Understanding Regressions with Observations Collected at High Frequency over Long Span*

Yoosoon Chang
Department of Economics
Indiana University

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

March 11, 2014

Abstract

In this paper, we consider regressions with observations collected at small time interval over long period of time. For the formal asymptotic analysis, we assume that samples are obtained from continuous time stochastic processes, and let the sampling interval δ shrink down to zero and the sample span T increase up to infinity. In this set-up, we show that the standard Wald statistic always diverges to infinity as long as $\delta \to 0$ sufficiently fast relative to $T \to \infty$, and regressions become spurious. This is indeed well expected from our asymptotics which shows that, in such a set-up, samples from any continuous time process become strongly dependent with their serial correlation approaching to unity, and regressions become spurious exactly as in the conventional spurious regression. However, as we show in the paper, the spuriousness of Wald test disappears if we account for strong persistency adequately using an appropriate longrun variance estimate. The empirical illustrations in the paper provide a strong and unambiguous support for the practical relevancy of our asymptotic theory.

JEL Classification: C13, C22

Keywords: high frequency regression, spurious regression, continuous time model, asymptotics

*We are grateful to Kyungsik Nam for his excellent research assistance. We would also like to thank Jihyun Kim and Ye Lu for helpful discussions.

1 Introduction

A great number of economic and financial time series are now collected and made available at high frequencies, and many empirical researchers find it difficult to decide at what frequency they collect the samples to estimate and test their models. Naturally we may think that we should use all available observations, since neglecting any available observations means a loss in information. Nevertheless, this is not the usual practice in applied empirical research. In most cases, samples used in practical applications are obtained at a frequency lower than the maximum frequency available. For instance, many time series models in financial economics are fitted using monthly observations, when their daily samples or even intra-day samples are available at no extra costs. Some researchers seem to believe, rather vaguely, that high frequency observations include excessive noise or erratic volatilities, and they do not bring in any significant amount of marginal information. Others keep silent on this issue, and simply choose the sampling frequency that yields sensible results.

In the paper, we formally investigate the effect of sampling frequency on the standard tests in regressions. For our analysis, we consider the standard regression model for continuous time stochastic processes, and assume that the regression is fitted by discrete time observations collected at varying time intervals. It is supposed that the discrete samples are collected at sampling interval δ over sample span T, and we let $\delta \to 0$ and $T \to \infty$ jointly to establish our asymptotics. Our asymptotics are therefore more relevant to regressions with observations collected at high frequency over long span. Both stationary and nonstationary continuous time regression models are analyzed. The former is the continuous time version of the standard stationary time series regression, whereas the latter is a continuous time analogue of the cointegrating regression model. Our assumptions are very mild and accommodate a wide class of regression models, and therefore, our asymptotics are applicable for virtually all regression models that are used in practical applications.

The most important finding from our analysis is that both type of regressions become spurious eventually as the sampling frequency increases. Even under the correct null hypothesis, the standard test statistics, such as the t-ratios and Wald statistics, increase up to infinity as the sampling interval decreases down to zero. Therefore, they would always lead us to reject the correct null hypothesis if the sampling interval is sufficiently small. This is completely analogous to the conventional spurious regression in econometrics, which was first studied through simulations by Granger and Newbold (1974) and studied theoretically later by Phillips (1986). The spuriousness in the conventional spurious regression is due to the presence of a unit root in the regression error that generates strong serial correlation. The same problem arises in the regressions we consider. The regression error from a continuous time process becomes strongly correlated as the sampling interval decreases, which yields the same type of spuriousness in the conventional spurious regression.

However, there is an important difference between our regressions and the conventional spurious regression. In contrast to the conventional spurious regression that does not represent any meaningful relationship, our regressions are well specified as signifying authentic relationships. Naturally, the spuriousness of our regressions is rectifiable. In fact, we show in the paper that the spuriousness of the t-tests or Wald tests disappears if we account for strong persistency in high frequency observations using an appropriate longrun variance estimate, in lieu of the usual variance estimate, to define the statistics. The longrun variance estimate takes all serial correlations in the samples into account, and therefore, it may effectively deal with the strong correlation in the regression errors of our regressions if it is used properly. Indeed, the usual HAC estimator obtained with the automatic bandwidth selection procedure proposed by Andrews (1991) works for our high frequency regressions, and the standard tests with such HAC estimators have well defined limit distributions.

The rest of the paper is organized as follows. Section 2 explains the background and motivation of our analysis in the paper. In particular, we provide some illustrative examples that are analyzed throughout the paper to show the practical relevancy of our asymptotic theory. Section 3 introduces the regression models, the set-up for our asymptotics and some preliminaries. The spuriousness of the high frequency regressions are derived and investigated in Section 4. In particular, we establish under fairly general conditions that the coefficient in the first order autoregression of the regression error converges to unity as the sampling interval decreases. Section 5 presents the limit theory for the modified test statistics defined with HAC estimators. We also demonstrate that the bandwidth selection is important, and the modified tests may or may not yield spurious results depending upon the bandwidth choice. Section 6 concludes the paper, and Appendices have mathematical proofs and additional figures.

2 Background and Motivation

It is widely observed that test results are critically dependent upon the choice of sample frequency in many time series regressions. To illustrate more explicitly the dependency on sampling interval of the test results, we consider a simple bivariate regression of (y_i) on (x_i) written as $y_i = \alpha + \beta x_i + u_i$, where α and β are respectively the intercept and slope parameters and (u_i) are the regression errors. For (y_i) and (x_i), we consider the following four pairs:

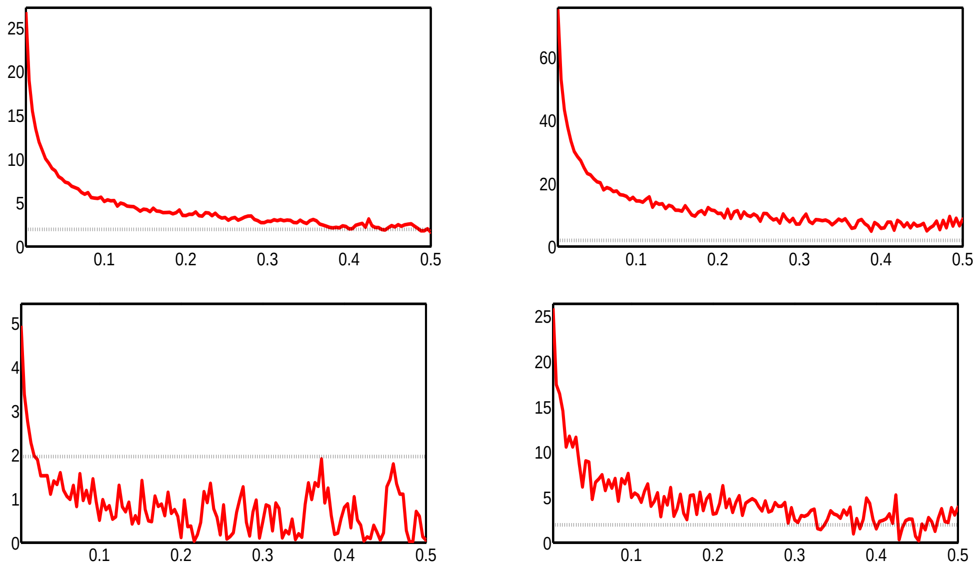
Model	(y_i)	(x_i)
I	20-Year T-bond rates	3-month T-bill rates
II	3-month eurodollar rates	3-month T-bill rates
III	log US/UK exchange rates forward	log US/UK exchange rates spot
IV	log SP500 index futures	log SP500 index

The plots for (y_i) and (x_i) in Models I-IV are given in Figure 5 in Appendix B. In all models, possibly except for Model I, two series (y_i) and (x_i) move very closely with each other. Therefore, the most natural hypotheses to be tested appear to be $\alpha = 0$ and $\beta = 1$. The hypotheses may not hold. In particular, the null hypothesis $\alpha = 0$ does not necessarily hold, in case where there are differences in the term or liquidity premium as well as the general risk premium between two assets represented by (y_i) and (x_i). However, we do not intend to provide any answers to whether or not the hypotheses should hold in any of the models we consider here. Instead, we simply analyze the dependence of test results on the sampling frequency. In what follows, we will mainly focus on testing for the null hypothesis $\beta = 1$, which seems to be much less controversial.

In Figure 1, we present the values of t-ratios for testing the null hypothesis $\beta = 1$ for Models I-IV. In the figure, we plot the t-values against various sampling intervals from six months with $\delta = 1/2$ to one day with $\delta = 1/250$ in yearly unit. As discussed, the t-values change extensively as the sampling interval δ varies. In particular, they tend to increase very rapidly as δ gets smaller and becomes near zero. The rate of increase in the t-values as δ approaches to zero varies across different models. However, it is common to all models that the t-values start to increase sharply as the sampling interval becomes approximately one month or shorter, and eventually at daily frequency, the t-tests in all models unambiguously and unanimously reject the null hypothesis $\beta = 1$. In contrast, the t-tests yield some mixed results for the same hypothesis as the sampling interval moves away from a neighborhood of zero and further increases. We have very similar results for the t-test for the null hypothesis $\alpha = 0$, as shown in Figure 7 in Appendix B.

