Persistent Stationary Process\textsuperscript{1}

Yoosoon Chang\textsuperscript{2}  Michael Chung\textsuperscript{3}  Joon Y. Park\textsuperscript{4}

Abstract

This paper presents a novel characterization of continuous time processes that captures the primary idiosyncratic features of many financial time series. We extend the analysis of unit root behaviors to incorporate series that have continuous sampling, but do so in such a way that the overall series does not tend towards explosive paths, as is implied by many unit root setups. In doing so, we enumerate a theory of asymptotics and regressions, presenting both simulation evidence and empirical backing for the existence, as well as the characteristic behavior of such a series in multiple real financial time series data.

\textsuperscript{1}The authors would like to thank Yongok Choi for many helpful discussions and suggestions.
\textsuperscript{2}Department of Economics, Indiana University.
\textsuperscript{3}Department of Economics, Indiana University.
\textsuperscript{4}Department of Economics, Indiana University and Sungkyunkwan University.

This version: April 20, 2012
JEL Classification: C13, C22

Keywords: persistency, near-unit root process, stationarity and nonstationarity, spurious regression, diffusion, markov switching
1. Introduction

The fitted models of many economic time series have characteristics dependent upon the sampling horizon $T$. In particular, it is often found that their degree of persistency increases with $T$, providing a strong motivation to use what is commonly referred to as the near-unit root model to specify their law of motion. There are numerous examples of the use of such a model in empirical applications, many of which have grown from implications of theory, as for future contracts (1965) and stock prices (1973) as postulated by Samuelson, for dividends and earnings (Kleidon, 1986), spot rates and exchange rates (Meese and Singleton, 1983) as well as a variety of economic aggregates. Systematic testing of multiple time series have failed to reject the possibility of such unit roots in discrete time frameworks, most convincingly in early work by Nelson and Plosser (1982), and later updated and reviewed with more sophisticated methodologies by Perron (1988). Though such evidences have occasioned skepticism in their results (see, e.g. Zivot and Andrews, 2002), our model’s behavior does not critically rely on changes in structural parameters, but instead characterizes the overall behavior of a series. At this juncture, the theories for such unit root behaviors in discrete have been well explored, as explicated in Dickey and Fuller (1979), extended and refined in Said and Dickey (1984) and Solo (1984), and further generalized by Philips and Perron (1988), and developed with regression theory by Phillips (1987).

Near-unit root models, however, have some unrealistic empirical implications. The processes generated by them are equally explosive as the exact unit root processes, growing stochastically at the rate of $\sqrt{T}$. This implies, among other things, that the marginal distribution of the underlying process becomes progressively more diffuse without any bound. Clearly, many economic time series do not show such a behavior. Though they are persistent, their marginal distributions change little over time with support which is naturally confined over a certain range of the real line regardless of $T$. The interest rates and the volatilities processes of many asset returns are good examples. Financial ratios such as dividend/price and earnings/ratio, which are used commonly in predictive regressions for stock returns, also show such a behavior.

We present some illustrative examples. First, we fit daily short rates to CIR model with various time horizons ($T$) and see if their mean reversion parameter decreases in absolute value as $T \to \infty$, and if their volatility parameter also decreases as $T \to \infty$. Second, we fit daily dividend/price and earnings/price ratios using Ornstein-Uhlenbeck process and see how their mean reversion and volatility parameters change as $T \to \infty$. Finally, we use the markov switching model to fit daily excess return volatility data and see if their persistency increases as $T \to \infty$.

In the paper, we introduce and analyze a novel process, called persistent stationary process, which has the properties of these and many other economic time series. By construction, the persistent stationary process becomes increasingly more persistent, though asymptotically it has time invariant marginal distribution, as $T$ gets large. The persistent stationary process can be defined more explicitly for diffusion and markov switching models, two types of models which are mostly widely used in practical applications. The process is generated if we set the drift and variance terms of a diffusion model to decrease at the same rate as $T$ increases. On the other hand, if the probability of changing states in a
markov switching model decreases proportionately with $T$, the switching process becomes a persistent stationary process.

Though it is stationary, law of large numbers does not apply for a persistent stationary process due to the presence of persistency. Therefore, we may well expect that the usual asymptotic theory is not applicable for persistent stationary processes. In particular, it is shown in the paper that the usual nonparametric density estimate obtained from the observations of a persistent stationary process does not converge to its invariant density. Instead, it converges to a random function, which may be interpreted as the local time of the limit process of the underlying persistent stationary process. The regressions with persistent stationary processes also yield unconventional asymptotics. Most importantly, they are meaningful only when their errors do not have persistency. Otherwise, they become spurious and yield all the well known consequences of spurious regressions.

The rest of the paper is organized as follows: a brief definition and basic terminology for the discussion of persistent stationary processes is presented, alongside the primary applications explored in this paper, followed by a set of empirical evidence for processes that we argue obey this behavior. Next, we develop theory regarding nonparametric density estimation of our process, followed by simulation evidence and an empirical illustration of interest rate behavior. Finally, we present a theory of regressions when these processes are involved, developing some results that are analogous to discrete time behavior in unit root models, as well as some results unique to the continuous time framework.

