other hand, under the alternative hypothesis of stationarity, it diverges to negative infinity in probability. Consequently, it provides a test for unit root nonstationarity that is consistent against a wide class of stationary time series. Therefore, we may say, loosely yet generally, that a unit root time series is nonstationary, and that a stationary time series does not have a unit root.

The Dickey-Fuller test has also routinely been applied to samples that are thought to be collected discretely from diffusion type continuous time processes. However, little is known about its asymptotic behavior for discrete samples obtained from general continuous time processes. In particular, except for simple models such as Brownian motion and Ornstein-Uhlenbeck process, it has been completely unknown whether the test can be used to effectively discriminate nonstationary diffusions from stationary diffusions. To facilitate our discussions on the asymptotic behavior of the Dickey-Fuller test, we will simply say in what follows that a diffusion has a unit root if and only if $t(\hat{\beta})$ is stochastically bounded and the test is expected to suggest the non-rejection of the unit root hypothesis at least with some positive probability.

In our asymptotics, we require that the sampling interval $\delta$ be sufficiently small relative to the extremal bounds of various functional transforms of $X$ over time interval $[0, T]$.

Following Aït-Sahalia and Park (2016), we define

$$T(f) = \max_{0 \leq t \leq T} |f(X_t)|$$

for some function $f : \mathcal{D} \to \mathbb{R}$. For the identity function, we have $T(\iota) = \max_{0 \leq t \leq T} |X_t|$ and $T(\iota)$ becomes the asymptotic order of extremal process of $X$. It is obvious that we have $[T(\iota)]^k = T(\iota^k)$ for any nonnegative $k$. More generally, for any regularly varying $f$, we may obtain the exact rate of $T(f)$ from the asymptotic behavior of extremal process. For instance, the extremal processes of Ornstein-Uhlenbeck process and Feller's square root process are respectively of orders $O_p(\sqrt{\log T})$ and $O_p(\log T)$, and the extremal process of the general driftless diffusion process is of order $O_p(T)$. Thus if $f$ is regularly varying and $c_T$ is the order of extremal process, then we have $T(f) = f(c_T)$. The order of extremal process is known for a wide class of diffusions. For instance, under some regularity conditions on $\mu$ and $\sigma^2$, it is well known that the extremal processes of positive recurrent diffusions are of order $O_p(s^{-1}(T))$, to which the reader is referred to, e.g., Davis (1982). Moreover, asymptotic orders for the extremal processes of general null recurrent diffusions are obtained by Stone (1963), Jeong and Park (2013) and Kim and Park (2015).

In what follows, we let $\iota, \mu, \sigma$ be all majorized by $\omega : \mathcal{D} \to \mathbb{R}$.

**Assumption 3.1.** $\delta T(\omega^4) T^2 \log(T/\delta) \to_p 0$.

Assumption 3.1 makes it necessary to have $\delta \to 0$. If we fix $T$, $\delta \to 0$ is indeed the necessary and sufficient condition for Assumption 3.1. Clearly, we may also allow $T \to \infty$ as long as $\delta \to 0$ sufficiently fast. In this case, our asymptotic results will be more relevant for the case where $\delta$ is sufficiently small relative to $T$. Our asymptotics in the paper are derived under the condition $\delta \to 0$ and $T \to \infty$ jointly. For Assumption 3.1 to hold, it suffices to have $\delta = O(T^{-2-\epsilon})$ for any $\epsilon > 0$, if $X$ is bounded so that $T(\omega^2)$ is a constant. The condition appears to be mild enough to yield asymptotics generally relevant for a very wide range of empirical analysis relying on samples collected from diffusion models. For daily observations over ten years, as an example, we have $\delta \approx 1/250$ and $T^{-2} = 1/100$. Our subsequent asymptotics hold jointly in $\delta$ and $T$ as long as they satisfy Assumption 3.1 as $\delta \to 0$ and $T \to \infty$. In particular, we do not use sequential asymptotics, requiring $\delta \to 0$ and $T \to \infty$ sequentially.

The primary asymptotics for $\hat{\beta}$ and $t(\hat{\beta})$ are given by the following lemma. Here and elsewhere in the paper, we let $\overline{X} = (\overline{X}_t)$, $\overline{X}_t = t^{-1} \int_0^t X_s ds$, be the sample mean process of $X$.

**Lemma 3.1.** *Let Assumption 3.1 hold. The we have*

$$\hat{\beta} \sim_p \frac{\delta \int_0^T (X_t - \overline{X}_T) dX_t}{\int_0^T (X_t - \overline{X}_T)^2 dt},$$

$$t(\hat{\beta}) \sim_p \frac{\sqrt{T} \int_0^T (X_t - \overline{X}_T) dX_t}{[X]_T^{1/2} \left( \int_0^T (X_t - \overline{X}_T)^2 dt \right)^{1/2}}$$

*for all $\delta$ sufficiently small relative to $T$.*

The limit theory in Lemma 3.1 holds for all small enough $\delta$ relative to $T$ as long as $\delta$ and $T$ satisfy Assumption 3.1. Therefore, we expect that they provide good approximations for finite sample distributions, whenever $\delta$ is relatively small compared with $T$. Note that we do not assume $T = \infty$ to obtain the asymptotics in Lemma 3.1. The asymptotics we have in Lemma 3.1 will be referred in the paper to as the primary asymptotics. The joint asymptotics for $\delta \to 0$ and $T \to \infty$, which are presented below, may be obtained simply by taking $T$-limits to our primary asymptotics.

Our primary asymptotics in Lemma 3.1 make it clear that we have $\hat{\beta} \to_p 0$ whenever $\delta \to 0$ fast enough relative to $T$. If, in particular, $T$ is fixed, we have $\hat{\beta} \to_p 0$ for any diffusion $X$ as long as $\delta \to 0$. For discrete samples from any diffusion, we will therefore always observe a root getting close to unity as we collect samples frequently enough and $\delta$ becomes sufficiently small. However, even in this case, the unit root test will not necessarily support the presence of a unit root. If we let $\delta \to 0$ with fixed $T$, $t(\hat{\beta})$ remains to be random and the unit root test will yield completely random conclusion, regardless of the asymptotic properties of the underlying diffusion. The unit root test will be totally uninformative in this case. Indeed, it is very natural and not surprising at all that the unit root test does not yield any information on the asymptotics, such as mean reversion and unit root properties, of the underlying diffusion unless $T$ becomes large. In discussions to follow, we let $T \to \infty$ and consider large $T$ asymptotics.

If used to test for nonstationarity of the underlying diffusion, the unit root test is consistent against all mean-reverting diffusions. However, the test becomes consistent and has discriminating power only when the time span $T$ goes to infinity. In particular, it is well expected from our theory that the test will not be consistent if we just increase the sample size $n = T/\delta$ by decreasing the sampling interval $\delta$ (or equivalently, increasing sampling frequency $1/\delta$) with fixed $T$. This was first observed in Shiller and Perron (1985), and further analyzed later by Perron (1989, 1991), for the test of a unit root in samples from Brownian motion against those from Ornstein-Uhlenbeck process.

### 3.2 Asymptotics of Unit Root Test

Now we establish the asymptotics for the unit root test. To effectively present our asymptotics, We let $(\lambda_T)$ be the normalizing sequence satisfying

$$T = \lambda_T [m_s](\lambda_T) \quad \text{or} \quad \lambda_T^2 m_s(\lambda_T) \tag{6}$$

depending upon whether or not ST holds. In case either ST or DD holds, we subsequently define

$$a_T = \begin{cases} \lambda_T [m_s \sigma_s^2](\lambda_T) & \\ & \text{if} \\ \lambda_T^2 (m_s \sigma_s^2)(\lambda_T) & \end{cases} \begin{array}{l} \text{DD holds} \\ \\ \text{ST holds and DD does not hold} \end{array}$$

$$b_T = \begin{cases} \lambda_T [m_s \iota_s^2](\lambda_T) & \\ & \text{if} \\ \lambda_T^2 (m_s \iota_s^2)(\lambda_T) & \end{cases} \begin{array}{l} \text{SI holds} \\ \\ \text{ST holds and SI does not hold} \end{array}$$

from $(\lambda_T)$, and let

$$P = \begin{cases} L(\tau, 0) & \\ & \text{if} \\ \int_0^\tau \overline{m_s \sigma_s^2}(B_t)dt & \end{cases} \begin{array}{l} \text{DD holds} \\ \\ \text{ST holds and DD does not hold} \end{array}$$

$$Q = \begin{cases} L(\tau, 0) \left[ 1 - \big(m(\iota)\big)^2 / m(\iota^2) \right] & \\ & \text{if} \\ \int_0^\tau \overline{m_s \iota_s^2}(B_t)dt & \end{cases} \begin{array}{l} \text{SI holds} \\ \\ \text{ST holds and SI does not hold,} \end{array}$$

where $\tau$ is a stopping time defined as

$$\tau = \inf \left\{ t \,\middle|\, L(t, 0) > 1 \right\} \quad \text{or} \quad \inf \left\{ t \,\middle|\, \int_{\mathbb{R}} L(t, x) \overline{m}_s(dx) > 1 \right\}, \tag{7}$$

depending upon whether or not ST holds, from the local time $L$ of Brownian motion $B$, and $\overline{f}$ denotes the limit homogeneous function of regularly varying $f$ on $\mathbb{R}$. Numerical sequences $(a_T)$ and $(b_T)$ and random variables $P$ and $Q$ introduced here will be used repeatedly in what follows.