The dependency of the test results on the sampling frequency is of course extremely undesirable, since in most cases the hypothesis of interest is not specific to sampling frequency and we expect it to hold for all samples collected at any sampling interval. Subsequently, we consider a continuous time regression model and build up an appropriate framework to analyze this dependency of the test results on the sampling frequency. We find that what we observe here as the common feature of the t-tests is not an anomaly. From our asymptotic analysis relying on $\delta \to 0$ as well as $T \to \infty$, it actually becomes clear that the test is ex-

Figure 1: t-Tests for β in Models I-IV



Notes: Presented in the top two panels are the absolute values of t-tests for testing $\beta = 1$ in Models I and II, and in the bottom two panels those from Models III and IV. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$. Each graph plots the absolute test values across different levels of frequency parameter δ on the horizontal axis. The black dotted horizontal line signifies the two-sided 5% standard normal critical value 1.96.

pected to diverge up to infinity with probability one as δ decreases down to zero. Roughly, this happens since the serial correlation at any finite lag of discrete samples converges to unity as the sampling frequency increases if the samples are taken from continuous time stochastic processes. We may allow for the presence of jumps, if the jump activity is regular and there are only a finite number of jumps in any time interval.

3 The Model, Set-Up and Preliminaries

Consider the standard regression model

$$y_i = x_i'\beta + u_i \tag{1}$$

for $i = 1, \ldots, n$, where (y_i) and (x_i) are respectively the regressand and regressor, β is the regression coefficient and (u_i) are the regression errors. Though it is possible to analyze more general regressions, the simple model we consider here is sufficient to illustrate the main issue dealt with in the paper. Throughout, we denote by $\hat{\beta}$ the OLS estimator of β. The general linear hypothesis on β, formulated typically as $R\beta = r$ with known matrix R and vector r of conformable dimensions, is often tested using the Wald statistic defined by

$$F(\hat{\beta}) = (R\hat{\beta} - r)' \left[R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - r)/\hat{\sigma}^2, \tag{2}$$

where $\hat{\sigma}^2$ is the usual estimator for the error variance obtained from the OLS residuals $(\hat{u}_i)$. In the presence of serial correlation in (u_i), the Wald statistic introduced in (2) is in general not applicable. Therefore, in this case, modified versions of the Wald statistics such as

$$G(\hat{\beta}) = (R\hat{\beta} - r)' \left[R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - r)/\hat{\omega}^2, \tag{3}$$

where $\hat{\omega}^2$ is a consistent estimator for the longrun variance of (u_i) based on $(\hat{u}_i)$, or

$$H(\hat{\beta}) = (R\hat{\beta} - r)' \left[R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \hat{\Omega} \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - r), \tag{4}$$

where $\hat{\Omega}$ is a consistent estimator for the longrun variance of $(x_i u_i)$ based on $(x_i \hat{u}_i)$. Subsequently, we consider two different types of regression given by (1): stationary type regression and cointegration type regression. The test based on (4) is generally more appropriate for the stationary type regression, whereas only the test based on (3) is sensible for the cointegration type regression.[1]

In the paper, we analyze regression (1), when (y_i) and (x_i) are high frequency observations.[2] Therefore, for the subsequent analysis, we let (y_i) and (x_i) be samples collected at discrete time intervals from the underlying continuous time processes denoted respectively by $Y = (Y_t)$ and $X = (X_t)$, i.e.,

$$y_i = Y_{i\delta} \quad and \quad x_i = X_{i\delta}$$

for $i = 1, \ldots, n$ be discrete samples from the continuous time processes Y and X over time $[0, T]$ collected at the sampling interval with length $\delta > 0$, where $T = n\delta$. Under our setup,

[1] Note that the longrun variance of $(x_i u_i)$ does not exist if (x_i) is nonstationary.

[2] In the paper, high frequency observations are defined to be samples collected at sampling intervals which are small relative to their time span. For instance, five years of daily observations are considered to be high frequency observationss.

it is clear that we may define the continuous time regression

$$Y_t = X_t'\beta + U_t \tag{5}$$

for $0 \leq t \leq T$ corresponding to the regression model introduced in (1), where Y and X are the regressand and regressor processes, and $U = (U_t)$ is the error process, from which (u_i) are defined similarly as (y_i) and (x_i) are defined from Y and X.

For any stochastic process $Z = (Z_t)$ appearing in the paper, we assume that $Z = Z^c + Z^d$, where Z^c is the continuous part and Z^d the jump part defined as $Z_t^d = \sum_{0\leq s\leq t} \Delta Z_s$ with $\Delta Z_t = Z_t - Z_{t-}$.

Assumption A. Let Z be any element in U, XX' or XU. We have

$$\sum_{0\leq t\leq T} \mathbb{E}|\Delta Z_t| = O(T).$$

Moreover, if we define $\Delta_{\delta,T}(Z) = \sup_{0\leq s,t\leq T} \sup_{|t-s|\leq\delta} |Z_t^c - Z_s^c|$, then

$$\max\left(\delta, \frac{\delta}{T} \sup_{0\leq t\leq T} |Z_t|\right) = O\Big(\Delta_{\delta,T}(Z)\Big)$$

as $\delta \to 0$ and $T \to \infty$.

The conditions in Assumption A is very mild and expected to be satisfied by a wide class of stochastic processes. The first condition is crucial in our asymptotic analysis. However, it is not stringent and met, for instance, for all processes with compound Poisson type jumps as long as their sizes are bounded in L^1 and their intensity is proportional to T. The second condition is just made to simplify our exposition. It is not critical and holds trivially for virtually all stochastic processes used in practical applications. Typically, we have

$$\Delta_{\delta,T}(Z) = \delta^{1/2-\epsilon}\lambda_T \tag{6}$$

for some $\epsilon \geq 0$ and a nonrandom sequence (λ_T) of T that is bounded away from zero and, as an example, the condition is clearly satisfied if $\sup_{0\leq t\leq T} |Z_t| = O_p(T)$. If T is fixed, $\Delta_{\delta,T}(Z)$ represents the usual modulus of continuity of the stochastic process Z. On the other hand, we let $T \to \infty$ in our set-up and therefore it may be regarded as the global modulus of continuity.

The following lemma allows us to approximate the sample moments in discrete time by the corresponding sample moments in continuous time. Here and elsewhere in the paper, we use $\|\cdot\|$ to denote the Euclidian norm for a vector or a matrix.

Lemma 3.1. *Let Assumption A hold. If we define $Z = U, XX'$ or XU and $z_i = Z_{i\delta}$ for $i = 1, \ldots, n$, we have*

$$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{T}\int_0^T Z_t dt + O_p\Big(\Delta_{\delta,T}(\|Z\|)\Big)$$

for all small δ and large T.

In our subsequent analysis, we impose a set of sufficient conditions to ensure the asymptotic negligibility of the approximation error $\Delta_{\delta,T}(\|Z\|)$, for $Z = U, XX'$ and XU, so that we may approximate all relevant sample moments by their continuous analog without affecting their asymptotics. Once the approximations are made, the rest of our asymptotics rely entirely on the asymptotics of moments in continuous time. This will be introduced below.

Assumption B. $T^{-1}\displaystyle\int_0^T U_t^2 dt \to_p \sigma^2$ for some $\sigma^2 > 0$ as $T \to \infty$.

Needless to say, Assumption B holds for a wide variety of asymptotically stationary stochastic processes.

As discussed, we consider two different types of regressions. Below we introduce assumptions for each of these regressions. We denote by $D[0,1]$ the space of cadlag functions endowed with the usual Skorohod topology.

Assumption C1. We assume that
(a) $T^{-1}\int_0^T X_t X_t' dt \to_p M$ as $T \to \infty$ for some nonrandom matrix $M > 0$, and
(b) we have

$$T^{-1/2}\int_0^T X_t U_t dt \to_d \mathbb{N}(0, \Pi)$$

as $T \to \infty$, where $\Pi = \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\int_0^T X_t U_t dt\right)\left(\int_0^T X_t U_t dt\right)' > 0$, which is assumed to exist.

Assumption C2. We assume that
(a) for X^T defined on $[0,1]$ with an appropriately defined nonsingular normalizing sequence (Λ_T) of matrices as

$$X_t^T = \Lambda_T^{-1} X_{Tt},$$

we have $X^T \to_d X^\circ$ in the product space of $D[0,1]$ as $T \to \infty$ with linearly independent limit process X°, and

(b) if we define U^T on $[0,1]$ as

$$U_t^T = T^{-1/2} \int_0^{Tt} U_s ds,$$

then $U^T \to_d U^\circ$ in $D[0,1]$ as $T \to \infty$, where U° is Brownian motion with variance $\pi^2 = \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\int_0^T U_t dt\right)^2 > 0$, which is assumed to exist.

Both Assumptions C1 and C2 are expected to hold for a wide class of stationary and nonstationary regressions. Assumption C1 is the continuous analog of the standard assumptions for stationary regressions in discrete time. Assumption C2(a) is satisfied for general null recurrent diffusions and jump diffusions, as shown by Jeong and Park (2011), Jeong and Park (2014) and Kim and Park (2014). Moreover, Assumption C2(b) is the continuous time version of the usual invariance principle. In parallel with Assumptions C1 and C2, respectively for the stationary and cointegrating regressions, we introduce Assumptions D1 and D2 below.

Assumption D1. $\Delta_{\delta,T}(U), \Delta_{\delta,T}(\|XX'\|) \to 0$ and $\sqrt{T}\Delta_{\delta,T}(\|XU\|) \to 0$ as $\delta \to 0$ and $T \to \infty$.