2. Definition and Basic Properties

2.1 Theory

For a continuous time process $X = (X_t), 0 \leq t \leq T$, we define a process $X^T = (X^T_t)$ by

$$X^T_t = X_{Tt}$$

for $0 \leq t \leq 1$.

**Definition 2.1** A stochastic process $X$ is called a persistent stationary process if

$$X^T \rightarrow_d \bar{X}$$

as $T \to \infty$ in $D[0,1]$, where $\bar{X}$ is a non-degenerate stationary process such that

$$\hat{\rho}_h = \frac{\mathbb{E}X_t \bar{X}_{t-h}}{\mathbb{E}X^2_t} \to 1$$

as $h \to 0$.

Throughout the paper, we call $\bar{X}$ the limit process of $X$. If we denote by $\rho$ the correlation function of $X$, correspondingly as the correlation function $\hat{\rho}$ for $\bar{X}$, then we have

$$\rho_h = \frac{\mathbb{E}X_t X_{t-h}}{\mathbb{E}X^2_t} \approx \frac{\mathbb{E}X^T_t \bar{X}_{(t-h)/T}}{\mathbb{E}X^2_{t/T}} = \hat{\rho}_h/T$$
for large $T$ under suitable regularity condition, which converges to unity for any $h$ as $T \to \infty$. Therefore, though it is asymptotically stationary, the process $X$ becomes more persistent as $T$ gets large.

We may define $X$ for each $T$, so that

$$X^T = d \tilde{X}$$

(1)

for some stationary process satisfying the conditions in Definition 2.1. Such a process $X$ clearly satisfies the conditions in Definition 2.1, and therefore, it is a persistent stationary process. Two simple examples. The first example is the Ornstein-Uhlenbeck process, which is given as the solution to the stochastic differential equation (SDE)

$$dX_t = \frac{\kappa}{T}(\mu - X_t)dt + \frac{\sigma}{\sqrt{T}}dW_t,$$

where $\mu \in \mathbb{R}, \kappa, \sigma \in \mathbb{R}_{++}$ are parameters and $W$ is standard Brownian motion. In this case, the required condition (1) holds with $\tilde{X}$ given by the solution of

$$d\tilde{X}_t = \tilde{\kappa}(\mu - \tilde{X}_t)dt + \tilde{\sigma}dW_t$$

with standard Brownian motion $\tilde{W}$, i.e., the Ornstein-Uhlenbeck process with parameters $\mu, \kappa$ and $\sigma$. More generally, if

$$dX_t = \mu_t dt + \sigma_t dW_t,$$

and $\bar{\mu}$ and $\bar{\sigma}$ are given by $\bar{\mu}_t = T\mu_{Tt}$ and $\bar{\sigma}_t = \sqrt{T}\sigma_{Tt}$, the condition (1) is satisfied with $\tilde{X}$

$$d\tilde{X}_t = \bar{\mu}_t dt + \bar{\sigma}_t d\tilde{W}_t,$$

where $W$ and $\tilde{W}$ are standard Brownian motions.

The second example is the homogeneous markov process $X$ on the state space consisting of two points $\{x_1, x_2\}$, $x_1, x_2 \in \mathbb{R}$, which has the transition matrix given by

$$P_t = \left( \begin{array}{cc} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{array} \right) + \exp\left(-\frac{\bar{\nu}}{T}t\right) \left( \begin{array}{cc} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{array} \right),$$

where $c > 0$ is a parameter and $0 < \pi_1, \pi_2 < 1$ are probabilities such that $\pi_1 + \pi_2 = 1$. For $i, j = 1, 2$, the $(i, j)$-th entry of the transition matrix $P$ represents the probability of transition of $X$ from $x_i$ to $x_j$. Clearly, the equilibrium distribution is given by two point probabilities $\pi_1$ and $\pi_2$ respectively on $x_1$ and $x_2$. Furthermore, if $X$ is started from its equilibrium distribution, it becomes stationary, i.e., it has the invariant distribution identical to its equilibrium distribution. For the process $X$ generated by this markov switching model, the condition in (1) is satisfied with a homogeneous markov process $\tilde{X}$, which is defined on the same state space as $X$ and has the transition matrix

$$\tilde{P}_t = \left( \begin{array}{cc} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{array} \right) + \exp(-\bar{\nu}t) \left( \begin{array}{cc} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{array} \right).$$

Note that $\tilde{X}$ has the same equilibrium and invariant distributions as $X$. 3
More generally, homogeneous markov processes with finite state spaces have transition matrices that can be represented as
\[
P_t = Q_0 + \sum_{k=1}^{m} \exp\left( -\frac{\bar{\nu}_k t}{T} \right) Q_k
\]
under appropriate regularity conditions, where $\bar{\nu}_k \in \mathbb{R}_{++}$, $(Q_k)$ are $m$-dimensional square matrices such that $\sum_{k=0}^{m} Q_k = I$, $Q_i^2 = Q_i$ and $Q_i Q_j = 0$ for all $0 \leq i \neq j \leq m$, and $Q_0$ has identical rows presenting the equilibrium and invariant distribution of the underlying process. For such a process, $X$ becomes the homogeneous markov process with transition matrix
\[
\tilde{P}_t = Q_0 + \sum_{k=1}^{m} \exp(-\bar{\nu}_k t)Q_k
\]
similarly as above.