**Lemma 3.2.** *Let Assumptions 2.1 and 2.2 hold. If either* ST *or* DD *holds, we have*

$$\frac{1}{a_T}[X]_T \to_d P, \quad \frac{1}{a_T} \int_0^T (X_t - \overline{X}_T) dX_t \to_d -\frac{P}{2}$$

*and*

$$\frac{1}{b_T} \int_0^T (X_t - \overline{X}_T)^2 dt \to_d Q$$

*as* $T \to \infty$*, and* $Ta_T/b_T \to \infty$ *as* $T \to \infty$*.*

If neither ST nor DD holds, we would have a quite different asymptotics. Let $Y = s(X)$ be the scale transformation of $X$ and define $Y^T$ by $Y_t^T = \lambda_T^{-1} Y_{Tt}$ with the normalizing sequence $(\lambda_T)$ in (6), we have $Y^T \to_d Y^\circ$ as $T \to \infty$ in the space $C[0,1]$ of continuous functions with uniform topology, where using Brownian motion $B$ and its local time $L$ we may represent the limit process $Y^\circ$ as

$$Y_t^\circ = B \circ \overline{A}_t \quad \text{with} \quad \overline{A}_t = \inf\left\{ s \,\middle|\, \int_{\mathbb{R}} L(s,x)\overline{m}_s(dx) > t \right\}. \tag{8}$$

To obtain the asymptotics for a general diffusion $X$, we write it as $X = s^{-1}(Y)$. Note that if $X$ does not satisfy DD, then $s^{-1} \in RV_p$ with $p = 1/(q+1) > 1/2$ since $s' \in RV_q$ with $q < 1$. Therefore, if we define $X^T$ as $X_t^T = X_{Tt}/s^{-1}(\lambda_T) = s^{-1}(Y_{Tt})/s^{-1}(\lambda_T)$, we may well expect that

$$X^T = \frac{s^{-1}(\lambda_T Y^T)}{s^{-1}(\lambda_T)} \to_d \overline{s^{-1}}(Y^\circ) = X^\circ,$$

in $C[0,1]$ as $T \to \infty$.

**Lemma 3.3.** *Let Assumptions 2.1 and 2.2 hold. If neither* ST *nor* DD *holds, we have* $\left(s^{-1}(\lambda_T)\right)^{-2}[X]_T \to_d [X^\circ]_1$ *and*

$$\frac{1}{\left(s^{-1}(\lambda_T)\right)^2}\int_0^T (X_t - \overline{X}_T)dX_t \to_d \int_0^1 (X_t^\circ - \overline{X}_1^\circ)dX_t^\circ$$
$$\frac{1}{T\left(s^{-1}(\lambda_T)\right)^2}\int_0^T (X_t - \overline{X}_T)^2 dt \to_d \int_0^1 (X_t^\circ - \overline{X}_1^\circ)^2 dt$$

*with* $\overline{X}_1^\circ = \int_0^1 X_t^\circ dt$*, as* $T \to \infty$*.*

The asymptotics for unit root test follow immediately from Lemmas 3.1, 3.2 and 3.3.

**Theorem 3.4.** *Let Assumptions 2.1, 2.2 and 3.1 hold. If neither* ST *nor* DD *holds, we have*

$$n\hat{\beta} \to_d \frac{\int_0^1 (X_t^\circ - \overline{X}_1^\circ)dX_t^\circ}{\int_0^1 (X_t^\circ - \overline{X}_1^\circ)^2 dt},$$
$$t(\hat{\beta}) \to_d \frac{\int_0^1 (X_t^\circ - \overline{X}_1^\circ)dX_t^\circ}{[X^\circ]_1^{1/2}\left(\int_0^1 (X_t^\circ - \overline{X}_1^\circ)^2 dt\right)^{1/2}}$$

*as $\delta \to 0$ and $T \to \infty$. On the other hand, if either* ST *or* DD *holds, we have*

$$\frac{b_T}{Ta_T} n\hat{\beta} \to_d -\frac{P}{2Q} \quad \textit{and} \quad \sqrt{\frac{b_T}{Ta_T}}\, t(\hat{\beta}) \to_d -\frac{1}{2}\sqrt{\frac{P}{Q}}$$

*as $\delta \to 0$ and $T \to \infty$, and $Ta_T/b_T \to \infty$ as $T \to \infty$.*

Theorem 3.4 shows that the tests based on $n\hat{\beta}$ and $t(\hat{\beta})$, two most commonly used statistics to test for a unit root, have full asymptotic discriminatory powers for the null and alternative hypotheses, $\mathrm{H}_0$: neither ST nor DD holds, and $\mathrm{H}_1$: either ST or DD holds. Under $\mathrm{H}_0$, neither ST nor DD holds, and $n\hat{\beta}$ and $t(\hat{\beta})$ have nondegenerate limit distributions. On the other hand, under $\mathrm{H}_1$, either ST or DD holds, $n\hat{\beta} \sim_d (Ta_T/b_T)(-P/2Q) \to_p -\infty$ and $t(\hat{\beta}) \sim_d (Ta_T/b_T)^{1/2} - (1/2)(P/Q)^{1/2}) \to_p -\infty$, since $P, Q > 0$ with probability one and $Ta_T/b_T \to \infty$. If under $\mathrm{H}_1$ both DD and SI hold, we have

$$P/Q = 1 \Big/ \Big[1 - \big(m(\iota)\big)^2/m(\iota^2)\Big] .$$

Note that our limit theory in Theorem 3.4 is completely general and holds for a truly broad class of recurrent diffusions. In particular, we do not impose any assumptions on serial dependence or existence of moments. We only rely on some basic regularity conditions in Assumptions 2.1 and 2.2.

It is clear from Theorem 3.4 that the unit root test cannot be used to test for nonstationarity. Null recurrent diffusion may not have a unit root if it satisfies DD and has a dominating drift. As an illustrative example, we consider the GHK model in (4) with $a = -7/5$ and $b = -1$. With the given set of parameter values, $X$ has a dominating drift though it is nonstationary, i.e., DD holds though ST is not satisfied. Therefore, the unit root test is expected to reject the unit root null hypothesis. To demonstrate this, we compute the actual rejection probabilities of the unit root test and report them in Table 1. The reported rejection probabilities are obtained by simulating 10,000 realizations of the diffusion with $\delta = 1/252$ and $T = 10, 20, 40, 80$, which correspond to daily observations for the periods of 10, 20, 40, 80 years, respectively. We use the 5% critical value from the Dickey-Fuller distribution. As expected, the actual rejection probabilities exceed the nominal level for all $T$, and they increase with $T$.

We should also note that the unit root test does not have any nontrivial power in discriminating diffusions with and without drift. A recurrent diffusion with linear drift is positive recurrent. Therefore, as long as it is recurrent, any diffusion is stationary if it has a linear drift. This, in turn, implies that the unit root test rejects the unit root hypothesis for any recurrent diffusion with linear drift. However, in general, the unit root test does not

Table 1: Rejection probabilities of Dickey-Fuller test for nonstationary diffusion having dominating drift

| $T$ | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| Rejection Probabilities | 10.1% | 14.2% | 20.3% | 29.9% |

have any discriminatory power for or against the presence of drift in diffusion. It is clear that for a driftless diffusion may also be mean-reverting, and has no unit root, as long as it is stationary or barely nonstationary.

Finally, we note that the unit root test becomes a test for nonstationarity if $X$ is a natural scale diffusion. Therefore, the test applied to the scale transformed version $s(X)$ of $X$ becomes a consistent test for nonstationarity of $X$.

# 4 Unit Root and Mean Reversion

## 4.1 Mean Reversion

If applied to samples from diffusion models, the standard unit root tests become tests for mean reversion. To show this, we define

**Definition 4.1.** We say that $X$ has *mean reversion* if and only if

$$\frac{1}{c_T}\int_0^T (X_t - \overline{X}_T)dX_t \to_d Z$$

as $T \to \infty$, for some normalizing sequence $(c_T)$ and a random variable $Z$ with support on a subset of $(-\infty, 0)$.

For $X$ with mean reversion as defined in Definition 4.1, we say that it has *strong* mean reversion if $Z$ has a point support, and *weak* mean reversion otherwise.

The motivation for our definition of mean reversion is clear. We have

$$M_T = \int_0^T (X_t - \overline{X}_T)dX_t \approx \sum_{i=1}^m (X_{t_{i-1}} - \overline{X}_T)(X_{t_i} - X_{t_{i-1}})$$

for $0 = t_0 < \cdots < t_m = T$ with $\max_{1\leq i\leq m} |t_i - t_{i-1}| \approx 0$. Therefore, negative $M_T$ implies that $(X_{t_{i-1}} - \overline{X}_T)$ has a negative sample correlation with $(X_{t_i} - X_{t_{i-1}})$, i.e., the deviation from sample mean in the current period is negatively correlated with the increment made in transition to the next period. This occurs if and only if whenever $X$ is observed below its sample mean it has tendency to increase, and vice versa. In describing the mean-reverting behavior of diffusion, we may use the recursive mean and define $(X_t - \overline{X}_t)$ as the deviation from mean, in place of $(X_t - \overline{X}_T)$ relying on the sample mean $\overline{X}_T$ over the entire time span. Though we do not show explicitly in the paper, all our subsequent theories are also applicable for this alternative definition of mean reversion only with some obvious minor changes.

Due to Lemmas 3.2 and 3.3, it follows straightforwardly from Definition 4.1 that

**Lemma 4.1.** *Let Assumptions 2.1 and 2.2 hold. Then $X$ has mean reversion if and only if either* ST *or* DD *holds.*

In case both ST and DD hold, we have

$$\frac{1}{T}\int_0^T (X_t - \overline{X}_T)dX_t \to_d -\frac{1}{2}\pi(\sigma^2)$$

as $T \to \infty$, and therefore, $X$ has strong mean reversion. On the other hand, if only one of ST and DD holds, $X$ has weak mean reversion.

The following lemma shows the reason why we have mean reversion if either ST or DD holds.