Assumption D2. $\Delta_{\delta,T}(U), \|\Lambda_T\|^2\Delta_{\delta,T}(\|XX'\|) \to 0$ and $\sqrt{T}\|\Lambda_T\|\Delta_{\delta,T}(\|XU\|) \to 0$ as $\delta \to 0$ and $T \to \infty$.

In our asymptotic analysis, we let $\delta \to 0$ and $T \to \infty$ jointly, satisfying Assumption D1 or D2. Our asymptotics are joint, not sequential, in δ and T. We allow $\delta \to 0$ and $T \to \infty$ jointly, as long as δ and T satisfy an appropriate condition in Assumption D. However, in all these assumptions, we require $\delta \to 0$ sufficiently fast relative to $T \to \infty$. It is therefore expected that our joint asymptotics yield the same results as the sequential asymptotics relying on $\delta \to 0$ followed by $T \to \infty$.

4 Spuriousness of Regression at High Frequency

In this section, we establish the asymptotics of OLS estimator $\hat{\beta}$ and analyze the asymptotic behaviors of the standard Wald statistic $F(\hat{\beta})$ under the null hypothesis.

Theorem 4.1. *Assume $R\beta = r$ and let Assumptions A and B hold.*

(a) Under Assumption C1, we have

$$\sqrt{T}(\hat{\beta}-\beta) \to_d N$$
$$\delta F(\hat{\beta}) \to_d N'R'\left(RM^{-1}R'\right)^{-1} RN/\sigma^2,$$

where $N =_d \mathbb{N}\left(0, M^{-1}\Pi M^{-1}\right)$*, as* $\delta \to 0$ *and* $T \to \infty$ *satisfying Assumption D1.*
(b) Under Assumptions C2, we have

$$\sqrt{T}\Lambda_T'(\hat{\beta}-\beta) \to_d P$$
$$\delta F(\hat{\beta}) \to_d P'R'\left(RQ^{-1}R'\right)^{-1} RP/\sigma^2$$

where $P = \left(\int_0^1 X_t^\circ X_t^{\circ\prime} dt\right)^{-1} \int_0^1 X_t^\circ dU_t^\circ$ *and* $Q = \int_0^1 X_t^\circ X_t^{\circ\prime} dt$*, as* $\delta \to 0$ *and* $T \to \infty$ *satisfying Assumption D2.*

For both stationary and cointegration type regressions, the OLS estimator $\hat{\beta}$ is generally consistent for β under our asymptotics relying on $\delta \to 0$ sufficiently fast relative to $T \to \infty$. It is crucial that we have $T \to \infty$ for the consistency of $\hat{\beta}$. If, for instance, T is fixed, $\delta \to 0$ alone is not sufficient for its consistency. On the other hand, for both stationary and cointegration type regressions, we have

$$F(\hat{\beta}) \to_p \infty$$

as $\delta \to 0$ and $T \to \infty$. This implies that the Wald test would always lead us to reject the null hypothesis when it is correct, and the asymptotic size would become unity. The regressions therefore become spurious.

It is easy to see why this happens. Suppose that the law of large numbers and the central limit theorem hold for U, as we assume in Assumption C1 or C2. Moreover, we let Assumption A hold for U, and set $\Delta_{\delta,T}(U) \to_p 0$ or more strongly $\sqrt{T}\Delta_{\delta,T}(U) \to_p 0$ if needed as $\delta \to 0$ and $T \to \infty$. We may easily deduce that

$$\frac{1}{n}\sum_{i=1}^n u_i = \frac{1}{T}\int_0^T U_t dt + o_p(1) \to_p 0$$

as $n \to \infty$ (with $\delta \to 0$ and $T \to \infty$), and therefore, the law of large numbers holds for (u_i). However, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i = \frac{1}{\sqrt{\delta}}\left[\frac{1}{\sqrt{T}}\int_0^T U_t dt + o_p(1)\right] \to_p \infty$$

as $n \to \infty$ (with $\delta \to 0$ and $T \to \infty$), and consequently, the central limit theory fails to hold for (u_i). In fact, in our setup, (u_i) becomes strongly dependent as $\delta \to 0$, since the correlation between u_i and u_{i-j} for any i and j becomes unity as $\delta \to 0$. Therefore, it is well expected that the central limit theory does not hold for (u_i).

Our results here are very much analogous to those from the conventional spurious regression, which was first investigated through simulations by Granger and Newbold (1974) and explored later analytically by Phillips (1986). As is now well known, the regression of two independent random walks, or more generally, integrated time series with no cointegration, yields spurious results, and the Wald statistic for testing no longrun relationship diverges to infinity, implying falsely the presence of cointegration. Granger and Newbold (1974) originally suggest that this is due to the existence of strong serial dependence in the regression error. On the other hand, we show in the paper that an authentic relationship in stationary time series or the presence of cointegration among nonstationary time series is always rejected if the test is based on the Wald statistic relying on observations collected at high frequencies. Our spurious regression here is therefore in contrast with the conventional spurious regression. True relationship is rejected and tested to be false in the former, while false relationship is rejected and tested to be true in the latter. However, our regression and the conventional spurious regression have the same reason why they yield nonsensical results: They both have the regression errors that are strongly dependent, and the central limit theory does not hold for them.

To further analyze the serial dependency in (u_i), we consider the AR(1) regression

$$u_i = \rho u_{i-1} + \varepsilon_i, \tag{7}$$

and introduce some additional assumptions

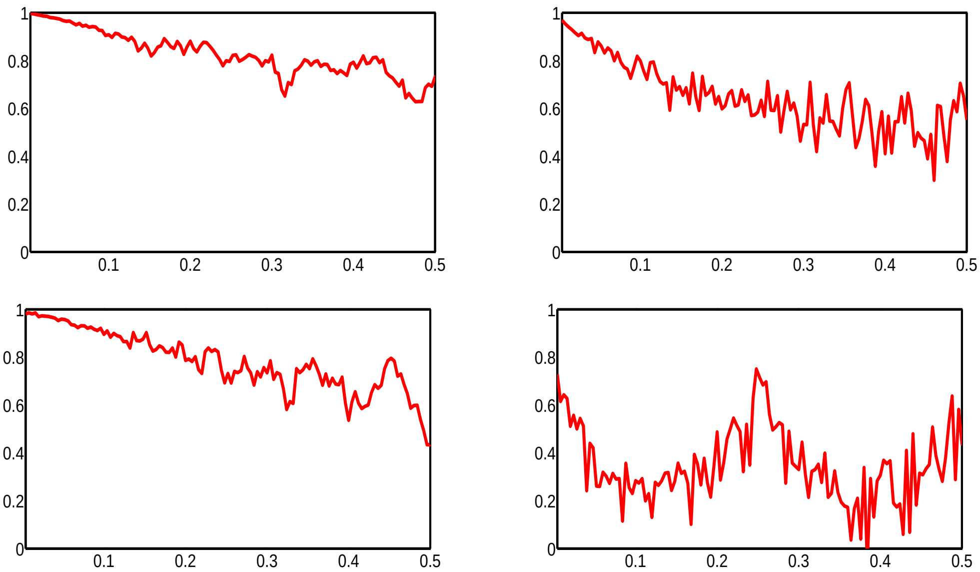
Assumption E. (a) We let U^c, the continuous part of U, be a semimartingale given by $U^c = A + B$, where A and B are respectively the bounded variation and martingale components of U^c satisfying

$$\sup_{0 \le s,t \le T} \frac{|A_t - A_s|}{|t-s|} = O_p(a_T) \quad \text{and} \quad \sup_{0 \le s,t \le T} \frac{\big|[B]_t - [B]_s\big|}{|t-s|} = O_p(b_T),$$

and $a_T \Delta_{\delta,T}(U) \to 0$ and $(b_T/\sqrt{T})\Delta_{\delta,T}(U) \to 0$ with $\delta = \Delta^2_{\delta,T}(U)$ as $\delta \to 0$ and $T \to \infty$. (b) Moreover, we assume that $\sum_{0 \le t \le T} \mathbb{E}(\Delta U_t)^4 = O(T)$ and $T^{-1}[U]_T \to_p \tau^2$ for some $\tau^2 > 0$ as $T \to \infty$.

Note that $[U]_T = [U^c]_T + \sum_{0 \le t \le T} (\Delta U_t)^2$ and $[U^c]_T = [B]_T$.

Figure 2: Estimated Residual AR Coefficients for Models I-IV



Notes: Figure 2 plots the estimated autoregressive coefficient in the first order autoregression of the fitted regression error from Models I-IV against various sampling intervals δ on the horizontal axis, from six month with $\delta = 1/2$ to one day with $\delta = 1/250$ in yearly unit. Two two panels present the estimated AR coefficients from Models I and II, and the bottom two those from Models III and IV.

Lemma 4.2. *Under Assumption E, we have*

$$\tilde{\rho} = 1 - \frac{\delta\tau^2}{\sigma^2} + o_p(\delta)$$

as $\delta \to 0$ *and* $T \to \infty$.

It follows immediately from Lemma 4.2 that

$$\tilde{\rho} \to_p 1$$

as $\delta \to 0$ (and $T \to \infty$ satisfying Assumption E(a)). Therefore, (u_i) becomes strongly dependent, and the regression becomes spurious as the sampling interval δ approaches to 0. Our regression is completely analogous to the conventional spurious regression, except that we let $\delta \to 0$ in contrast with the conventional spurious regression requiring $n \to \infty$. Therefore, the results in Theorem 4.1 may well be expected.