The process $X^T$ may be regarded as an embedded version of the given process $X$ on the unit interval. For a wide class of nonstationary null recurrent markov processes, such an embedded version weakly converges if properly standardized. In fact, Jeong and Park (2010) shows that for a general null recurrent diffusion $X$ in natural scale we have
\[
T^{-1/(r+2)} X^T \to_d \tilde{X}
\]
as $T \to \infty$, where the index $r > -1$ is the asymptotic degree of homogeneity of the speed density of $X$ and the limit process $\tilde{X}$ becomes with an appropriate scaling a skew Bessel process in natural scale of dimension $2(r + 1)/(r + 2)$. The persistent stationary process may be viewed as the limit case of this with $r = \infty$.

For a persistent stationary process $X$, we have
\[
\frac{1}{T} \int_0^{T} X_t dt = \int_0^{1} X_t^T dt \to_d \int_0^{1} \tilde{X}_t dt
\]
for its limit process $\tilde{X}$. Therefore, the conventional law of large numbers for a stationary process does not hold. The sample mean of a persistent stationary process converges in distribution to a random variable, which is defined as an integral of its limit process. Obviously, this is due to the presence of persistency in a persistent stationary process. In contrast, we have
\[
\frac{1}{T} \int_0^{T} X_t dt \to_{a.s.} \mathbb{E}(X_t)
\]
for a stationary and ergodic process $X$. Moreover, we would expect
\[
\frac{1}{T} \int_0^{T} X_t dt \to_p \infty
\]
in general for a nonstationary null recurrent process.
2.2 Empirical Evidence

The existence of such a persistent stationary process can be illustrated by a few cogent examples. Here, we take series that we imagine should asymptotically achieve time invariant marginal distributions based upon our intuitions regarding their long run behavior and subsequently run a series of estimations that elucidate their natures as persistent stationary processes. In particular, we present results from an three primary estimations: of interest rates using a CIR diffusion framework, of dividend/price ratios using an OU process and of VIX data modeled using Markov switching. Additional examples can be found in the Appendix.

Consider the behavior of a data series in which we militate that the mean reversion parameter must obey a linear decrease with respect to the time horizon of estimation; that is, for linear drift, a series in which the drift term obeys the behavior \( \dot{\mu} = \mu - \lambda X \). We posit that this presumption, once imposed on a persistent stationary process, generates a commensurate requirement on the resultant volatilities. To show this result, we enforce this rule throughout our estimation process by conducting the full sample estimation, from \([0, T_{\text{max}}]\) to generate the requisite reference mean \( \hat{\mu} \) and mean reversion parameter \( \hat{\kappa} \). Subsequently, for every estimation at time horizons shorter than \( T_{\text{max}} \), we decompose the time series into a sequence of disjoint intervals over which we run another set of estimations. For these, we fix the drift parameter to \( \frac{\hat{\mu} - X}{\hat{\kappa}} \), where \( T \) ranges from a 5 year observation span up to \( T_{\text{max}} \). The resultant estimation values of the volatility are then averaged within each grouping of subsamples sharing the same time horizon \( T \). If, indeed, our model correctly specifies the relational behavior of the drift and volatility, and the processes in question are of the persistent stationary type, we should observe a commensurate drop in volatility as the time horizon lengthens. We can see that this indeed follows in Figures 1 and 2. We use in these estimations daily treasury 3 month bond yield data from the Federal Reserve Economics Database (DTB3) running from 1954-2011 and dividend price ratio data constructed from CRSP data, with dividends computed, as usual, from the 1 year accumulations of the difference in value-weighted returns, including dividends (VWRETD) and excluding dividends (VWRETX), adjusted by total market value (TOTVAL) running from 1927-2009.