**Lemma 4.2.** *Let Assumptions 2.1 and 2.2 hold. Then* ST *or* DD *holds if and only if*

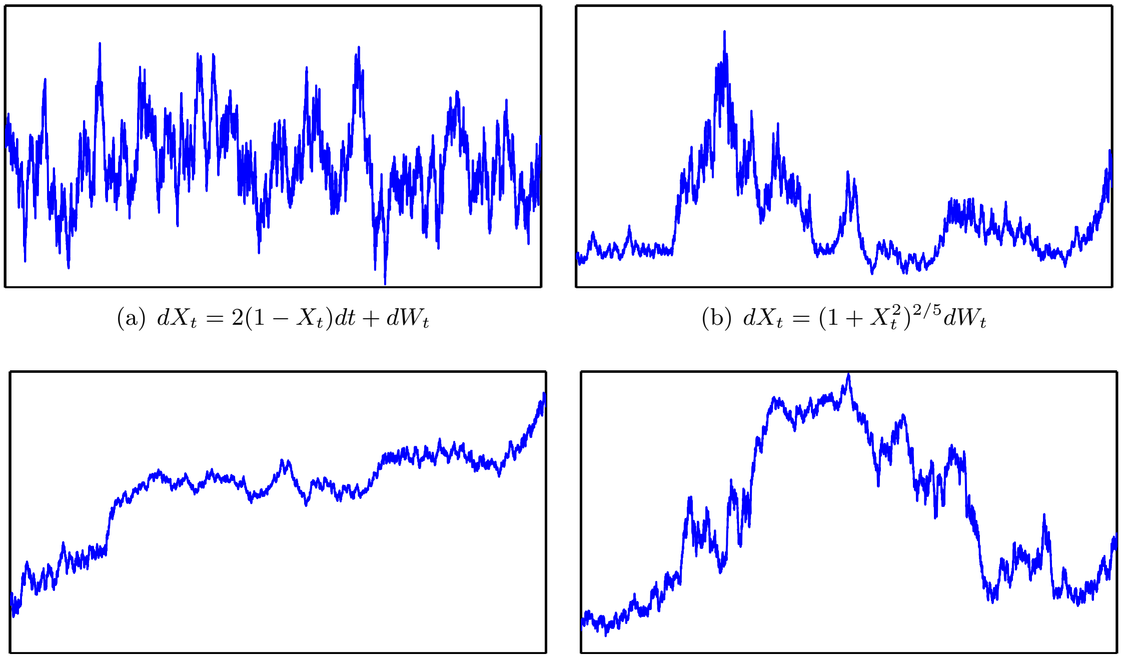
$$\int_0^T (X_t - \overline{X}_T)dX_t \sim_p -\frac{1}{2}\int_0^T \sigma^2(X_t)dt$$

*as $T \to \infty$.*

Note that we have $(X_t - \overline{X}_T)dX_t = (1/2)\big(d(X_t - \overline{X}_T)^2 - d[X]_t\big)$, due to Ito's formula, and $d[X]_t = \sigma^2(X_t)dt$. Therefore, the only if part of Lemma 4.2 follows straightforwardly, if ST holds with stationary $X$ having all required moments, since in this case $X_T, X_0$ and $\overline{X}_T$ are all of $O_p(1)$, and $(X_T - \overline{X}_T)^2$ and $(X_0 - \overline{X}_T)^2$ become negligible in the limit compared to $[X]_T = \int_0^T \sigma^2(X_t)dt$. In fact, Lemma 4.2 shows that $(X_T - \overline{X}_T)^2$ and $(X_0 - \overline{X}_T)^2$ are asymptotically negligible if and only if either ST or DD holds.

**Example 4.1.** In Figure 1, we provide the simulated sample paths of four different diffusions with $\delta = 1/252$ and $T = 40$, which correspond to daily observations for 40 years. The sample path of stationary Ornstein-Uhlenbeck process is in Part (a). It is a stationary process with strong mean reversion, satisfying both ST and DD. Parts (b) and (c) present the sample paths of driftless diffusions. The driftless diffusion in Part (b) satisfies ST but not DD, and becomes a stationary process with weak mean reversion. On the other hand, the driftless diffusion in Part (c) satisfies neither ST nor DD, which implies that it is a

Figure 1: Sample Paths of Various Diffusions



(a) $dX_t = 2(1 - X_t)dt + dW_t$

(b) $dX_t = (1 + X_t^2)^{2/5} dW_t$

(c) $dX_t = (1 + X_t^2)^{1/5} dW_t$,

(d) $dX_t = -(7/5)X_t(1+X_t^2)^{-2}dt+(1+X_t^2)^{-1/2}dW_t$

nonstationary process with no mean reversion. Finally, Part (d) presents the sample path of the GHK model with parameter values $a = -1.4$ and $b = -1$. It satisfies DD but not ST, and provides an example of a nonstationary process with mean reversion.

### 4.2 Testing for Mean Reversion

Now we reformulate the null and alternative hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$, introduced in Section 3.2 respectively as, $\mathrm{H}_0$: no mean reversion and $\mathrm{H}_1$: mean reversion, in $X$. Note that, for the reformulated null and alternative hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$, the unit root tests based on $n\hat{\beta}$ and $t(\hat{\beta})$ become consistent. Therefore, we may use them to test for the absence of mean reversion against the presence of mean reversion. Unfortunately, however, their null limit distributions are heavily model-dependent. In general, the null limit distributions of the unit root tests based on $n\hat{\beta}$ and $t(\hat{\beta})$ are different from the Dickey-Fuller distributions, and therefore, the standard critical values are not applicable. Of course, we may use the standard critical values in case that $X$ becomes Brownian motion asymptotically. Clearly, $X^\circ$ becomes Brownian motion if and only if $\overline{\sigma^2}$ and $\overline{s'}$ are constant functions. Therefore, if we let $s' \in RV_p$ and $m \in RV_q$, then we should have $p = 0$ and $q = 0$ at both boundaries.

Moreover, it is required that $\sigma^2(-\lambda)/\sigma^2(\lambda) \to 1$, and

$$\frac{s'(-\lambda)}{s'(\lambda)} = \exp\left(-\int_{|x|\leq\lambda} \frac{2\mu(x)}{\sigma^2(x)}dx\right) \to 1,$$

which holds if and only if

$$\int_{|x|\leq\lambda} \frac{\mu(x)}{\sigma^2(x)}dx \to 0,$$

as $\lambda \to \infty$. For instance, the diffusion defined as

$$dX_t = \frac{X_t}{1+X_t^2}dt + \left(1 + \log^{1/2}(1+|X_t|)\right)dW_t$$

becomes Brownian motion in the limit.

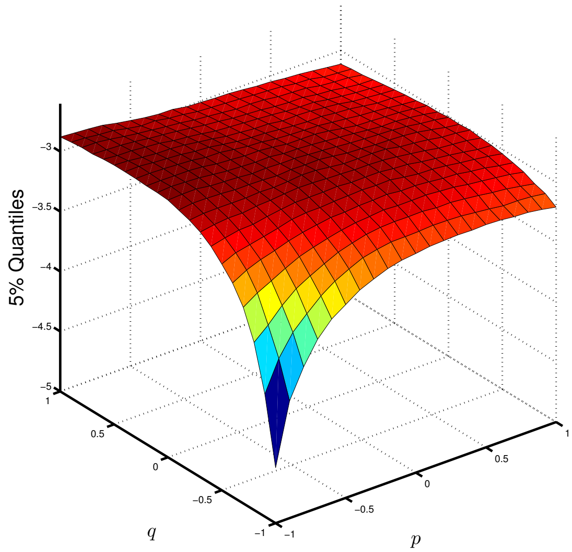
**Example 4.2.** Let $s' \in RV_p$ and $m \in RV_q$. For the GHK diffusion in Example 2.1, we have $p = -2a$ and $q = 2a - 2b$. As discussed, different choice of $(p, q)$ gives different limiting distributions. We conduct simulations for the process (4) with various combinations of $(p, q) \in [-1, 1] \times [-1, 1]$ to examine the model dependency in null distributions. For each experiment, we simulate 10,000 realizations with $\delta = 1/252$ and $T = 40$ which correspond to daily data for 40 years. Figure 2 shows the 5% quantiles of $t$-statistics. There appears to be considerable amount of model dependency in the finite sample distribution. As we can see in Figure 2, the critical values from the Dickey-Fuller distributions may give misleading conclusions. In particular, if both $p$ and $q$ are close to, but not less than, -1, then 5% quantiles of $t$-statistics from (4) are significantly smaller than the critical value from the Dickey-Fuller distributions. For example, if $p = -0.9$ and $q = -0.9$, then the 5% quantile of $t$-statistics from (4) is about $-3.7$, whereas corresponding critical value from the Dickey-Fuller distributions is about $-2.9$.