Though we let $T \to \infty$, as well as $\delta \to 0$, to get more explicit limit of $\tilde{\rho}$ as in Lemma 4.2, the condition $T \to \infty$ is not essential for the spuriousness in regression (1). This is clear from our proof of Lemma 4.2. Of course, it is also possible to formulate and analyze the classical spurious regression in continuous time. If the underlying regression error process U is indeed nonstationary and has a stochastic trend, we have $T^{-1}\int_0^T U_t^2 dt \to_p \infty$ as $T \to \infty$. Therefore, we expect that our regression becomes spurious as long as T is large even if δ is not small. However, we assume $T^{-1}\int_0^T U_t^2 dt \to_p \sigma^2$ as $T \to \infty$, and let $\delta \to 0$ to analyze the spuriousness generated by high frequency observations. In our setup, the regression error (u_i) becomes strongly persistent and the regression becomes spurious, simply because we collect samples too frequently.

The speed at which $\tilde{\rho}$ diverges from the unity as δ increases depends on the ratio τ^2/σ^2. The larger the value of the ratio is, $\tilde{\rho}$ more quickly moves away from the unity. Roughly, τ^2 measures the mean local variation, while σ^2 represents the mean global variation of the error process U. Therefore, we may say that at a given value of δ, $\tilde{\rho}$ is more distinct from the unity if the underlying error process U has more local variation relative to its global variation. As an immediate consequence, we may well expect that the spuriousness of regression becomes less severe as the ratio gets larger, since then $\tilde{\rho}$ is more likely further away from one. Loosely put, if U fluctuate more locally compared with its overall scale, it is less likely that we have spurious results from our regression.

The actual estimates of the autoregressive coefficients for the fitted residuals from Models I-IV are plotted in Figure 2 against various values of the sampling interval. It is clearly seen that the estimates tend to increase as the sampling interval shrinks. In particular, except for Model IV, the estimates approach to unity as the decrease in sampling interval. This is exactly what we expect from Lemma 4.2. Model IV is rather exceptional. For Model IV, the estimated autoregressive coefficients do not have any monotonous increasing trend, unlike all other models. We believe that this is due to the irregular and frequent jump activities existing in stock prices. The existence of such a jump component in the underlying continuous time stochastic process is not allowed in the paper. Note in particular that we assume the jump intensity is proportional to time span.

We assume that U is stationary at least asymptotically. As a result, U cannot be a martingale if it is continuous. For a continuous martingale, we have $\mathbb{E}U_t^2 - \mathbb{E}U_0^2 = \mathbb{E}[U]_t$, and therefore, U cannot be stationary if it is not a constant process. Consequently, U must have either a jump component or a bounded variation component, as well as the martingale component, to induce the mean reversion and make it stationary. If U is the stationary Ornstein-Uhlenbeck process given by $dU_t = -aU_t + bdW_t$ with $a, b > 0$, we have $\sigma^2 = \mathbb{E}U_t^2 = b^2/2a$ and $\tau^2 = b^2$. Therefore, $\tilde{\rho} = 1 - 2a\delta$. The rate of divergence of $\tilde{\rho}$ from the unity is determined by the mean reversion parameter a. If a is large, $\tilde{\rho}$ quickly gets smaller as δ increases and less likely generates spuriousness.

Needless to say, all our analysis for (u_i) applies also to any linear combination of the vector time series $(x_i u_i)$, $(c' x_i u_i)$ for an arbitrary nonrandom vector c, say, if we assume the vector process $c'XU$ satisfies the same conditions as those we impose on U above in Assumption E.

5 Asymptotics of Modified Tests

Now we investigate the asymptotic behaviors of the modified Wald statistics, $G(\hat{\beta})$ and $H(\hat{\beta})$. To analyze the asymptotic longrun variance estimates $\hat{\omega}^2$ and $\hat{\Omega}$ in the statistics, we momentarily assume that (u_i) are observed and let $\tilde{\omega}^2$ and $\tilde{\Omega}$ denote the longrun variance estimates based on (u_i), in place of $(\hat{u}_i)$. Given the consistency of the OLS estimator $\hat{\beta}$ we established in the previous section, it is well expected that $\hat{\omega}^2$ and $\tilde{\omega}^2$, and also $\hat{\Omega}$ and $\tilde{\Omega}$, are asymptotically equivalent. The commonly used longrun variance estimators $\tilde{\omega}^2$ and $\tilde{\Omega}$ may be written as

$$\tilde{\omega}^2 = \sum_{|j| \leq m} K\left(\frac{j}{m}\right) \tilde{\gamma}(j) \quad \text{and} \quad \tilde{\Omega} = \sum_{|j| \leq m} K\left(\frac{j}{m}\right) \tilde{\Gamma}(j) \tag{8}$$

where K is the kernel function, $\tilde{\gamma}(j) = T^{-1} \sum_i u_i u_{i-j}$ and $\tilde{\Gamma}(j) = T^{-1} \sum_i x_i u_i x'_{i-j} u_{i-j}$ are the sample autocovariances, and m is the lag truncation parameter or bandwidth parameter that determines the number of sample autocovariances included in the estimators. As will be shown below, the asymptotics of $G(\hat{\beta})$ and $H(\hat{\beta})$ statistics are crucially dependent upon the choice of lag length m in the estimation of the longrun variance estimators in (8).

The longrun variance estimators $\tilde{\omega}^2$ and $\tilde{\Omega}$ are not consistent estimators for the longrun variances π^2 and Π of U and XU defined in Assumptions C1 and C2. However, for a general mean zero stationary process V and its discrete samples (v_i) with $v_i = V_{i\delta}$ for $i = 1, \ldots, n$, we have

$$\frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^{n} v_i \approx \frac{1}{\sqrt{T}} \int_0^T V_t dt$$

as shown earlier, and therefore, we may expect that

$$\delta \tilde{\omega}^2 \approx \pi^2 \qquad \text{and} \qquad \delta \tilde{\Omega} \approx \Pi.$$

In fact, under very mild regularity conditions, Lu and Park (2014) show that $\delta \tilde{\omega}^2$ and $\delta \tilde{\Omega}$ become consistent for π^2 and Π as $\delta \to 0$ and $T \to \infty$, if we choose m such that $m\delta \to \infty$ and $m/n \to 0$. Here we just assume their consistency, instead of introducing all the necessary low level assumptions, which involve the decaying rate of the autocovariance functions $\tilde{\gamma}$

and $\tilde{\Gamma}$ and the global modulus of continuity of $V = U$ or XU.

Assumption F. If $m\delta \to \infty$ and $m/n \to 0$ as $\delta \to 0$ and $T \to \infty$, then we have $\delta\tilde{\omega}^2 \to_p \pi^2$ and $\delta\tilde{\Omega} \to_p \Pi$ as $\delta \to 0$ and $T \to \infty$.

If we let $S = m\delta$, then the conditions $m\delta \to \infty$ and $m/n \to 0$ imply that $S \to \infty$ and $S/T \to 0$.

The following theorem establishes the asymptotic distributions of $G(\hat{\beta})$ and $H(\hat{\beta})$ statistics. We let q be the number of restrictions, and χ_q^2 denotes the chi-square distribution with q degrees of freedom.

Theorem 5.1. *Assume $R\beta = r$ and let Assumptions A, B and F hold.*
(a) Under Assumption C1, we have

$$H(\hat{\beta}) \to_d \chi_q^2$$

as $\delta \to 0$ and $T \to \infty$ jointly satisfying Assumption D1.
(b) Under Assumption C2, we have

$$G(\hat{\beta}) \to_d \overline{P}' R' \left(RQ^{-1}R'\right)^{-1} R\overline{P}$$

where $\overline{P} = \left(\int_0^1 X_t^\circ X_t^{\circ\prime} dt\right)^{-1} \int_0^1 X_t^\circ d\overline{U}_t^\circ$ with $U^\circ = \pi\overline{U}^\circ$ and $Q = \int_0^1 X_t^\circ X_t^{\circ\prime} dt$ using the notations in Theorem 4.1, as $\delta \to 0$ and $T \to \infty$ jointly satisfying Assumption D2.

Both of the statistics $H(\hat{\beta})$ and $G(\hat{\beta})$ have well defined limit null distributions respectively for general stationary and nonstationary regressions. This is in sharp contrast to the standard Wald statistic $F(\hat{\beta})$ that diverges to infinity. The reason is simple and obvious. As the sampling interval shrinks, we have stronger dependencies and the $H(\hat{\beta})$ and $G(\hat{\beta})$ statistics take care of these dependencies by using the longrun variance estimates instead of the standard error variance estimates. The $H(\hat{\beta})$ statistic has the standard chi-square limit null distribution for stationary regressions. On the other hand, the limit null distribution of the $G(\hat{\beta})$ statistic is generally nonnormal and nonstandard. If, however, the limit processes X° and U° are independent, then its limit null distribution reduces to chi-square distribution.