3. Density Estimation of Persistent Stationary Process

3.1 Theory

We consider how the standard nonparametric kernel density estimator behaves if samples are collected from a persistent stationary process. For this, we assume that a set of discrete time observations are made from the continuous time process \( X \) over time interval \([0, T]\), and set \( x_i = X_{i\delta} \) for \( i = 1, \ldots, n \) with \( T = n\delta \). Then we define

\[
\hat{\pi}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x_i - x}{h} \right),
\]

where \( h \) is the bandwidth parameter and \( K \) is the kernel function such that \( \int_{-\infty}^{\infty} K(x)dx = 1 \). In our subsequent analysis, we assume that \( K \) is Lipschitz continuous. Moreover, we let
\[ \delta \to 0, \text{ as well as } h \to 0, \text{ as } T \to \infty. \]

For a stationary ergodic process \( X \), we have under general regularity conditions

\[ \hat{\pi}(x) \to_p \pi(x) \]

as \( T \to \infty \), where \( \pi \) is the time invariant marginal density of \( X \).\(^5\)

In case of a persistent stationary process, the kernel density estimator \( \hat{\pi} \) is in general consistent for its time invariant marginal distribution. Once again, this is due to the presence of persistency. To characterize the limit behavior of \( \hat{\pi} \) for the persistent stationary process, we need to introduce some technical assumptions. In what follows, we assume that the local times \( \ell, \ell^T \) and \( \bar{\ell} \) respectively of \( X, X^T \) and \( \bar{X} \) exist and are well defined, and denote by \( \mathcal{D} \) the state space of \( X, X^T \) and \( \bar{X} \). Note that

\[
\ell(t, x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|X_s - x| < \varepsilon\} ds
\]

\[
= T \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^{t/T} 1\{|X_{Ts} - x| < \varepsilon\} ds
\]

\[
= T \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^{t/T} 1\{|X^T_{Ts} - x| < \varepsilon\} ds
\]

\[
= T \ell^T(t/T, x),
\]

and therefore, we have in particular \( \ell^T(1, x) = \ell(T, x)/T \) for all \( x \in \mathcal{D} \).

\(^5\)In this case, we may let \( \delta \) be either fixed or decreasing down to zero as \( T \to \infty \).
In what follows, we write $X = X^c + X^d$, where $X^c$ is the continuous part of $X$ and $X^d$ is the jump part of $X$ defined as $X^d_t = \sum_{0 \leq s \leq t} \Delta X_s$ with $\Delta X_t = X_t - X_{t-}$.

**Assumption 3.1** Let $X$ be a persistent stationary process such that
(a) $\sup_{|t-s| \leq \delta} |X^c_t - X^c_s| = O_p(\delta^\kappa a_T)$ for all $0 \leq s, t \leq T$ with some $\kappa > 0$ and $a_T$,
(b) $\mathbb{E} \left( \sum_{0 \leq t \leq T} |\Delta X_t| \right)^2 = O(b_T^2)$,
(c) $h, \delta \to 0$ and $T \to \infty$ in such a way that $\delta^\kappa a_T/h^2 \to 0$ and $\delta b_T/h^2 T \to 0$, and
(d) $\ell^T(1, x) \to \ell(1, x)$ for all $x \in \mathcal{D}$, and $\ell^T(1, \cdot)$ is continuous and bounded a.s. on $\mathcal{D}$ uniformly in $T$.

The required conditions in Assumption 3.1 are mild and satisfied by a wide class of stochastic processes. Condition (a) holds with $\kappa = 1/2$ and $a_T = 1$ for Brownian motions and more generally for general diffusions with bounded diffusion functions. It is well known that $\max_{0 \leq t \leq \delta} |W(s + t) - W(s)| = \tilde{\omega} = d |\mathcal{N}(0, \delta)|$ for all $s \geq 0$. Condition (b) holds with $b_T = \sqrt{T}$ for a general persistent process $X$ whose jump part is given by the general Lévy jump process
$$X^d_t = \int_0^t \int_{\mathbb{R}} x \lambda(dt, dx),$$
for which we have
$$\int_{\mathbb{R}} (|x| + |x|^2) \lambda(dx) < \infty,$$
where $\Lambda$ is the Poisson random measure with $\mathbb{E}\Lambda(dt, dx) = \lambda(dx)dt$. The reader is referred to, e.g., Applebaum (2009) for the representation and relevant theory of the Lévy jump process. We have two conditions in (c) to make the errors negligible in approximating the discrete sample moments by their continuous time analogues respectively for the continuous and jump parts of $X$. They are innocuous, since they are met as long as the bandwidth parameter $h$ is chosen appropriately. In fact, the condition for the continuous part is not absolutely necessary and may be even weakened if we define $X$ more specifically as in Aït-Sahalia and Park (2010). Condition (d) is also not very restrictive. We may naturally expect that the weak convergence of local times follows from that of the underlying processes, which hold by our definition. Furthermore, it is well known that the local time is continuous in its spatial argument for a large class of semimartingales. Finally, for a general stationary ergodic process $\ell^T(1, \cdot)$ converges as $T \to \infty$ to its time invariant marginal density, and therefore, it will be bounded uniformly for all large $T$ as long as the time invariant marginal density of the underlying process is bounded. The reader is referred to Bosq (1998) for more details.

**Theorem 3.1** Let Assumption 3.1 hold. Then we have

$$\hat{\pi}(x) \to_d \ell(1, x)$$

as $T \to \infty$ for all $x \in \mathcal{D}$.