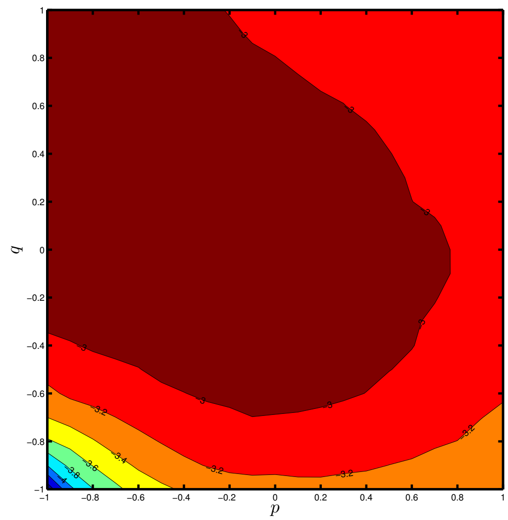
Under the presence of mean reversion, the unit root tests based on $n\hat{\beta}$ and $t(\hat{\beta})$ diverge up to $-\infty$ as $T \to \infty$. Their divergence rates are given by $Ta_T/b_T$ and $\sqrt{Ta_T/b_T}$, respectively. If $s' \in RV_p$ and $m \in RV_q$, then (i) ST holds if and only if $q \leq -1$, (ii) DD holds if and only if $p \geq 1$, and (iii) SI holds if and only if $q \leq -3$. Moreover, if $p \neq -1$, then $s^{-1}$ becomes regularly varying. In this case, it follows from Karamata's theorem that for some slowly varying function $\ell$

$$a_T \sim \begin{cases} \lambda_T \ell(\lambda_T) & \text{if DD holds} \\ \lambda_T^{2/(p+1)}\ell(\lambda_T) & \text{otherwise,} \end{cases} \qquad b_T \sim \begin{cases} \lambda_T \ell(\lambda_T) & \text{if SI holds} \\ \lambda_T^{(p+q+4)/(p+1)}\ell(\lambda_T) & \text{otherwise,} \end{cases}$$

where $\lambda_T \sim T\ell(T)$ or $T^{(p+1)/(p+q+2)}\ell(T)$ depending upon whether or not ST holds, and

Figure 2: Asymptotic Critical Values of Unit Root Test for GHK Diffusions




Notes: The data were generated according to the GHK model (4). The left plot presents 5% quantiles of $t$-statistics for $(p,q) \in [-1,1] \times [-1,1]$, and its corresponding Contour plot is on the right.

therefore, we can easily obtain the rates of divergence as shown in Table 2. It is easy to see that $Ta_T/b_T \prec T$ for a nonstationary diffusion satisfying DD.

Table 2: The rates $Ta_T/b_T$ for $X$ having $s' \in RV_p$ and $m \in RV_q$ with $p \neq -1$.

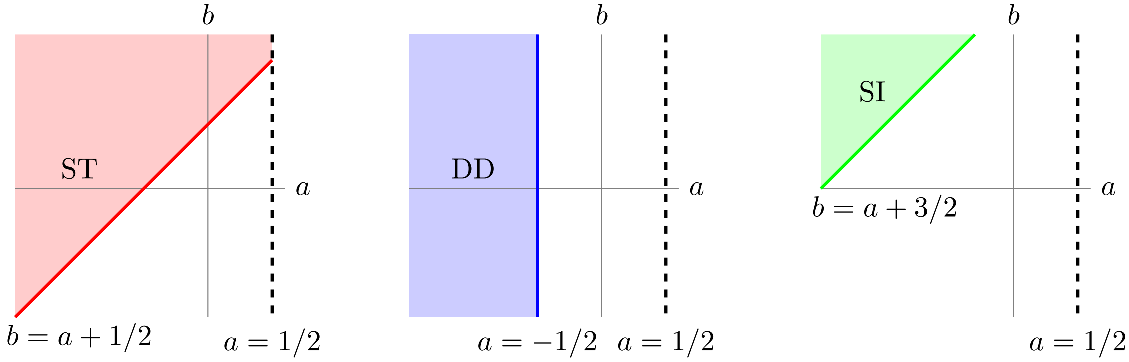
| | ST and SI | ST and NSI | NST |
|---|---|---|---|
| DD | $T\ell(T)$ | $T^{(p-q+2)/(p+1)}\ell(T)$ | $T^{(p-1)/(p+q+2)}\ell(T)$ |
| NDD | $T^{2/(p+1)}\ell(T)$ | $T^{(-q-1)/(p+1)}\ell(T)$ | 1 |

**Remark 4.1.** In the paper, we mainly consider stock variables, for which we assume that the discrete samples $(x_i)$ are obtained from the underlying diffusion at equally spaced time intervals and let $x_i = X_{i\delta}$. However, we may also analyze the discrete samples for flow variables, which are given as

$$x_i = \int_{(i-1)\delta}^{i\delta} X_t dt$$

from the underlying diffusion $X$. The asymptotics of the unit root test for the discrete samples from continuous time flow variables is investigated by Chambers (2004, 2008), under the assumption that they are generated by either Brownian motion or Ornstein-Uhlenbeck process. For some related simulation studies of the unit root test for the samples from continuous time processes, the reader is referred to Choi (1992), Choi and Chung

Figure 3: Mean Reversion Characteristics of GHK Diffusions




(1995) and Boswijk and Klaassen (2012).

## 5 Conclusion

In this paper, we study the longrun behaviors of general diffusion models and find many interesting results. In particular, we show that the Dickey-Fuller statistic cannot be used to test for nonstationarity of the underlying diffusion. Rather, it becomes a test for no mean reversion of the underlying diffusion, if applied to discrete samples from a diffusion. A diffusion has a unit root in its discrete samples if and only if it does not have mean reversion, and a nonstationary diffusion may also be mean-reverting if it has a drift dominating its diffusion. The test would at least have the perfect discriminatory power against all stationary diffusions, if we only consider nonstationary and non-mean-reverting diffusions under the null hypothesis. However, even in this case, the limit distribution of the test becomes model-dependent and the usual Dickey-Fuller critical values are not applicable. For the unit root test applied to discrete samples from a diffusion model to be strictly valid as a test for nonstationarity of the underlying diffusion, therefore, we should further restrict the class of diffusion models considered under the null hypothesis to be nonstationary diffusions that become Brownian motions in the limit.

# Appendix A Useful Lemmas

In the following, we assume that $X$ is recurrent diffusion satisfying Assumption 2.1 and 2.2. For a recurrent diffusion $X$, we define its natural scale diffusion $s(X)$ by $Y$, and denote $m_s$ by the speed measure of $Y$, which is given by $m_s = 1/(s'\sigma)^2 \circ s^{-1}$. We assume that $m_s$ and $s^{-1}$ have $+\infty$ as their dominating boundary, and the reader is referred to Section 4 of Kim and Park (2015) for more discussion about the dominating boundary and asymptotics of diffusion functionals. Moreover, we let $s' \in RV_p$ with $p \neq -1$ due to Assumption 2.2.

For $\lambda_T$ in (6), we note that $\lambda_T \sim T/m(\mathcal{D})$ for a stationary $X$, whereas $\lambda_T = o(T)$ for a nonstationary $X$. We define $\nu_T$ by $s^{-1}(\lambda_T)$. It then follows from the recurrence property, we have $\nu_T \to \infty$ as $\lambda_T \to \infty$. For a brevity of notations, we sometimes write I, BN and SN instead of integrable, barely nonintegrable and strongly nonintegrable, respectively. For instance, we write $f$ is BN when it is barely nonintegrable.

Finally, some of our subsequent asymptotics rely on various properties of regularly varying function. Among them, we frequently use Karamata's theorem (see, e.g., Bingham et al. (1993)) given by

$$\frac{\lambda f(\lambda)}{\int_\omega^\lambda f(x)dx} \to \kappa + 1$$

as $\lambda \to \infty$, for a nonnegative regularly varying function $f$ defined on $(0, \infty)$ with index $-1 \leq \kappa < \infty$, and for any $\omega > 0$. Moreover, Karamata's theorem implies that $\int_\omega^x f(y)dy$ becomes a regularly varying function with index $\kappa + 1$ for any $\omega > 0$.

**Lemma A1.** *Let $f$ be regularly varying on $\mathcal{D}$. If $f$ is $m$-regularly nonintegrable, then $s^{-1} \in RV_\kappa$ with $0 < \kappa < \infty$ at $+\infty$. Moreover, if $f$ is $m$-barely nonintegrable/strongly nonintegrable on $\mathcal{D}$, then $f_s$ is $m_s$-barely nonintegrable/strongly nonintegrable on $\mathbb{R}$.*

*Proof of Lemma A1.* By our construction, $f$ is $m$-regularly nonintegrable means that $mf = f/(s'\sigma^2)$ is regularly nonintegrable. Then $s'$ becomes regularly varying. If not and $s'$ is rapidly varying, then $s'$ is rapidly increasing as $x \to x_B$ due to the recurrence property, and hence, $f$ is $m$-integrable since $\sigma^2$ is regularly varying by Assumption 2.2. Therefore, if $f$ is $m$-regularly nonintegrable, then $s'$ is regularly varying on $\mathcal{D}$ with index $-1 < p < \infty$ at $x_B = \pm\infty$ and $-\infty < p < -1$ at $x_B = 0$ due to Assumption 2.2 and the recurrence conditions. Then, by the property of regularly varying function (see, e.g., Proposition 1.5.7. in Bingham et al. (1993)), $s^{-1} \in RV_\kappa$ with $0 < \kappa = 1/(p+1) < \infty$ at $+\infty$ since $-1 < p < \infty$ in this case, which completes the first part of the proof.