It should be emphasized that Assumption F, which is crucial to derive the asymptotics in Theorem 5.1, may not hold if the bandwidth parameter m is not carefully chosen in the usual discrete time set-up. If, for instance, we let

$$m = n^\kappa$$

for some $0 < \kappa < 1$, Assumption F does not hold and the tests based on the $H(\hat{\beta})$ and $G(\hat{\beta})$ statistics yield spurious results. In this case, we have

$$m\delta = n^\kappa \delta = T^\kappa \delta^{1-\kappa} \to 0,$$

if $\delta \to 0$ fast enough so that $\delta = o(T^{-\kappa/(1-\kappa)})$. In this case, we have $\delta\hat{\omega}^2 \to_p 0$ and $\delta\hat{\Omega} \to_p 0$, and consequently, $H(\hat{\beta}) \to_p \infty$ and $G(\hat{\beta}) \to_p \infty$, exactly as for the standard Wald statistic $F(\hat{\beta})$.

The longrun variance estimators like those in (8) are typically implemented in practical applications with an optimal choice of bandwidth parameter m. We first consider the estimator $\tilde{\omega}^2$ for the longrun variance of (u_i). The optimal bandwidth parameter m^*, which balances off the asymptotic variance and the squared asymptotic bias and hence minimizes the asymptotic mean squared error variance, is given by

$$m_\nu^* = \left(\frac{\nu K_\nu^2 C_\nu^2}{\int K(x)^2 dx} n \right)^{1/(2\nu+1)}, \tag{9}$$

where ν is the so-called characteristic exponent, $K_\nu = \lim_{x\to 0}(1 - K(x))/|x|^\nu$ and $C_\nu = \sum_j |j|^\nu \gamma(j) / \sum_j \gamma(j)$. We have $\nu = 1$ for the Bartlett kernel, and $\nu = 2$ for all other commonly used kernels such as Parzen, Tukey-Hanning and Quadratic Spectral kernels. Note that ν, K_ν, as well as $\int K(x)^2 dx$, are determined entirely by the choice of kernel function K, whereas C_ν is given by the covariance structure of the underlying random sequence (u_i). If, in particular, (u_i) is AR(1) with the autoregressive coefficient ρ, then we have

$$C_1^2 = \frac{4\rho^2}{(1-\rho)^2(1+\rho)^2}, \qquad C_2^2 = \frac{4\rho^2}{(1-\rho)^4} \tag{10}$$

for $\nu = 1, 2$. In fact, Andrews (1991) suggests that we fit AR(1) model to $(\hat{u}_i)$ and obtain the OLS estimate $\hat{\rho}$ of the autoregressive coefficient ρ and use it to estimate C_ν, $\nu = 1, 2$, in (10), and get an estimate of the optimal bandwidth parameter m^* in (9) from these estimates of C_ν, $\nu = 1, 2$.

We may follow Andrews (1991) by fitting the AR(1) regression as in (7) to obtain $\tilde{\rho}$, and use it to estimate the constants C_ν in (10). In this case, we have

$$\tilde{C}_1^2 = \frac{\tau^4}{\delta^2\sigma^4} + o_p(\delta^{-2}), \qquad \tilde{C}_2^2 = \frac{4\tau^8}{\delta^4\sigma^8} + o_p(\delta^{-4}),$$

and in particular $\tilde{C}_\nu^2 = O_p(\delta^{-2\nu})$ as $\delta \to 0$ and $T \to \infty$ satisfying Assumption E(a). With these choices of $\tilde{C}_1^2$ and $\tilde{C}_2^2$ for C_1^2 and C_2^2, the optimal bandwidth m^* in (9) becomes

$$\tilde{m}_1^* = \left(\frac{\tau^4 K_1^2}{\sigma^4 \int K(x)^2 dx} n\right)^{1/3} \delta^{-2/3}\Big(1 + o_p(1)\Big)$$
$$\tilde{m}_2^* = \left(\frac{8\tau^8 K_2^2}{\sigma^8 \int K(x)^2 dx} n\right)^{1/5} \delta^{-4/5}\Big(1 + o_p(1)\Big),$$

and therefore,

$$\tilde{m}_1^*\delta = \left(\frac{\tau^4 K_1^2}{\sigma^4 \int K(x)^2 dx} T\right)^{1/3} (1 + o_p(1)),\ \ \tilde{m}_2^*\delta = \left(\frac{8\tau^8 K_2^2}{\sigma^8 \int K(x)^2 dx} T\right)^{1/5} (1 + o_p(1))$$

as $\delta \to 0$ and $T \to \infty$ satisfying Assumption E(a). Consequently, we have $\tilde{m}_\nu^*\delta \to \infty$ and $\tilde{m}_\nu^*/n \to_p 0$, for $\nu = 1, 2$, and the conditions in Assumption F are automatically satisfied if we use the Andrews' automatic bandwidth selection based on AR(1) regression.

The choice of the optimal bandwidth in estimating the longrun variance Ω of a vector time series $(x_i u_i)$ can be made similarly as in the scalar case above, once we take a particular linear combination of $(x_i u_i)$, as in Newey and West (1994). More generally, we may introduce an arbitrary weight function to define a matrix norm we can use to measure the bias and variance of the longrun variance matrix of $(x_i u_i)$ as in Andrews (1991). The Andrews' automatic bandwidth selection procedure is typically implemented in practice using the formula (9) is given by

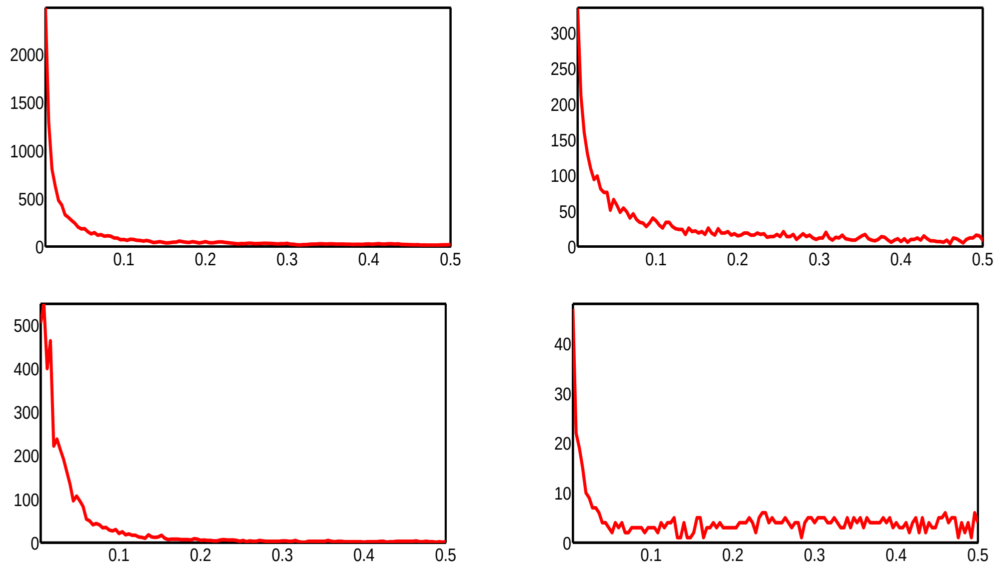
$$C_\nu^2 = \frac{\sum_{k=1}^p \Big(\sum_j |j|^\nu \gamma_k(j)\Big)^2}{\sum_{k=1}^p \Big(\sum_j \gamma_k(j)\Big)^2},$$

where $\gamma_k(j)$ is the autocovariance function of the k-th component of the p-dimensional vector time series $(x_i u_i)$, which is assumed to be AR(1) with autoregressive coefficient ρ_k, say, for $k = 1, \ldots, p$. It is straightforward to show that Assumption F holds in this case as well.

6 Concluding Remarks

To be written.

Figure 3: Estimated Optimal Bandwidth Parameters for Residuals in Models I-IV



Notes: The optimal bandwidth parameters used in the estimation of the longrun variance Ω of the bivariate time series $(\hat{u}_i, \hat{u}_i x_i)'$ from Models I and II are presented in the top two panels. Similarly, the bottom two panels present the optimal bandwidth parameters used to estimate the longrun variance ω^2 of the fitted residuals $\hat{u}_i$ from Models III and IV. The optimal bandwidth parameter is computed following the automatic selection procedure by Andrews (1991) for various sampling intervals δ, and plotted against δ on the horizontal axis, from six month with $\delta = 1/2$ to one day with $\delta = 1/250$ in yearly unit.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

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

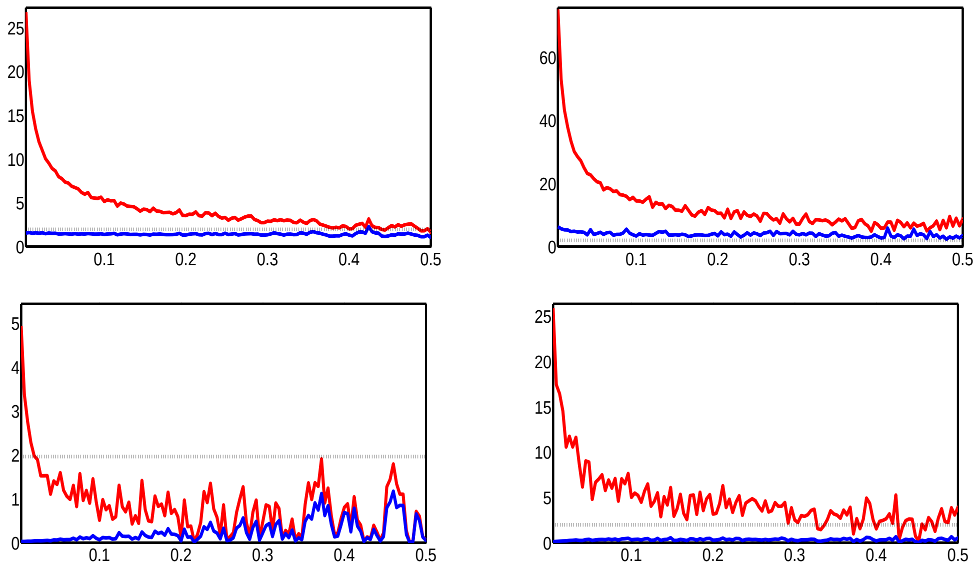
Jeong, M., and J. Y. Park (2011): "Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model," Working Paper, Indiana University.