Theorem 3.1 shows that the standard kernel density estimator obtained from the observations of a persistent stationary process does not converge to its time invariant marginal density. This is in sharp contrast with a stationary ergodic process. The density estimator of a persistent stationary process therefore remains random even we increase the sampling horizon up to infinity.

### 3.2 Simulations

Here we emphasize the contrasting characteristics of a typical stationary process and one with a persistent stationary nature through the use of simulations. We first use as a data generating process the CIR model to simulate an interest rate series. Full sample estimates, roughly 50 years in length, of 3 month T-bill data are used to construct plausible parameters for mean reversion $\kappa$, mean $\mu$ and volatility $\sigma$. By construction, such a series produces stationary behavior, and as such, its kernel density estimate asymptotically converges to its time invariant marginal density, well known to be a Gamma distribution with shape and scale parameters given respectively by $\frac{2\mu}{\sigma^2}$ and $\frac{\sigma^2}{2\kappa}$. We then simulate 10,000 such series for varying time horizons $T$, and compute the 95% confidence band for each $T$. We can clearly see the bands asymptotically approach the Gamma distribution in Figure 3. To generate the contrasting series, in which we hypothesize that interest rates in fact possess a persistent stationary nature, we appeal to the work of Phillips and Yu (2005) and suppose that the true mean reversion parameter may be as low as one fourth of its estimated value, which will in turn generate a persistent process of the type we wish. This persistent stationary process will subsequently have differing mean reversions and volatilities which will explicitly
depend on the time horizon, given as $\bar{\kappa}/T$ and $\bar{\sigma}/\sqrt{T}$. Thus, to generate a set of simulations for varying $T$ for this persistent stationary process, we scale the 50-year mean reversion and volatility such that $\bar{\kappa}/50 = \kappa/4$ and $\bar{\sigma}/\sqrt{50} = \sigma$. The values of $\bar{\kappa}$ and $\bar{\sigma}$ are then used to generate series of the time horizons we wish by appropriately scaling according to $T$. The results of these simulations can be seen in Figure 4. The confidence bounds for any horizon $T$ are indistinguishable, and most notably, fail to improve as the simulation horizon lengthens. Indeed we should expect this result as a consequence of the developed theory, as instead of convergence to the purported time invariant marginal density, we should expect convergence to a random function given by the local time of the persistent process.

![Figure 3: Stationary Process 95% Confidence Bounds](image)

**Figure 3: Stationary Process 95% Confidence Bounds**

### 3.3 Empirical Illustrations

We would also like to extend our analysis to actual interest rate data to see if we achieve a similar conformity in kernel densities as compared to the simulated data. We consider the kernel density estimate for a variety of time horizons. If, indeed, the interest rate obeys a stationary process, we should expect that the shape of the kernel density should be insensitive to the estimation time horizon, once $T$ is sufficiently large. That is, the invariant marginal distribution, over a sufficiently long time span, would generate a precise estimate of the true density. We can see, however, that this fails to be the case for actual 3 month T-bill rates. We generate the nonparametric kernel densities by drawing the first $T$ years
of interest rate data (starting at 1954) and computing a single density estimate, comparing results for various horizons $T$. There remains no evident pattern of convergence as we can see in Figure 5.

4 Regressions with Persistent Stationary Processes

4.1 Theory

Now we consider the regressions with persistent stationary processes. The regressions behave quite differently depending upon whether or not the persistent stationary processes involved in the regressions have a common persistent stochastic trend. If a common persistent stochastic trend exists, the corresponding regression represents a meaningful relationship and can be estimated consistently. On the other hand, in the lack of such a common trend, the regression becomes spurious and yields nonsensical results. This is entirely analogous to the regressions involving unit root nonstationary time series in discrete time. As is well known, the regression with unit root processes defines a cointegrating relationship if the regressand and regressor share a common integrated stochastic trend, whereas it reduces to a spurious regression if there does not exist a common trend in the regressand and regressor. For the regressions with persistent stationary processes, the regressand and regressor are
said to be *co-persistent* if they have a common persistent stochastic trend and their linear combination yields a transient stationary process.

In what follows, we let $Y$ and $X$ be two processes, respectively of dimensions 1 and $m$, which generate the regressand and regressor for the regressions with persistent stationary processes. More explicitly, we suppose that the observations

$$y_i = Y_i \delta \quad \text{and} \quad x_i = X_i \delta$$

are available for $i = 1, \ldots, n$ in discrete time, where we set $T = n\delta$ as before, and we analyze the regression

$$y_i = x_i' \beta + u_i,$$

where $\beta$ is an $m$-dimensional parameter and $(u_i)$, $u_i = U_i \delta$ for $i = 1, \ldots, n$, is the regression error. The asymptotics of regression (2) depends crucially on whether $Y$ and $X$ are co-persistent.