For the second part, without loss of generality, we let $\sigma^2$ and $f$ be regularly varying on $\mathcal{D}$ with indexes $q$ and $r$, respectively. Then $mf$ is regularly varying on $\mathcal{D}$ with index

$r-p-q$, whereas $m_s f_s$ is regularly varying on $\mathbb{R}$ with index $(r-2p-q)/(p+1)$ since $f_s m_s = (f/(s'\sigma)^2)\circ s^{-1}$. It then easy to see that $(r-2p-q)/(p+1) \lesseqgtr -1$ if and only if $r-p-q \lesseqgtr -1$ at $x_B = \pm\infty$ and $r-p-q \gtreqless -1$ at $x_B = 0$. Then the stated result follows immediately from the representation theorem for regularly varying functions (see, e.g., pages 21-22 in Bingham et al. (1993)). $\square$

**Lemma A2.** *(a) If $f$ and $g^2$ are $m$-integrable, then*

$$\frac{1}{\lambda_T}\int_0^T f(X_t)dt \to_d L(\tau,0),$$
$$\frac{1}{\sqrt{\lambda_T}}\int_0^T g(X_t)dW_t \to_d \sqrt{L(\tau,0)}N$$

*jointly, where $N$ is a standard normal random variate independent of $L(\tau,0)$.*

*(b) If $f$ and $g^2$ are $m$-strongly nonintegrable, then*

$$\frac{1}{\lambda_T^2(m_s f_s)(\lambda_T)}\int_0^T f(X_t)dt \to_d \int_0^\tau \overline{m_s f_s}(B_t)dt,$$
$$\frac{1}{\lambda_T(m_s^{1/2} g_s)(\lambda_T)}\int_0^T g(X_t)dW_t \to_d \int_0^\tau \overline{m_s^{1/2} g_s}(B_t)dB_t$$

*jointly.*

*(c) If $f$ and $g^2$ are $m$-barely nonintegrable, then*

$$\int_0^T f(X_t)dt \sim_d \lambda_T[m_s f_s](\lambda_T)L(\tau,0) + \lambda_T^2(m_s f_s)(\lambda_T)\lim_{\varepsilon\to 0}\int_0^\tau \frac{1}{B_t}1\{|B_t|>\varepsilon\}dt,$$

*or*

$$\frac{1}{\lambda_T[m_s f_s](\lambda_T)}\int_0^T f(X_t)dt \to_d L(\tau,0),$$

*depending upon whether $\overline{m_s f_s}(x) = 1/x$ or $\overline{m_s f_s}(x) \neq 1/x$, and*

$$\frac{1}{\sqrt{\lambda_T[m_s g_s^2](\lambda_T)}}\int_0^T g(X_t)dW_t \to_d \sqrt{L(\tau,0)}N$$

*jointly, where $N$ is a standard normal random variate independent of $L(\tau,0)$.*

*(d) If $f$ is $m$-barely nonintegrable, then*

$$\int_0^T f(X_t)dt = O_p(\lambda_T[m_s|f_s|](\lambda_T)).$$

*(e) If ST holds, then $L(\tau,0)=1$ with probability one.*

*Proof for Lemma A2.* Due to Lemma A1, the results (a), (b) and (c) follows respectively from Theorems 3.3, 3.4 and 3.5 of Kim and Park (2015).

As for (d), we note that if $f$ is $m$-barely nonintegrable, then $|f|$ is $m$-barely nonintegrable and $\overline{m_s|f_s|}\neq 1/x$. It then follows from the part (c) of this lemma that

$$\frac{1}{\lambda_T[m_s|f_s|](\lambda_T)}\int_0^T f(X_t)dt \leq \frac{1}{\lambda_T[m_s|f_s|](\lambda_T)}\int_0^T |f|(X_t)dt \to_d L(\tau,0)$$

which completes the proof of part (d). Finally, the part (e) follows immediately from the construction of $\tau$. □

**Lemma A3.** *Let $X$ be null recurrent, and define $X^T$ by $X_t^T=\nu_T^{-1}X_{Tt}$ for $0\leq t\leq 1$. Then we have the followings.*

*(a) We have $X^T\to_d X^\circ$, where $X^\circ=\overline{s^{-1}}(Y^\circ)$ for a strongly nonstationary $X$, and $X^\circ=0$ a.s. for a barely nonstationary $X$. Here $Y^\circ$ is defined in (8).*

*(b) If $X$ is strongly nonstationary, then $X^\circ$ becomes a semimartingale with quadratic variation $[X^\circ]$ is given by $[X^\circ]_t=\int_0^{\overline{A}_t}\left((\overline{s^{-1}})'_-\right)^2(B_s)ds$, where $(\overline{s^{-1}})'_-$ is the left-hand derivative of $\overline{s^{-1}}$, and $\overline{A}_t=\inf\left\{s|\int_{\mathbb{R}}\overline{m}_s(x)L(s,x)>t\right\}$.*

*Proof for Lemma A3 (a).* If $X$ is null recurrent, then $m$ is regularly nonintegrable on $\mathcal{D}$ due to Assumption 2.2. It then follows from Lemma A1 with $f=1$ that $m_s$ is regularly nonintegrable. In particular. $m_s$ is not integrable at $+\infty$ since $m_s$ has $+\infty$ as its dominating boundary. Moreover, we have $s^{-1}\in RV_\kappa$ with $\kappa>0$ due to Lemma A1, and hence,

$$X_t^T=\frac{s^{-1}(\lambda_T Y_{Tt}/\lambda_T)}{s^{-1}(\lambda_T)}=\overline{s^{-1}}(Y_{Tt}/\lambda_T)(1+o_p(1))\to_d \overline{s^{-1}}(Y_t^\circ)$$

by Proposition 3.2 of Kim and Park (2015), the continuous mapping theorem and the uniform convergence of regularly varying functions (see, e.g., pages 21-22 in Bingham et al. (1993)). This completes the proof. □

*Proof for Lemma A3 (b).* Note that $\overline{s^{-1}}$ is nondecreasing and can be represented as the difference of two convex functions. As in Kim and Park (2015), we may apply Itô-Tanaka formula to $Y^\circ=B\circ\overline{A}$ to deduce

$$\begin{aligned}X_t^\circ-X_0^\circ&=\overline{s^{-1}}(B\circ\overline{A}_t)-\overline{s^{-1}}(B\circ\overline{A}_0)\\&=\int_0^t(\overline{s^{-1}})'_-(B\circ\overline{A}_s)d(B\circ\overline{A}_s)+\frac{1}{2}\int_{\mathbb{R}}L(\overline{A}_t,x)(\overline{s^{-1}})''(dx),\end{aligned}\tag{A.1}$$

where $(\overline{s^{-1}})''$ is the second derivative of $\overline{s^{-1}}$ in the sense of distributions.

In (A.1), the first and second terms represent respectively the martingale and bounded variation components of a semimartingale $X^\circ$, for which we have $[X^\circ]_t = \int_0^{\overline{A}_t} \left( (\overline{s^{-1}})'_- \right)^2 (B_s)ds$. This completes the proof. □

**Lemma A4.** *If ST holds, we have $X_T = o_p(\nu_T)$ and $\overline{X}_T = o_p(\nu_T)$ for large $T$.*

*Proof for Lemma A4.* If ST holds, then $X$ is either stationary or barely nonstationary. For a stationary $X$, we have $X_T = O_p(1) = o_p(\nu_T)$ since $\nu_T = s^{-1}(\lambda_T) \to \infty$ as $T \to \infty$. On the other hand, $X_T = o_p(\nu_T)$ for a barely nonstationary $X$ due to Lemma A3 (a), which completes the first part of proofs.

For the second part, we note that

$$\overline{X}_T = \frac{1}{T}\int_0^T X_t dt = \begin{cases} O_p(\lambda_T/T), & \text{if } \iota \text{ is } m\text{-integrable} \\ O_p(\lambda_T[m_s|\iota_s|](\lambda_T)/T), & \text{if } \iota \text{ is } m\text{-barely nonintegrable} \\ O_p(\lambda_T^2(m_s\iota_s)(\lambda_T)/T), & \text{if } \iota \text{ is } m\text{-strongly nonintegrable} \end{cases} \tag{A.2}$$

by Lemma A2. Trivially, if $\iota$ is $m$-integrable, then $\overline{X}_T = o_p(\nu_T)$ since $\lambda_T = O(T)$ and $\nu_T \to \infty$.

On the other hand, if $\iota$ is $m$-barely nonintegrable, then $|\iota_s|$ is $m_s$-barely nonintegrable, and hence, $[m_s|\iota_s|]$ becomes slowly varying due to Karamata's theorem. In a similar argument, $[m_s]$ is slowly varying since $m_s$ is either integrable or barely nonintegrable. Therefore, for some slowly varying function $\ell$ we have that

$$\frac{\lambda_T[m_s|\iota_s|](\lambda_T)}{\nu_T T} = \frac{[m_s|\iota_s|](\lambda_T)}{s^{-1}(\lambda_T)[m_s](\lambda_T)} \leq \frac{\ell(\lambda_T)}{s^{-1}(\lambda_T)} \to 0, \tag{A.3}$$

since $T = \lambda_T[m_s](\lambda_T)$ and $s^{-1} \in RV_p$ with $0 < p < \infty$ at $+\infty$ due to Lemma A1.

Now let $\iota$ be $m$-strongly nonintegrable. In this case, we have

$$\frac{\lambda_T^2(m_s\iota_s)(\lambda_T)}{\nu_T T} = \frac{\lambda_T(m_s\iota_s)(\lambda_T)}{s^{-1}(\lambda_T)[m_s](\lambda_T)} = \frac{\lambda_T m_s(\lambda_T)}{[m_s](\lambda_T)} \to 0, \tag{A.4}$$

where, in particular, the last convergence follows from Karamata's theorem since $m_s$ is either integrable or barely nonintegrable. This completes the proof. □

**Lemma A5.** *If a recurrent diffusion $X$ is defined on $\mathcal{D} = (0, \infty)$, then $(\iota/s')(x) \to 0$ and $1/s'(x) \to 0$ as $x \to 0$.*

*Proof for Lemma A5.* Since $X$ is recurrent on $(0, \infty)$, $s(0) = \int_\omega^0 s'(x)dx = -\infty$ for any

$\omega > 0$, and hence, $s' \in RV_p$ with $p < -1$ at $x_B = 0$ due to Assumption 2.2. This completes the proof. $\square$

**Lemma A6.** *Let DD hold. Then we have the followings.*

*(a)* $1/s'(x) \to 0$ *as* $x \to \underline{x}, \overline{x}$.

*(b)* $\iota/s' \circ s^{-1}(\lambda) \prec [m_s\sigma_s^2](\lambda)$ *and* $\iota/s' \circ s^{-1}(-\lambda) \prec [m_s\sigma_s^2](\lambda)$.

*(c)* $[m_s\sigma_s^2](\lambda) \sim -2[m_s\iota_s\mu_s](\lambda)$.

*(d)* $\nu_T^2 = o(\lambda_T[m_s\sigma_s^2](\lambda_T))$.