——— (2014): "Asymptotic Theory of Maximum Likelihood Estimator for Jump Diffusion Model," Working Paper, Indiana University.

Kim, J., and J. Y. Park (2014): "Asymptotics for Recurrent Diffusions with Application to High Frequency Regression," Working Paper, Indiana University.

Lu, Y., and J. Y. Park (2014): "Estimation of Longrun Variance of Continuous Time Stochastic Process Using Discrete Sample," Working Paper, Indiana University.

Figure 4: Modified Tests for β in Models I-IV



Notes: Presented in the top two panels are the absolute values of the modified t-tests for $\beta = 1$ based on the nonparametric estimator $\hat{\Omega}$ of longrun variance Ω of the bivariate process $(\hat{u}_i, x_i\hat{u}_i)'$ from Models I and II. Similarly, presented in the bottom two panels are the modified t-tests based on the nonparametric estimator $\hat{\omega}^2$ of longrun variance ω^2 of the fitted error $(\hat{u}_i)$ from Models III and IV. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$ in yearly unit, and displayed as blue line across different levels of frequency parameter δ on the horizontal axis. For easy comparison, the conventional t-test is also presented in each graph as red line along with the black dotted line signifying the two-sided 5% standard normal critical value 1.96.

Newey, W. K., and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

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

Appendices

A Mathematical Proofs

Proof of Lemma 3.1 We may assume without loss of generality that Z is univariate by looking at each component separately. Note that

$$\frac{1}{T}\int_0^T Z_t dt = \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} Z_t dt$$
$$\frac{1}{n}\sum_{i=1}^n z_i = \frac{1}{T}\sum_{i=1}^n \delta Z_{(i-1)\delta} + \frac{\delta}{T}(Z_T - Z_0),$$

from which it follows that

$$\frac{1}{T}\int_0^T Z_t dt - \frac{1}{n}\sum_{i=1}^n z_i = \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} (Z_t - Z_{(i-1)\delta})dt + O_p\left(\frac{\delta}{T}\sup_{0\le t\le T}|Z_t|\right).$$

However, we have

$$|Z_t - Z_{(i-1)\delta}| \le |Z_t^c - Z_{(i-1)\delta}^c| + \sum_{(i-1)\delta < s \le t} \Delta Z_t$$

for all $i = 1, \ldots, n$ and t such that $(i-1)\delta < t \le i\delta$. Consequently, we have

$$\left|\frac{1}{n}\sum_{i=1}^n z_i - \frac{1}{T}\int_0^T Z_t dt\right| \le \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} |Z_t - Z_{(i-1)\delta}|dt + O_p\left(\frac{\delta}{T}\sup_{0\le t\le T}|Z_t|\right)$$

and

$$\begin{aligned}
\frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} |Z_t - Z_{(i-1)\delta}|dt &\le \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} \left(|Z_t^c - Z_{(i-1)\delta}^c| + \sum_{(i-1)\delta < s \le t} |\Delta Z_t|\right) dt \\
&\le \frac{1}{T}\sum_{i=1}^n \int_{(i-1)\delta}^{i\delta} \left(|Z_t^c - Z_{(i-1)\delta}^c| + \sum_{(i-1)\delta < s \le i\delta} |\Delta Z_t|\right) dt \\
&\le \left(\sup_{|t-s|\le\delta} |Z_t^c - Z_s^c|\right) + \frac{\delta}{T}\sum_{0<t\le T} |\Delta Z_t| \\
&= O_p\Big(\Delta_{\delta,T}(Z)\Big) + O_p(\delta),
\end{aligned}$$

and we may deduce the stated result immediately. □

Proof of Theorem 4.1 Under Assumptions A and B, we have

$$\frac{1}{n}\sum_{i=1}^{n} u_i^2 = \frac{1}{T}\int_0^T U_t^2 dt + o_p(1) \to_p \sigma^2$$

as $\delta \to 0$ and $T \to \infty$ with $\Delta_{\delta,T}(U) = o(1)$ as in Assumption D1 or D2.

For the proof of part (a), we write

$$\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i,$$

and note that, under Assumptions A and C1, we have

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' = \frac{1}{T}\int_0^T X_t X_t' dt + o_p(1) \to_p M > 0$$

$$\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i = \frac{1}{\sqrt{T}}\int_0^T X_t U_t dt + o_p(1) \to_d \mathbb{N}(0, \Pi)$$

as $\delta \to 0$ and $T \to \infty$ satisfying Assumption D1, and that

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \frac{1}{T}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} u_i x_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} u_i^2 + O_p(T^{-1}).\end{aligned}$$

Therefore, the stated results follow immediately.

The proof of part (b) is completely analogous. We write

$$\sqrt{T}\Lambda_T'(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} \Lambda_T^{-1} x_i x_i' \Lambda_T^{1\prime}\right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} \Lambda_T^{-1} x_i u_i,$$

and note that, under Assumptions A and C2, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} &= \frac{1}{T}\int_0^T \Lambda_T^{-1} X_t X_t' \Lambda^{-1\prime} dt + o_p(1)\\ &= \int_0^1 X_t^T X_t^{T\prime} dt + o_p(1) \to_d \int_0^1 X_t^\circ X_t^{\circ\prime} dt\end{aligned}$$

and

$$\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_T^{-1}x_iu_i = \frac{1}{\sqrt{T}}\int_0^T \Lambda_T^{-1}X_tU_tdt + o_p(1)$$
$$= \int_0^1 X_t^T dU_t^T + o_p(1) \to_d \int_0^1 X_t^\circ dU_t^\circ$$

as $\delta \to 0$ and $T \to \infty$ satisfying Assumption D2, and that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}u_i^2 - \frac{1}{T}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n}u_ix_i'\Lambda_T^{-1\prime}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda_T^{-1}x_ix_i'\Lambda_T^{1\prime}\right)^{-1}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_T^{-1}x_iu_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}u_i^2 + O_p(T^{-1}),$$

from which the stated results readily follow. □

Proof of Lemma 4.2 Write

$$\tilde{\rho} - 1 = \frac{\displaystyle\sum_{i=1}^{n}u_{i-1}(u_i - u_{i-1})}{\displaystyle\sum_{i=1}^{n}u_{i-1}^2}. \tag{11}$$

As shown earlier, we have $n^{-1}\sum_{i=1}^{n}u_{i-1}^2 \to_p \sigma^2$. Moreover, note that

$$u_{i-1} = \frac{1}{2}\left[(u_i + u_{i-1}) - (u_i - u_{i-1})\right],$$

and therefore, we may deduce that

$$\sum_{i=1}^{n}u_{i-1}(u_i - u_{i-1}) = \frac{1}{2}\left[\sum_{i=1}^{n}(u_i^2 - u_{i-1}^2) - \sum_{i=1}^{n}(u_i - u_{i-1})^2\right]$$
$$= \frac{1}{2}\left(u_n^2 - u_0^2\right) - \frac{1}{2}\sum_{i=1}^{n}(u_i - u_{i-1})^2,$$

which will be further analyzed subsequently.

We have

$$(U_{i\delta} - U_{(i-1)\delta})^2 = 2\int_{(i-1)\delta}^{i\delta}(U_{t-} - U_{(i-1)\delta})dU_t^c + \left([U^c]_{i\delta} - [U^c]_{(i-1)\delta}\right)$$
$$+ \sum_{(i-1)\delta < t \le i\delta}\Delta(U_t - U_{(i-1)\delta})^2, \tag{12}$$

and

$$\begin{aligned}\Delta(U_t - U_{(i-1)\delta})^2 &= (U_t - U_{(i-1)\delta})^2 - (U_{t-} - U_{(i-1)\delta})^2 \\ &= (U_t - U_{t-})(U_t + U_{t-} - 2U_{(i-1)\delta}) \\ &= (U_t - U_{t-})\Big[(U_t - U_{t-}) + 2(U_{t-} - U_{(i-1)\delta})\Big] \\ &= 2(U_{t-} - U_{(i-1)\delta})\Delta U_t + (\Delta U_t)^2 \end{aligned} \tag{13}$$

for $i = 1, \ldots, n$. Therefore, it follows from (12) and (13) that

$$\sum_{i=1}^{n} (U_{i\delta-} - U_{(i-1)\delta})^2 = [U]_T + 2Z_T, \tag{14}$$

where $Z = Z^c + Z^d$ with

$$\begin{aligned} Z_T^c &= \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_{t-} - U_{(i-1)\delta}) dU_t^c \\ Z_T^d &= \sum_{i=1}^{n} \sum_{(i-1)\delta < t \le i\delta} (U_{t-} - U_{(i-1)\delta}) \Delta U_t. \end{aligned}$$

Note that

$$[U]_T = \sum_{i=1}^{n} \Big([U^c]_{i\delta} - [U^c]_{(i-1)\delta}\Big) + \sum_{0<t\le T} (\Delta U_t)^2$$

for any n and δ such that $T = n\delta$. In what follows, we use

$$U_{t-} - U_{(i-1)\delta} = (U_t^c - U_{(i-1)\delta}^c) + \sum_{(i-1)\delta < s < t} \Delta U_s, \tag{15}$$

which holds for t, $(i-1)\delta < t \le i\delta$, and all $i = 1, \ldots, n$.