For the co-persistent regression, we assume that they are generated as

$$Y_t = X_t' \beta + U_t$$

with an $m$-dimensional parameter $\beta$, where $U$ is a mean zero transient stationary process. Note that $Y$ and $X$ are individually persistent, but their linear combination given by $(1, -\beta)'$ yields a non-persistent stationary process. Therefore, $Y$ and $X$ share a common persistent stochastic trend and are co-persistent. To analyze the spurious regression, on the other hand, we let $Y$ and $X$ be independent for the spurious regression. As discussed, the
regression (2) becomes a spurious regression, if there is no common persistent stochastic trend between \(Y\) and \(X\).

We let \(Z = Y\) or any component in \(X\), and write \(Z = Z^c + Z^d\) similarly as earlier, where \(Z^c\) is the continuous part of \(Z\) and \(Z^d\) is the jump part of \(Z\) defined as \(Z^d_t = \sum_{0 \leq s \leq t} \Delta Z_s\) with \(\Delta Z_t = Z_t - Z_{t-}\). We assume that

**Assumption 4.1** Let \(Z\) a persistent stationary process such that
(a) \(\sup_{|t-s| \leq \delta} |Z^c_t - Z^c_s| = O_p(\delta^\kappa a_T)\) for all \(0 \leq s, t \leq T\) with some \(\kappa > 0\) and \(a_T\),
(b) \(\mathbb{E}\left(\sum_{0 \leq t \leq T} |\Delta Z_t|\right)^2 = O(\delta^2)\), and
(c) \(\sup_{0 \leq t \leq T} |Z_t| = O_p(c_T)\) for some \(c_T\).

The conditions in Assumption 4.1 are not stringent. As discussed earlier, many of persistent stationary processes that are of interest in economic applications are of stochastically bounded, as required in condition (a). Also, condition (b) is expected to hold widely for continuous processes. As can be seen clearly in our proofs, we may allow for unbounded \(X\) and \(Y\) in condition (a), if we require stronger condition in (b) for the modulus of continuity of \(X\) and \(Y\). Throughout this section, we let Assumption 4.1 hold.

**Lemma 4.1** Let Assumption 4.1 hold. Then we have
\[
\frac{1}{n} \sum_{i=1}^{n} x_i x'_i = \frac{1}{T} \int_{0}^{T} X_t X'_t dt + O_p(\delta^\kappa a_T) + O_p\left(\frac{\delta b_T c_T}{T}\right)
\]
\[
\frac{1}{n} \sum_{i=1}^{n} x_i y_i = \frac{1}{T} \int_{0}^{T} X_t Y'_t dt + O_p(\delta^\kappa a_T) + O_p\left(\frac{\delta b_T c_T}{T}\right)
\]
as \(\delta \to 0\) and \(T \to \infty\).

To establish the asymptotics for regression (2) with co-persistence as specified in (3), we further assume that

**Assumption 4.2** We let
\[
\frac{1}{T} \int_{0}^{T} U^2_t dt \to_p \sigma^2 > 0
\]
as \(T \to \infty\).

**Assumption 4.3** If we define a process \(U_T = (U^T_t)\) on the unit interval \([0, 1]\) by
\[
U^T_t = \frac{1}{\sqrt{T}} \int_{0}^{tT} U_s ds,
\]
then we have \(U_T \to_d \bar{U}\) as \(T \to \infty\), where \(\bar{U}\) is Brownian motion with variance
\[
\omega^2 = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left(\int_{0}^{T} U_t dt\right)^2 > 0,
\]
which is assumed to exist.

Assumptions 4.2 and 4.2 are continuous time analogues of standard assumptions we routinely impose for the analysis of discrete time series models. The condition in Assumption 4.2 is not stringent and holds, for instance, for all stationary ergodic processes with finite second moment. The weak convergence of $U^T$ to $\bar{U}$ in Assumption 4.3 is a continuous time version of the invariance principle for partial sum process in discrete time that is widely used in analyzing asymptotic behaviors of models with unit root nonstationary time series. Therefore, it will be referred to simply as continuous time invariance principle. The continuous time invariance principle we invoke here is expected to hold for a wide class of mean zero stationary processes. For instance, it holds for general positive recurrent diffusion processes, as shown in Theorem 2.6 of Bhattacharya (1982).

For regression (2), we define $\hat{\beta}$ to be the OLS estimator of $\beta$. Moreover, we denote by $F_{R}(\hat{\beta})$ the standard Wald statistic for the null hypothesis $H_{0}: R\beta = r$, where $R$ and $r$ are the restriction matrix and vector respectively of dimensions $q \times m$ and $q$. For notational simplicity, we signify by $F(\hat{\beta})$ the Wald statistic for the null hypothesis $H_{0}: \beta = 0$. Below, we present the asymptotics of $\hat{\beta}$, $F_{R}(\hat{\beta})$ and $F(\hat{\beta})$ for the co-persistent regression and the spurious regression.