*Proof for Lemma A6 (a) and (b).* Due to Lemma A5, the stated results in (a) and (b) hold at $x_B = 0$ regardless of DD, and hence, it suffice to prove the statements at $x_B = \pm\infty$ when DD holds. We note that DD holds if and only if $1/s'$ is either integrable or barely nonintegrable since $m\sigma^2 = 1/s'$. Therefore, we have $1/s'(x) \to 0$ as $x \to x_B = \pm\infty$ which completes the proof of part (a).

For the part (b), we write

$$[m_s\sigma_s^2](\lambda) = \int_{-\lambda}^{\lambda} \frac{1}{(s')^2 \circ s^{-1}(x)} dx = \int_{s^{-1}(-\lambda)}^{s^{-1}(\lambda)} \frac{1}{s'(x)} dx. \tag{A.5}$$

Then, by Karamata's theorem, we have

$$\iota/s' \circ s^{-1}(\lambda)/[m_s\sigma_s^2](\lambda),\ \iota/s' \circ s^{-1}(-\lambda)/[m_s\sigma_s^2](\lambda) \to 0$$

which completes the proof of part (b). $\square$

*Proof for Lemma A6 (c).* We have

$$\begin{aligned} -2[m_s\iota_s\mu_s](\lambda) &= -2\int_{-\lambda}^{\lambda} m_s\iota_s\mu_s(x)dx = -2\int_{s^{-1}(-\lambda)}^{s^{-1}(\lambda)} m\iota\mu(x)dx = \int_{s^{-1}(-\lambda)}^{s^{-1}(\lambda)} \frac{xs''(x)}{[s'(x)]^2}dx \\ &= \int_{s^{-1}(-\lambda)}^{s^{-1}(\lambda)} \frac{1}{s'(x)}dx - \left[\frac{s^{-1}(\lambda)}{s' \circ s^{-1}(\lambda)} - \frac{s^{-1}(-\lambda)}{s' \circ s^{-1}(-\lambda)}\right] \end{aligned} \tag{A.6}$$

due to the integration by parts, from which joint with Lemma A6 (b) and (A.5) we have the stated result. $\square$

*Proof for Lemma A6 (d).* Due to Assumption 2.2 with the recurrence condition, we have $s' \in RV_p$ with $-1 < p \leq \infty$ at $+\infty$. If $s'$ is rapidly varying with $p = \infty$, then $s^{-1}$ becomes slowly varying, and hence,

$$\frac{\nu_T}{\sqrt{\lambda_T[m_s\sigma_s^2](\lambda_T)}} = \frac{s^{-1}(\lambda_T)}{\sqrt{\lambda_T[m_s\sigma_s^2](\lambda_T)}} \to 0,$$

since $[m_s\sigma_s^2]$ is monotonically increasing.

On the other hand, if $-1 < p < \infty$, then $\nu_T s'(\nu_T) \sim (p+1)s(\nu_T)$ by Karamata's theorem, and hence, we have

$$\frac{\nu_T}{\sqrt{\lambda_T[m_s\sigma_s^2](\lambda_T)}} \leq \frac{\nu_T}{\sqrt{s(\nu_T)\int_0^{\nu_T} 1/s'(x)dx}} \sim \left(\frac{(p+1)\nu_T}{s'(\nu_T)\int_0^{\nu_T} 1/s'(x)dx}\right)^{1/2} \to 0 \qquad \text{(A.7)}$$

since $\int_0^{\nu_T} 1/s'(x)dx \leq [m_s\sigma_s^2](\lambda_T)$ by (A.5) and $\int_0^{\nu_T} 1/s'(x)dx$ is slowly varying satisfying $(\nu_T/s'(\nu_T))/\int_0^{\nu_T} 1/s'(x)dx \to 0$ due to Karamata's theorem. This completes the proof. $\square$

**Lemma A7.** *If DD holds, then*

$$\int_0^T X_t\sigma(X_t)dW_t \prec_p \int_0^T X_t\mu(X_t)dt$$

*for large $T$.*

*Proof for Lemma A7.* It follows from Lemma A2 that

$$\int_0^T X_t\sigma(X_t)dW_t = \begin{cases} O_p\big(\sqrt{\lambda_T[m_s\iota_s^2\sigma_s^2](\lambda_T)}\big), & \text{if } \iota^2\sigma^2 \text{ is } m\text{-I or } m\text{-BN} \\ O_p\big(\lambda_T(m_s^{1/2}\iota_s\sigma_s)(\lambda_T)\big), & \text{if } \iota^2\sigma^2 \text{ is } m\text{-SN}. \end{cases} \qquad \text{(A.8)}$$

Moreover, we can deduce from Lemmas A2 and A6 (c) that

$$\frac{1}{\lambda_T[m_s\sigma_s^2](\lambda_T)}\int_0^T X_t\mu(X_t)dt \to_d -\frac{1}{2}L(\tau,0). \qquad \text{(A.9)}$$

Note that if $\iota^2\sigma^2$ is either $m$-I or $m$-BN, $[m_s\iota_s^2\sigma_s^2]$ is slowly varying. In this case, the stated result follows immediately from (A.8) and (A.9) since $\sqrt{\lambda_T[m_s\iota_s^2\sigma_s^2](\lambda_T)}/\lambda_T[m_s\sigma_s^2](\lambda_T) \to 0$ due to Karamata's theorem.

Now let $\iota^2\sigma^2$ be $m$-SN. Then we have

$$\frac{\lambda_T(m_s^{1/2}\iota_s\sigma_s)(\lambda_T)}{\lambda_T[m_s\sigma_s^2](\lambda_T)} = \frac{\nu_T}{s'(\nu_T)[m_s\sigma_s^2](\lambda_T)} \leq \frac{\nu_T}{s'(\nu_T)\int_0^{\nu_T} 1/s'(x)dx} \to 0, \qquad \text{(A.10)}$$

where the first equality is due to $m_s^{1/2}\iota_s\sigma_s = (\iota/s')\circ s^{-1}$, the second inequality follows from $\int_0^{\nu_T} 1/s'(x)dx \leq [m_s\sigma_s^2](\lambda_T)$ as in (A.7), and the convergence to zero follows from Karamata's theorem with the fact that $\int_0^{\nu_T} 1/s'(x)dx$ is slowly varying. The stated result for $m$-SN $\iota^2\sigma^2$ follows from (A.8), (A.9) and (A.10), which completes the proof. $\square$

**Lemma A8.** *If DD does not hold, then*

$$\int_0^T X_t\sigma(X_t)dW_t \gtrsim_p \int_0^T X_t\mu(X_t)dt$$

*for large $T$.*

*Proof for Lemma A8.* If DD does not hold, then $\iota^2\sigma^2$ is $m$-strongly nonintegrable, and hence, it follows from Lemma A2 that

$$\frac{1}{\lambda_T(m_s^{1/2}\iota_s\sigma_s)(\lambda_T)}\int_0^T X_t\sigma(X_t)dW_t \to_d \int_0^\tau \overline{m_s\iota_s\sigma_s}(B_t)dB_t. \tag{A.11}$$

We also have

$$\int_0^T X_t\mu(X_t)dt = \begin{cases} O_p\big(\lambda_T[m_s\iota_s\mu_s](\lambda_T)\big), & \text{if } \iota\mu \text{ is } m\text{-I or } m\text{-BN} \\ O_p\big(\lambda_T^2(m_s\iota_s\mu_s)(\lambda_T)\big), & \text{if } \iota\mu \text{ is } m\text{-SN.} \end{cases} \tag{A.12}$$

We first let $\iota\mu$ be either $m$-I or $m$-BN. It then follows from (A.6) and $m_s^{1/2}\iota_s\sigma_s = (\iota/s')\circ s^{-1}$ that

$$\frac{\lambda_T[m_s\iota_s\mu_s](\lambda_T)}{\lambda_T(m_s^{1/2}\iota_s\sigma_s)(\lambda_T)} \sim \frac{\int_{-\nu_T}^{\nu_T} 1/s'(x)dx}{(\iota/s')(\nu_T)} + O(1) = O(1)$$

due to Karamata's theorem with the fact that $1/s'$ is strongly nonintegrable.

Now let $\iota\mu$ be $m$-SN. We note that $s' \in RV_p$ at $x_B = +\infty$ for $p \in \mathbb{R}\cup\{+\infty, -\infty\}$ if and only if $x\mu(x)/\sigma^2(x) \to -p/2$ as $x \to +\infty$ due to the representation theorem of regularly and rapidly varying functions. Moreover, if a recurrent diffusion $X$ satisfies Assumption 2.2 without satisfying DD, then $-1 < p < 1$. Therefore, we have

$$\frac{\lambda_T^2(m_s\iota_s\mu_s)(\lambda_T)}{\lambda_T(m_s^{1/2}\iota_s\sigma_s)(\lambda_T)} = \frac{\lambda_T(m_s^{1/2}\mu_s)(\lambda_T)}{\sigma_s(\lambda_T)} = \frac{s(\nu_T)\mu(\nu_T)}{s'(\nu_T)\sigma^2(\nu_T)} \to \frac{-p}{2(1+p)}$$

since $s(\nu_T)/(\nu_T s'(\nu_T)) \to 1/(1+p)$ by Karamata's theorem, which completes the proof.

$\square$

**Lemma A9.** *If ST holds, then*

$$\int_0^T (X_t - \overline{X}_T)dX_t \sim_p -\frac{1}{2}\int_0^T \sigma^2(X_t)dt$$

*for large $T$.*

*Proof for Lemma A9.* We write

$$\int_0^T (X_t - \overline{X}_T)dX_t = \frac{1}{2}\left[(X_T^2 - X_0^2) - 2\overline{X}_T(X_T - X_0) - \int_0^T \sigma^2(X_t)dt\right]. \tag{A.13}$$

By applying Lemma A2 to the last term in (A.13), we have

$$\frac{1}{a_T}\int_0^T \sigma^2(X_t)dt \to_d P, \tag{A.14}$$

where $a_T$ and $P$ are defined in Section 3.2. Moreover, we know that if ST holds, then

$$(X_T^2 - X_0^2) - 2\overline{X}_T(X_T - X_0) = o_p(\nu_T^2) \tag{A.15}$$

due to Lemma A4.