To consider Z^c, we write

$$Z^c = Z^a + Z^b, \tag{16}$$

where

$$\begin{aligned} Z_T^a &= \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_{t-} - U_{(i-1)\delta}) dA_t \\ &= \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t^c - U_{(i-1)\delta}^c) dA_t + \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(\sum_{(i-1)\delta < s < t} \Delta U_s \right) dA_t \end{aligned}$$

and

$$
\begin{aligned}
Z_T^b &= \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_{t-} - U_{(i-1)\delta}) dB_t \\
&= \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t^c - U_{(i-1)\delta}^c) dB_t + \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(\sum_{(i-1)\delta < s < t} \Delta U_s \right) dB_t,
\end{aligned}
$$

which we analyze subsequently. For Z_a, we have

$$
\begin{aligned}
\left| \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t^c - U_{(i-1)\delta}^c) dA_t \right| &\leq \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} |U_t^c - U_{(i-1)\delta}^c| |dA_t| \\
&\leq a_T \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} |U_t^c - U_{(i-1)\delta}^c| dt \\
&\leq a_T \Delta_{\delta,T}(U) \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} dt = O_p\Big(a_T T \Delta_{\delta,T}(U)\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
\left| \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(\sum_{(i-1)\delta < s < t} \Delta U_s \right) dA_t \right| &\leq \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(\sum_{(i-1)\delta < s < t} |\Delta U_s| \right) |dA_t| \\
&\leq a_T \sum_{i=1}^{n} \left(\sum_{(i-1)\delta < t \leq i\delta} |\Delta U_t| \right) \int_{(i-1)\delta}^{i\delta} dt \\
&= a_T \delta \sum_{0 < t \leq T} |\Delta U_t| = O_p(a_T T \delta),
\end{aligned}
$$

from which it follows that

$$
Z_T^a = O_p(a_T T \Delta_{\delta,T}(U)) + O_p(a_T T \delta) = O_p(a_T T \Delta_{\delta,T}(U)), \tag{17}
$$

since $\delta = O(\Delta_{\delta,T}(U))$.

For Z^b, it suffices to look at its quadratic Variation, since it can be embedded into a continuous martingale. However, we have

$$
\begin{aligned}
\sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t^c - U_{(i-1)\delta}^c)^2 d[B]_t &\leq b_T \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t^c - U_{(i-1)\delta}^c)^2 dt \\
&\leq b_T \Delta_{\delta,T}^2(U) \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} dt = O_p\left(b_T T \Delta_{\delta,T}^2(U)\right).
\end{aligned}
$$

Furthermore, it follows that

$$\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)^2 d[B]_t \le b_T\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)^2 dt,$$

and that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)^2 dt\right] &= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\mathbb{E}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)^2 dt\\
&= \sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left[\sum_{(i-1)\delta<s<t}\mathbb{E}(\Delta U_s)^2\right] dt\\
&\le \sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\mathbb{E}(\Delta U_t)^2\int_{(i-1)\delta}^{i\delta} dt\\
&= \delta\sum_{0<t\le T}\mathbb{E}(\Delta U_t)^2 = O(\delta T).
\end{aligned}$$

Therefore, we may deduce that

$$Z_T^b = O_p\left(\sqrt{b_T T}\Delta_{\delta,T}(U)\right) + O_p\left(\sqrt{b_T T\delta}\right) = O_p\left(\sqrt{b_T T}\Delta_{\delta,T}(U)\right), \tag{18}$$

since $\sqrt{\delta} = O(\Delta_{\delta,T}(U))$. The order of Z^c may now be easily obtained as

$$\begin{aligned}
Z_T^c &= O_p\left(a_T T\Delta_{\delta,T}(U)\right) + O_p\left(\sqrt{b_T T}\Delta_{\delta,T}(U)\right)\\
&= T\left[O_p\left(a_T\Delta_{\delta,T}(U)\right) + O_p\left(\sqrt{b_T/T}\Delta_{\delta,T}(U)\right)\right]
\end{aligned}\tag{19}$$

from (17) and (18).

To analyze Z^d, we let

$$\begin{aligned}
Z_T^d &= \sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}(U_{t-}-U_{(i-1)\delta})\Delta U_t\\
&= \sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}(U_t^c-U_{(i-1)\delta}^c)\Delta U_t + \sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)\Delta U_t
\end{aligned}$$

We have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}(U_t^c-U_{(i-1)\delta}^c)\Delta U_t\right]^2\\
&=\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\mathbb{E}(U_t^c-U_{(i-1)\delta}^c)^2\mathbb{E}(\Delta U_t)^2\\
&\le\left[\max_{1\le i\le n}\sup_{(i-1)\delta<t\le i\delta}\mathbb{E}(U_t^c-U_{(i-1)\delta}^c)^2\right]\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\mathbb{E}(\Delta U_t)^2\\
&=\left[\max_{1\le i\le n}\sup_{(i-1)\delta<t\le i\delta}\mathbb{E}(U_t^c-U_{(i-1)\delta}^c)^2\right]\sum_{0<t\le T}\mathbb{E}(\Delta U_t)^2=O\left(T\Delta_{\delta,T}^2(U)\right). \qquad (20)
\end{aligned}
$$

Moreover, we may easily deduce that

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)\Delta U_t\right]^2\\
&=\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\mathbb{E}\left(\sum_{(i-1)\delta<s<t}\Delta U_s\right)^2\mathbb{E}(\Delta U_t)^2\\
&=\sum_{i=1}^{n}\sum_{(i-1)\delta<t\le i\delta}\left[\sum_{(i-1)\delta<s<t}\mathbb{E}(\Delta U_s)^2\right]\mathbb{E}(\Delta U_t)^2\\
&\le\sum_{i=1}^{n}\sum_{(i-1)\delta<s,t\le i\delta}\mathbb{E}(\Delta U_s)^2\mathbb{E}(\Delta U_t)^2\\
&\le\sum_{i=1}^{n}\sum_{(i-1)\delta<s,t\le i\delta}\mathbb{E}(\Delta U_t)^4=\sum_{0<t\le T}\mathbb{E}(\Delta U_t)^4=O(T). \qquad (21)
\end{aligned}
$$

Therefore, it follows from (20) and (21) that

$$
Z_T^d=O_p\left(\sqrt{T}\Delta_{\delta,T}(U)\right)+O_p(\sqrt{T})=O_p(\sqrt{T})=o_p(T) \qquad (22)
$$

as $\delta\to 0$ and $T\to\infty$.

Now we have, due in particular to (19) and (22),

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} u_{i-1}(u_i - u_{i-1}) &= \frac{1}{2n}(u_n^2 - u_0^2) - \frac{1}{2n}\sum_{i=1}^{n}(u_i - u_{i-1})^2 \\
&= \delta\left[\frac{1}{2T}(U_T^2 - U_0^2) - \frac{1}{2T}\sum_{i=1}^{n}(U_{i\delta} - U_{(i-1)\delta})^2\right] \\
&= -\delta\left(\frac{1}{2T}[U]_T + o_p(1)\right) = -\frac{\delta\tau^2}{2} + o_p(\delta)
\end{aligned}$$

as $\delta \to 0$ and $T \to \infty$. Consequently, the stated result follows immediately from (11), and the proof is complete. $\square$

Proof of Theorem 5.1 Given Theorem 4.1, it is straightforward to show that

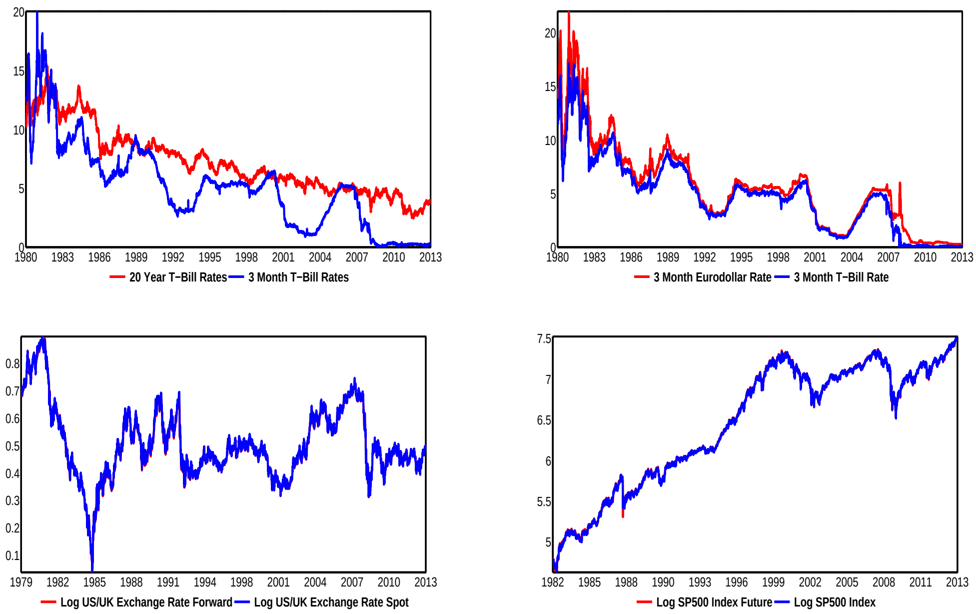
$$\delta\hat{\omega}^2 = \delta\tilde{\omega}^2 + o_p(1) \qquad \text{and} \qquad \delta\hat{\Omega} = \delta\tilde{\Omega} + o_p(1),$$

and therefore, we have $\delta\hat{\omega}^2 \to_p \pi^2$ and $\delta\hat{\Omega} \to_p \Pi$, due to Assumption F. Once this is established, the rest of the proof is entirely analogous to the proof of Theorem 4.1. Therefore, we do not provide the details to save the space. $\square$