For the co-persistent regression, we have

**Theorem 4.2 (Co-Persistent Regression)** Let $Y$ and $X$ be persistent stationary processes satisfying Assumption 4.1 with

$$\delta = o \left( \min \left( \frac{1}{(a_Tb_T\sqrt{T})^{1/\alpha}}, \frac{\sqrt{T}}{b_Tc_T} \right) \right).$$

If they are generated as in (3) and $U$ satisfies Assumption 4.2, we have

$$\sqrt{T}(\hat{\beta} - \beta) \to_d \left( \int_{0}^{1} \bar{X}_t\bar{X}_t' dt \right)^{-1} \int_{0}^{1} \bar{X}_t d\bar{U}_t$$

and, under the null hypothesis,

$$\delta F_{R}(\hat{\beta}) \to_d \frac{1}{\sigma^2} \int_{0}^{1} d\bar{U}_t \bar{X}_t' \left( \int_{0}^{1} \bar{X}_t\bar{X}_t' dt \right)^{-1} \int_{0}^{1} \bar{X}_t d\bar{U}_t \int_{0}^{1} \bar{X}_t' d\bar{U}_t$$

and

$$\left( \int_{0}^{1} \bar{X}_t\bar{X}_t' dt \right)^{-1} \int_{0}^{1} \bar{X}_t d\bar{U}_t$$

as $T \to \infty$.

For the co-persistent regression, the OLS estimator $\hat{\beta}$ of $\beta$ is consistent at the rate of $\sqrt{T}$. Note that the convergence rate does not depend upon $\delta$. Therefore, in particular, $\hat{\beta}$ does not converge to $\beta$, if we fix $T$ and just collect samples more frequently over the fixed time span to increase the sample size. The limit distribution of $\hat{\beta}$ is in general non-Gaussian.
It becomes Gaussian only when the limit process $\tilde{X}$ of the regressor becomes independent of the limit process $\tilde{U}$ of the regression error. If they are independent of each other, the limit distribution of $\hat{\beta}$ reduces to

$$\sqrt{T}(\hat{\beta} - \beta) \to_d MN \left(0, \omega^2 \left(\int_0^1 \tilde{X}_t \tilde{X}_t' dt\right)^{-1}\right)$$

and becomes mixed-normal.

Under the null hypothesis, the Wald statistic $F_R(\hat{\beta})$ diverges as $\delta \to 0$, and the size of the Wald test based on $F_R(\hat{\beta})$ increases up to unity in the limit. The Wald test would therefore reject the null hypothesis with probability one as $\delta$ decreases down to zero, even if the null hypothesis is true. Therefore, the test should not be used with high frequency observations. The limit distribution of $F_R(\hat{\beta})$ is generally non-Gaussian. As for $\hat{\beta}$, it reduces normal if the limit process $\tilde{X}$ of the regressor becomes independent of the limit process $\tilde{U}$ of the regression error. In this case, we have

$$\delta F_R(\hat{\beta}) \to_d \frac{\omega^2 T}{\sigma^2} \chi_q^2$$

as $T \to \infty$. This can be readily deduced from (4).

For the spurious regression, we have

**Theorem 4.3 (Spurious Regression)** Let $Y$ and $X$ be persistent stationary processes satisfying Assumption 4.1 with

$$\delta = o \left( \min \left( \frac{1}{(a_T b_T)^{1/\kappa}}, \frac{T}{b_T c_T} \right) \right).$$

If they are independent of each other, we have

$$\hat{\beta} \to_d \left( \int_0^1 \tilde{X}_t \tilde{X}_t' dt \right)^{-1} \int_0^1 \tilde{X}_t \tilde{Y}_t dt$$

and

$$\frac{\delta}{T} F(\hat{\beta}) \to_d \frac{\int_0^1 \tilde{Y}_t \tilde{X}_t' dt \left( \int_0^1 \tilde{X}_t \tilde{X}_t' dt \right)^{-1} \int_0^1 \tilde{X}_t \tilde{Y}_t dt}{\int_0^1 \tilde{Y}_t^2 dt - \int_0^1 \tilde{Y}_t \tilde{X}_t' dt \left( \int_0^1 \tilde{X}_t \tilde{X}_t' dt \right)^{-1} \int_0^1 \tilde{X}_t \tilde{Y}_t dt}$$

as $T \to \infty$.