If DD holds, $a_T = \lambda_T[m_s\sigma_s^2](\lambda_T)$ and $\nu_T^2 = o\left(\lambda_T[m_s\sigma_s^2](\lambda_T)\right)$ due to Lemma A6 (d). On the other hand, if DD does not hold, then $a_T = \lambda_T^2(m_s\sigma_s^2)(\lambda_T)$ and $\nu_T^2 = O(\lambda_T^2(m_s\sigma_s^2)(\lambda_T))$ because

$$\frac{\nu_T^2}{\lambda_T^2(m_s\sigma_s^2)(\lambda_T)} = \left(\frac{\nu_T s'(\nu_T)}{s(\nu_T)}\right)^2 \to (1+p)^2 \tag{A.16}$$

since $s(\lambda) \sim \lambda s'(\lambda)/(1+p)$ and $p \neq -1, \pm\infty$. In all cases, we have $\nu_T^2 = O(a_T)$, and hence, the stated result follows from (A.13)-(A.15). $\square$

**Lemma A10.** *If DD holds, then*

$$\int_0^T (X_t - \overline{X}_T)dX_t \sim_p -\frac{1}{2}\int_0^T \sigma^2(X_t)dt$$

*for large $T$.*

*Proof for Lemma A10.* Due to Lemma A9, it suffice to prove the lemma for a strongly nonstationary $X$. If $X$ is strongly nonstationary, then

$$\overline{X}_T = O_p(\lambda_T^2(m_s\iota_s)(\lambda_T)/T) = O_p(\nu_T)$$

by Lemma A2, and hence, we have

$$(X_T^2 - X_0^2) - 2\overline{X}_T(X_T - X_0) = O_p(\nu_T^2) \tag{A.17}$$

due to Lemma A3 (a). However, if DD holds, $\nu_T^2 = o\left(\lambda_T[m_s\sigma_s^2](\lambda_T)\right)$ due to Lemma A6 (d), and hence, the stated result follows from (A.13), (A.14) and (A.17). □

**Lemma A11.** *We have*

$$\sum_{i=1}^{n}\left(\Delta x_i - \overline{\Delta x}_n\right)^2 = \sum_{i=1}^{n}(\Delta x_i)^2 + O_p(\delta T(\mu^2)T) + O_p(\delta T(\sigma^2)) + O_p(\delta T(\mu\sigma)T^{1/2}),$$

*where* $\overline{\Delta x}_n = n^{-1}\sum_{i=1}^{n}\Delta x_i$*. Moreover, if* $\delta \to 0$*, then*

$$\begin{aligned}\sum_{i=1}^{n}(\Delta x_i)^2 = \int_0^T \sigma^2(X_t)dt + O_p(\delta T(\mu^2)T) + O_p(\delta T(\mu\sigma)T^{1/2}) \\ + O_p\left(\delta^{1/2}T(\mu\sigma)T\sqrt{\log(T/\delta)}\right) + O_p\left(\delta^{1/2}T(\sigma^2)T^{1/2}\sqrt{\log(T/\delta)}\right).\end{aligned}$$

*Proof for Lemma A11.* We write

$$\sum_{i=1}^{n}\left(\Delta x_i - \overline{\Delta x}_n\right)^2 = \sum_{i=1}^{n}(\Delta x_i)^2 + n(\overline{\Delta x}_n)^2. \tag{A.18}$$

For the second term in the right hand side of (A.18), we have

$$\overline{\Delta x}_n = \frac{X_T - X_0}{n} = \frac{\delta}{T}\left[\int_0^T \mu(X_t)dt + \int_0^T \sigma(X_t)dW_t\right] = \frac{\delta}{T}\left[O_p(T(\mu)T) + O_p(T(\sigma)T^{1/2})\right],$$

and hence,

$$n(\overline{\Delta x}_n)^2 = O_p(\delta T(\mu^2)T) + O_p(\delta T(\sigma^2)) + O_p(\delta T(\mu\sigma)T^{1/2}) \tag{A.19}$$

which together with (A.18) provides the first part.

Now, for the first term in the right hand side of (A.18), we have

$$\begin{aligned}\sum_{i=1}^{n}(x_i - x_{i-1})^2 - [X]_T &= \sum_{i=1}^{n}\Big(\big(X_{i\delta} - X_{(i-1)\delta}\big)^2 - \big([X]_{i\delta} - [X]_{(i-1)\delta}\big)\Big)\\ &= 2\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\big(X_t - X_{(i-1)\delta}\big)\,dX_t\\ &= 2\,(P_T + Q_T + R_T + S_T)\end{aligned}$$

by Itô's formula, where

$$\begin{aligned}P_T &= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\int_{(i-1)\delta}^{t}\mu(X_s)ds\right)\mu(X_t)dt = O_p(\delta T(\mu^2)T),\\ Q_T &= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\int_{(i-1)\delta}^{t}\mu(X_s)ds\right)\sigma(X_t)dW_t = O_p(\delta T(\mu\sigma)T^{1/2}),\\ R_T &= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\int_{(i-1)\delta}^{t}\sigma(X_s)dW_s\right)\mu(X_t)dt = O_p\left(\delta^{1/2}T(\mu\sigma)T\sqrt{\log(T/\delta)}\right),\\ S_T &= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\int_{(i-1)\delta}^{t}\sigma(X_s)dW_s\right)\sigma(X_t)dW_t = O_p\left(\delta^{1/2}T(\sigma^2)T^{1/2}\sqrt{\log(T/\delta)}\right)\end{aligned}$$

due to Lemmas B4 and B5 of Kim and Park (2015). This completes the proof of the second part since $[X]_T = \int_0^T \sigma^2(X_t)dt$. □

# Appendix B Proofs for Main Results

*Proof for Lemma 2.1.* The stated result follows immediately from Lemmas A7 and A8. □

*Proof for Lemma 3.1.* We note that

$$\sum_{i=1}^{n} x_{i-1}\delta = \int_0^T X_t dt + O_p(\delta T(\mu)T) + O_p(\delta T(\sigma)T^{1/2}), \tag{B.1}$$

$$\sum_{i=1}^{n} x_{i-1}^2\delta = \int_0^T X_t dt + O_p(\delta T(\iota\mu)T) + O_p(\delta T(\sigma^2)T) + O_p(\delta T(\iota\sigma)T^{1/2}) \tag{B.2}$$

due to Lemma B1 of Kim and Park (2015). Moreover, we may deduce from Lemma B3 of

Kim and Park (2015) that

$$\begin{aligned}\sum_{i=1}^{n} x_{i-1}\Delta x_i &= \sum_{i=1}^{n} X_{(i-1)\delta}\int_{(i-1)\delta}^{i\delta}\mu(X_t)dt + \sum_{i=1}^{n} X_{(i-1)\delta}\int_{(i-1)\delta}^{i\delta}\sigma(X_t)dW_t \\ &= \int_0^T X_t\mu(X_t)dt + \int_0^T X_t\sigma(X_t)dW_t + R_T + S_T \\ &= \int_0^T X_t dX_t + R_T + S_T \end{aligned} \tag{B.3}$$

where

$$\begin{aligned} R_T &= O_p(\delta T(\mu^2)T) + O_p(\delta T(\mu\sigma)T^{1/2}), \\ S_T &= O_p(\delta T(\mu\sigma)T^{1/2}) + O_p\left(\delta^{1/2}T(\sigma^2)T^{1/2}\sqrt{\log(T/\delta)}\right). \end{aligned}$$

The stated result for $\hat{\beta}$ follows immediately from (B.1)-(B.3) given Assumption 3.1.

As for $t(\hat{\beta})$, we write

$$\begin{aligned}\frac{\hat{\sigma}^2}{\delta} &= \frac{1}{n\delta}\sum_{i=1}^{n}\left[\Delta x_i - (\hat{\alpha} - \hat{\beta}x_{i-1})\delta\right]^2 \\ &= \frac{1}{n\delta}\sum_{i=1}^{n}\left(\Delta x_i - \overline{\Delta x}_n\right)^2 - \frac{1}{n\delta}\frac{\left[\sum_{i=1}^{n}(x_{i-1} - \overline{x}_n)\Delta x_i\right]^2}{\sum_{i=1}^{n}(x_{i-1} - \overline{x}_n)^2}. \end{aligned} \tag{B.4}$$

For the second term in (B.4), we may deduce from (B.1)-(B.3) with Assumption 3.1 that

$$\frac{1}{n\delta}\frac{\left[\sum_{i=1}^{n}(x_{i-1} - \overline{x}_n)\Delta x_i\right]^2}{\sum_{i=1}^{n}(x_{i-1} - \overline{x}_n)^2} = \frac{\delta}{T}\frac{\left(\int_0^T (X_t - \overline{X}_T)dX_t\right)^2}{\int_0^T (X_t - \overline{X}_T)^2 dt}(1 + o_p(1)), \tag{B.5}$$

where

$$\begin{aligned}\int_0^T (X_t - \overline{X}_T)dX_t &= \frac{1}{2}\left((X_T - \overline{X}_T)^2 - (X_0 - \overline{X}_T)^2 - \int_0^T \sigma^2(X_t)dt\right) \\ &= O_p(T(\iota^2)) + O_p(T(\sigma^2)T). \end{aligned}$$

For $b_T$ defined in Section 3.2, $b_T^{-1}\int_0^T (X_t - \overline{X}_T)^2 dt$ has a well defined and strictly positive limit a.s. due to Lemma A2. For the leading term in (B.5), we may therefore easily show that

$$\frac{\delta}{T}\frac{\left[\int_0^T (X_t-\overline{X}_T)dX_t\right]^2}{\int_0^T (X_t-\overline{X}_T)^2dt} = O_p\left(\delta T(\iota^4)/(Tb_T)\right) + O_p\left(\delta T(\sigma^4)/(Tb_T)\right) = o_p(1), \qquad \text{(B.6)}$$

where, in particular, the last equality is due to Assumption 3.1. It follows from (B.4), (B.5) and (B.6) with Lemma A11 that

$$\frac{\hat{\sigma}^2}{\delta} = \frac{1}{T}\sum_{i=1}^n \left[\Delta x_i - (\hat{\alpha}-\hat{\beta}x_{i-1})\delta\right]^2 = \frac{1}{T}\int_0^T \sigma^2(X_t)dt(1+o_p(1)) \qquad \text{(B.7)}$$

due to Assumption 3.1. The stated result for $t(\hat{\beta})$ follows immediately from (B.7) with the proof of the first part of this lemma.