B Additional Figures

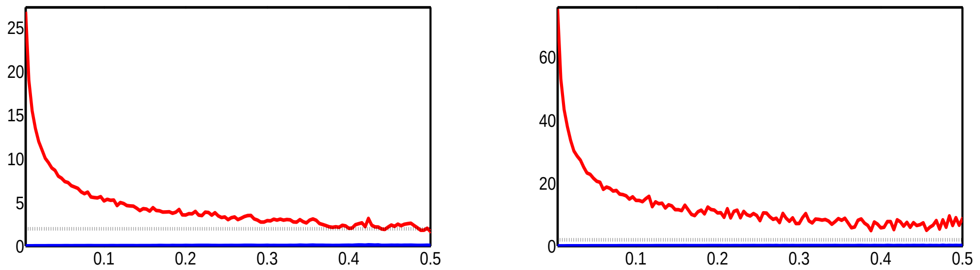
In this section, we present some additional figures. In Figure 5, we plot the regressands and regressors for each of our Models I-IV. On the other hand, Figure **??** presents the values of the G-statistics at different frequencies for Models I and II. As discussed, the underlying stochastic processes generating regressands and regressors for Models I and II appear to be at the boundary of the stationary and nonstationary regions.

Figure 5: Data Plots



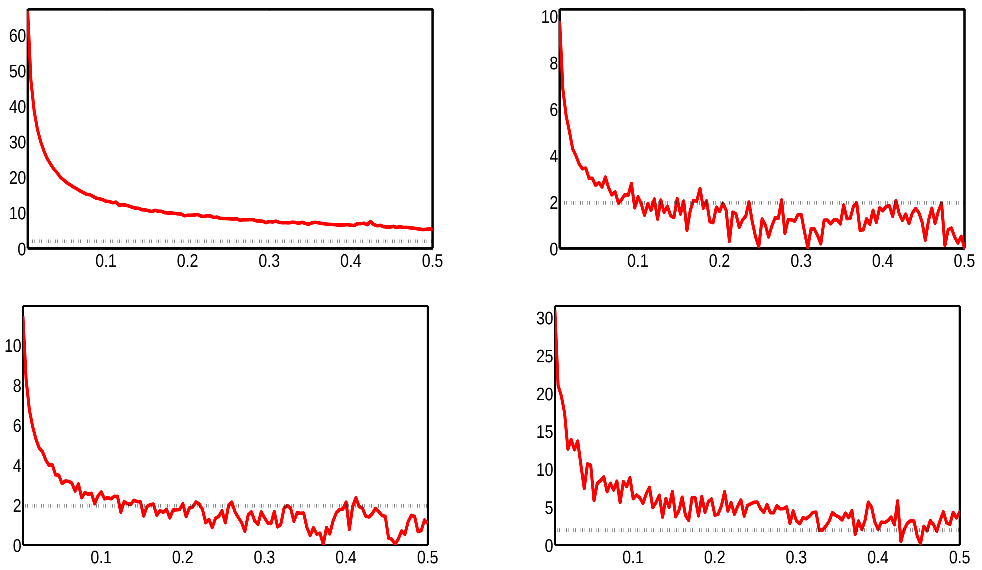
Notes: Each graph in Figure 5 presents the sample paths of the regressand y_i displayed in red line and the regressor x_i in blue line used for empirical illustrations of Models I-IV. Top left panel presents sample paths of 20-Year T-Bond rate and 3-Month T-Bill rate as y_i and x_i for Model I. Similarly, 3-Month Eurodollar rate and 3-Month T-Bill rate are presented in top right panel for Model II; log US/UK forward exchange rate and log US/UK spot exchange rate in bottom left panel for Model III, and log SP500 Index future and log SP500 Index in bottom right panel.

Figure 6: G-Tests for β in Models I and II



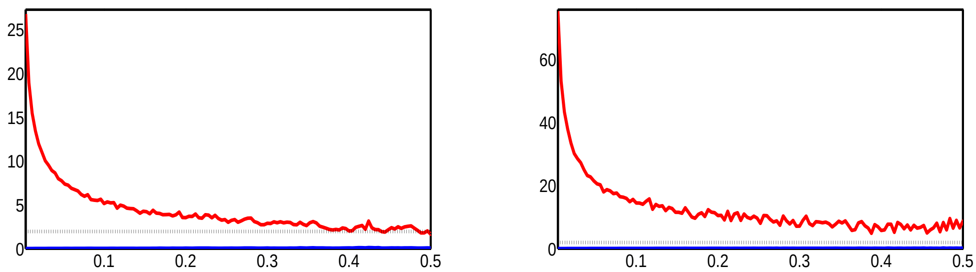
Notes: The modified t-tests for $\beta = 1$ based on the nonparametric estimator $\hat{\omega}^2$ of longrun variance ω^2 of the fitted error $(\hat{u}_i)$ from Models I and II are presented respectively in the left and right panels. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$ in yearly unit, and displayed as blue line across different levels of frequency parameter δ on the horizontal axis. For easy comparison, the conventional t-test is also presented in each graph as red line along with the black dotted line signifying the two-sided 5% standard normal critical value 1.96.

Figure 7: t-Tests for α in Models I-IV



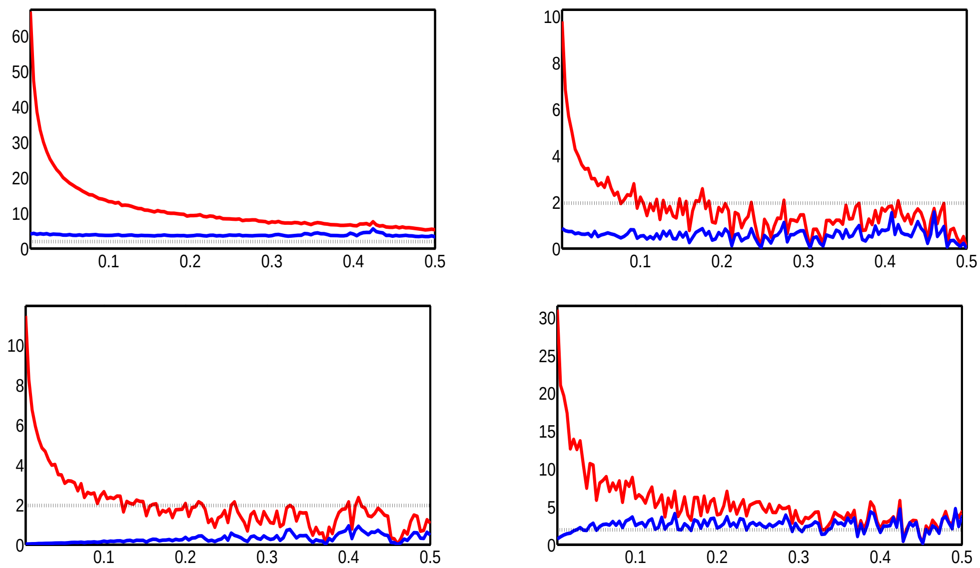
Notes: Presented in the top two panels are the absolute values of t-tests for $\alpha = 0$ in Models I and II, and in the bottom two panels those from Models III and IV. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$ in yearly unit. Each graph plots the absolute test values across different levels of frequency parameter δ on the horizontal axis. The black dotted horizontal line signifies the two-sided 5% standard normal critical value 1.96.

Figure 8: G-Tests for α in Models I and II



Notes: The modified t-tests for $\alpha = 0$ based on the nonparametric estimator $\hat{\omega}^2$ of longrun variance ω^2 of the fitted error $(\hat{u}_i)$ from Models I and II are presented respectively in the left and right panels. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$ in yearly unit, and displayed as blue line across different levels of frequency parameter δ on the horizontal axis. For easy comparison, the conventional t-test is also presented in each graph as red line along with the black dotted line signifying the two-sided 5% standard normal critical value 1.96.

Figure 9: Modified Tests for α in Models I-IV



Notes: Presented in the top two panels are the absolute values of the modified t-tests for $\alpha = 0$ based on the nonparametric estimator $\hat{\Omega}$ of longrun variance Ω of the bivariate process $(\hat{u}_i, x_i\hat{u}_i)'$ from Models I and II. Similarly, presented in the bottom two panels are the modified t-tests based on the nonparametric estimator $\hat{\omega}^2$ of longrun variance ω^2 of the fitted error $(\hat{u}_i)$ from Models III and IV. They are computed from the samples of varying frequency, from daily observations with the sampling interval $\delta = 1/250$ to semi-annual observations with $\delta = 1/2$ in yearly unit, and displayed as blue line across different levels of frequency parameter δ on the horizontal axis. For easy comparison, the conventional t-test is also presented in each graph as red line along with the black dotted line signifying the two-sided 5% standard normal critical value 1.96.