The spurious regression between independent persistent stationary processes yields the same results as in the spurious regression resulting from two independent random walks. The OLS estimator $\hat{\beta}$ of $\beta$ is inconsistent and remains to be random as $T \to \infty$. Moreover, the Wald test based on the Wald statistic $F(\hat{\beta})$ diverges at the rate of $n = T/\delta$. We may therefore expect that the actual size of the Wald test increases up to unity as $n = T/\delta$ gets large. We would therefore be led to falsely reject the null hypothesis of no relationship between $Y$ and $X$, which are generated independently from each other.
Mathematical Appendix

Proof of Theorem 3.1  We denote by \( c \) the Lipschitz constant for the kernel function \( K \), and note that

\[
\left| \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x_i - x}{h} \right) \right| - \frac{1}{Th} \int_0^T K \left( \frac{X_t - x}{h} \right) dt \leq \frac{1}{Th} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left| K \left( \frac{X_t - x}{h} \right) - K \left( \frac{X_{i\delta} - x}{h} \right) \right| dt
\]

\[
\leq c \sup_{|t-s| \leq \delta} |X^c_t - X^c_s| + \frac{c\delta}{Th^2} \sum_{0 \leq t \leq T} |\Delta X_t|
\]

\[
= O_p \left( \frac{\delta^a T}{h^2} \right) + O_p \left( \frac{\delta b T}{h^2 T} \right) = o_p(1)
\]
due to conditions (a), (b) and (c) of Assumption 3.1. However, we have

\[
\frac{1}{Th} \int_0^T K \left( \frac{X_t - x}{h} \right) dt = \frac{1}{Th} \int_{-\infty}^\infty K \left( \frac{y - x}{h} \right) \ell(T, y) dy
\]

\[
= \int_{-\infty}^\infty K(y) \ell(T, x + hy) \frac{dy}{T}
\]

\[
= \int_{-\infty}^\infty K(y) \ell^T(1, x + hy) dy
\]

\[
= \ell^T(1, x) + o_{a.s.}(1)
\]

\[
\rightarrow d \ell(1, x)
\]

by the successive applications of occupation times formula, change of variables and dominated convergence, using condition (d) of Assumption 3.1. The stated result now follows immediately and the proof is complete.

\[ \square \]

Proof of Lemma 4.1  We have

\[
\left| \frac{1}{n} \sum_{i=1}^{n} Z^2_{i\delta} - \frac{1}{T} \int_0^T Z^2_t dt \right| \leq \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} |Z^2_t - Z^2_{i\delta}| dt
\]

\[
\leq 2 \left( \sup_{0 \leq t \leq T} |Z_t| \right) \left( \sup_{|t-s| \leq \delta} |Z^c_t - Z^c_s| + \frac{\delta}{T} \sum_{0 \leq t \leq T} |\Delta Z_t| \right)
\]

\[
= O_p(\delta^a T c_T) + O_p \left( \frac{\delta b c_T}{T} \right)
\]

as was to be shown, under conditions (a), (b) and (c) of Assumption 4.1.

\[ \square \]
Proof of Theorem 4.2 Write
\[ \sqrt{T}(\hat{\beta} - \beta) = \left( \frac{1}{n} \sum_{i=1}^{n} x_i x'_i \right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i. \]
We have
\[ \frac{1}{n} \sum_{i=1}^{n} x_i x'_i = \frac{1}{T} \int_{0}^{T} X_t X'_t dt + o_p(1) \]
due to Lemma 4.1, under the given condition for \( \delta \), and we have
\[ \frac{1}{T} \int_{0}^{T} X_t X'_t dt \to_d \int_{0}^{1} \bar{X}_t \bar{X}'_t dt \]
by the definition of persistent stationary process and the continuous mapping theorem. Therefore, the limit distribution of \( \hat{\beta} \) readily follows from the continuous mapping theorem, if we show that
\[ \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \to_d \int_{0}^{1} \bar{X}_t d\bar{U}_t \]
as \( \delta \to 0 \) and \( T \to \infty \). Note that
\[ \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i = \sqrt{T} \frac{\delta}{T} \sum_{i=1}^{n} x_i u_i = \frac{1}{\sqrt{T}} \int_{0}^{T} X_t U_t dt + o_p(1), \]
due to Lemma 4.1 and the given condition for \( \delta \). However, we have
\[ \frac{1}{\sqrt{T}} \int_{0}^{T} X_t U_t dt = \frac{1}{\sqrt{T}} \int_{0}^{1} X_{Tt} U_{Tt} d(Tt), \]
and it follows from the definitions of the processes \( X^T \) and \( U^T \) that
\[ X^T_t = X_{Tt}, \quad dU^T_t = \frac{1}{\sqrt{T}} U_{Tt} d(Tt). \]
Therefore, we have
\[ \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i = \int_{0}^{1} X^T_t dU^T_t + o_p(1) \to_d \int_{0}^{1} \bar{X}_t d\bar{U}_t, \]
due to, e.g., Kurtz and Protter (1991). This was to be shown. \( \Box \)

Proof of Theorem 4.3 The stated results follow immediately from the continuous mapping theorem, once we note that
\[ \frac{1}{n} \sum_{i=1}^{n} x_i x'_i = \frac{1}{T} \int_{0}^{T} X_t X'_t dt + o_p(1) \]
and
\[ \frac{1}{n} \sum_{i=1}^{n} x_i y_i = \frac{1}{T} \int_{0}^{T} X_t Y_t dt + o_p(1) \]
due to Lemma 4.1, under the given condition for \( \delta \). The details are therefore omitted. \( \Box \)