□

*Proof for Lemma 3.2.* Due to Lemma A2, we have the desired convergences for $[X]_T$, $\int_0^T (X_t - \overline{X}_T)dX_t$ and $\int_0^T (X_t-\overline{X}_T)dt$. To complete the proof, it suffice to show that $Ta_T/b_T \to \infty$. As defined in Section 3.2, $a_T$ is determined by DD, whereas $b_T$ is depending upon SI. If SI holds, then $b_T = \lambda_T \ell(\lambda_T)$ for some slowly varying function $\ell$, and hence, $Ta_T/b_T = Ta_T/(\lambda_T\ell(\lambda_T)) \to \infty$ since $\lambda_T \precsim T$ and $a_T/\ell(\lambda_T) \to \infty$.

Now we let SI do not hold. If both DD and ST hold, then

$$\begin{aligned}\frac{Ta_T}{b_T} = \frac{[m_s](\lambda_T)[m_s\sigma_s^2](\lambda_T)}{(m_s\iota_s^2)(\lambda_T)} &= \frac{[m_s](\lambda_T)}{\lambda_T m_s(\lambda_T)}\ \frac{[m_s\sigma_s^2](\lambda_T)}{\lambda_T(m_s\sigma_s^2)(\lambda_T)}\ \frac{\lambda_T^2(m_s\sigma_s^2)(\lambda_T)}{\iota_s^2(\lambda_T)} \\ &\equiv A_TB_TC_T \to \infty \qquad \text{(B.8)}\end{aligned}$$

since $A_T, B_T \to \infty$ and $C_T = \left(s(\nu_T)/\nu_T s'(\nu_T)\right)^2 \to 1/(p+1)^2$ by Karamata's theorem, where $-1 < p < \infty$ due to Lemma A1 with SI.

If DD holds and ST does not hold, then

$$\frac{Ta_T}{b_T} = \frac{\lambda_T[m_s\sigma_s^2](\lambda_T)}{\iota_s^2(\lambda_T)} = B_TC_T \to \infty$$

due to the same argument in (B.8).

On the other hand, if DD does not hold, then ST should be satisfied. In this case, we have

$$\frac{Ta_T}{b_T} = \frac{\lambda_T[m_s](\lambda_T)\sigma_s^2(\lambda_T)}{\iota_s^2(\lambda_T)}, = A_TC_T \to \infty$$

due to the same argument in (B.8). In all cases, we have $Ta_T/b_T \to \infty$, which completes the proof. $\square$

*Proof for Lemma 3.3.* We have

$$\frac{1}{\lambda_T^2 (m_s\sigma_s^2)(\lambda_T)}[X]_T \to_d \int_0^\tau \overline{m_s\sigma_s^2}(B_t)dt$$

and $\tau = \overline{A}_1$ due to Lemma A2, and

$$\frac{1}{\nu_T^2}(X_T^2 - X_0^2) - 2\overline{X}_T(X_T - X_0) \to_d (X_1^{\circ 2} - X_0^{\circ 2}) - 2\overline{X}_1^\circ (X_1^\circ - X_0^\circ)$$
$$\frac{1}{T\nu_T^2}\int_0^T (X_t^\circ - \overline{X}_T)dt \to_d \int_0^1 (X_t - \overline{X}_1^\circ)dt$$

due to Lemma A3 (a). Moreover, we have $\nu_T^2 \sim (1+p)^2\lambda_T^2(m_s\sigma_s^2)(\lambda_T)$ with $p \neq -1, \pm\infty$ as shown in (A.16), and

$$\int_0^1 (X_t^\circ - \overline{X}_1^\circ)dX_t^\circ = \frac{1}{2}\left[(X_1^{\circ 2} - X_0^{\circ 2}) - 2\overline{X}_1^\circ (X_1^\circ - X_0^\circ) - [X^\circ]_t\right],$$

where $[X^\circ]_t = \int_0^{\overline{A}_t}\left(\overline{(s^{-1})'_-}\right)^2(B_s)ds$, due to Itô's formula with Lemma A3 (b). Therefore, the stated result follows immediately if we show

$$\left(\overline{(s^{-1})'_-}\right)^2 = \frac{1}{(1+p)^2}\overline{m_s\sigma_s^2}. \tag{B.9}$$

To show (B.9), we note that $m_s\sigma_s^2 = \left((s^{-1})'\right)^2$ since $m_s\sigma_s^2 = (1/s' \circ s^{-1})^2$ and $(s^{-1})' = 1/s' \circ s^{-1}$. It then follows from the property of regularly varying function and Assumption 2.2 that $(s^{-1})' \in RV_\kappa$ with $\kappa = -p/(p+1) > -1$, from which joint with Karamata's theorem we can show that $\overline{(s^{-1})'} = (1+p)\overline{(s^{-1})'_-}$. Therefore, we have (B.9), which completes the proof. $\square$

*Proof for Theorem 3.4.* The stated results follow immediately from Lemmas 3.2 and 3.3 with Lemma 3.1. $\square$

*Proof for Lemmas 4.1 and 4.2.* The stated results follow immediately from Lemmas 3.2 and 3.3. $\square$

## References

Aït-Sahalia, Y. and J. Y. Park (2016): "Bandwidth Selection and Asymptotic Properties of Local Nonparametric Estimators in Possibly Nonstationary Continuous-Time Models," *Journal of Econometrics*, 192, 119–138.

Baillie, R. T. and T. Bollerslev (1989): "Common Stochastic Trends in a System of Exchange Rates," *Journal of Finance*, 44, 167–181.

——— (1994): "Cointegration, Fractional Cointegration, and Exchange Rate Dynamics," *Journal of Finance*, 49, 737–745.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1993): *Regular Variation*, Cambridge, UK: Cambridge University Press.

Bossaerts, P. (1988): "Common Nonstationary Components of Asset Prices," *Journal of Economic Dynamics and Control*, 12, 347–364.

Bossaerts, P. and R. Green (1989): "A General Equilibrium Model of Changing Risk Premia: Theory and Evidence," *Review of Financial Studies*, 2, 467–493.

Boswijk, H. P. and F. Klaassen (2012): "Why Frequency Matters for Unit Root Testing in Financial Time Series," *Journal of Business & Economic Statistics*, 30, 351–357.

Chambers, M. J. (2004): "Testing for unit roots with flow data and varying sampling frequency," *Journal of Econometrics*, 119, 1–18.

——— (2008): "Corrigendum to: Testing for unit roots with flow data and varying sampling frequency," *Journal of Econometrics*, 144, 524–525.

Chen, X., L. P. Hansen, and M. Carrasco (2010): "Nonlinearity and Temporal Dependence," *Journal of Econometrics*, 155, 155–169.

Choi, I. (1992): "Effects of data aggregation on the power of tests for a unit root: A simulation study," *Economics Letters*, 40, 397 – 401.

Choi, I. and B. S. Chung (1995): "Sampling frequency and the power of tests for a unit root: A simulation study," *Economics Letters*, 49, 131–136.

Davis, R. A. (1982): "Maximum and minimum of one-dimensional diffusions," *Stochastic Processes and their Applications*, 13, 1–9.

Diebold, F. X., J. Gardeazabal, and K. Yilmaz (1994): "On Cointegration and Exchange Rate Dynamics," *Journal of Finance*, 49, 727–735.

Fama, E. F. and K. R. French (1988): "Permanent and Transitory Component of Stock Prices," *Journal of Political Economy*, 96, 246–273.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst (2006): "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Review of Financial Studies*, 19, 797–827.

Höpfner, R. and Y. Kutoyants (2003): "On a Problem of Statistical Inference in Null Recurrent Diffusions," *Statistical Inference for Stochastic Processes*, 6, 25–42.

Jeong, M. and J. Y. Park (2013): "Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model," Working Paper, Indiana University.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

Kim, J. and J. Y. Park (2015): "Asymptotics for Recurrent Diffusions with Application to High Frequency Regression," Forthcoming in *Journal of Econometrics*.

Kim, M. J., C. R. Nelson, and R. Startz (1991): "Mean Reversion in Stock Prices: A Reappraisal of the Empirical Evidence," *Review of Economic Studies*, 58, 515–528.

Perron, P. (1989): "Testing for a random walk: A simulation experiment when the sampling interval is varied. In B. Raj (ed.)," in *Advances in Econometrics and Modeling*, Dordrecht: Kluwer Academic Publisher, 47–68.

——— (1991): "Test Consistency with Varying Sampling Frequency," *Econometric Theory*, 7, 341–368.

Poterba, J. M. and L. H. Summers (1988): "Mean Reversion in Stock Prices: Evidence and Implications," *Journal of Financial Economics*, 22, 27–59.

Shiller, R. J. and P. Perron (1985): "Testing the random walk hypothesis: Power versus frequency of observation," *Economics Letters*, 18, 381 – 386.

Stone, C. (1963): "Limit Theorems for Random Walks, Birth and Death Processes, and Diffusion Processes," *Illinois Journal of Mathematics*, 7, 638–660.