

THE UNIVERSITY OF  
**SYDNEY**

Economics Working Paper Series

2018 - 10

Understanding Regressions with Observations  
Collected at High Frequency over Long Span

Yoosoon Chang, Ye Lu and Joon Y. Park

July 2018

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

Yoosoon Chang

Department of Economics  
Indiana University

Ye Lu

School of Economics  
University of Sydney

Joon Y. Park

Department of Economics  
Indiana University  
and Sungkyunkwan University

July 16, 2018

## Abstract

In this paper, we analyze 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  $\delta$  shrink down to zero and the sample span  $T$  increase up to infinity. In this setup, we show that the standard Wald statistic diverges to infinity and the regression becomes spurious as long as  $\delta \rightarrow 0$  sufficiently fast relative to  $T \rightarrow \infty$ . Such a phenomenon is indeed what is frequently observed in practice for the type of regressions considered in the paper. In contrast, our asymptotic theory predicts that the spuriousness disappears if we use the robust version of the Wald test with an appropriate longrun variance estimate. This is supported, strongly and unambiguously, by our empirical illustration.

JEL Classification: C13, C22

Keywords and phrases: high frequency regression, spurious regression, continuous time model, asymptotics, longrun variance estimation

---

\*Earlier versions of this paper have been circulated since 2014. We would like to thank Donald Andrews, Bandi Federico and Jihyun Kim 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 seem to 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  $\delta$  over sample span  $T$ , and we let  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  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 dependence. The same problem arises in the regressions we consider. The regression error from a con-

tinuous time process becomes strongly dependent as the sampling interval decreases, which yields the same type of spuriousness in the conventional spurious regression.

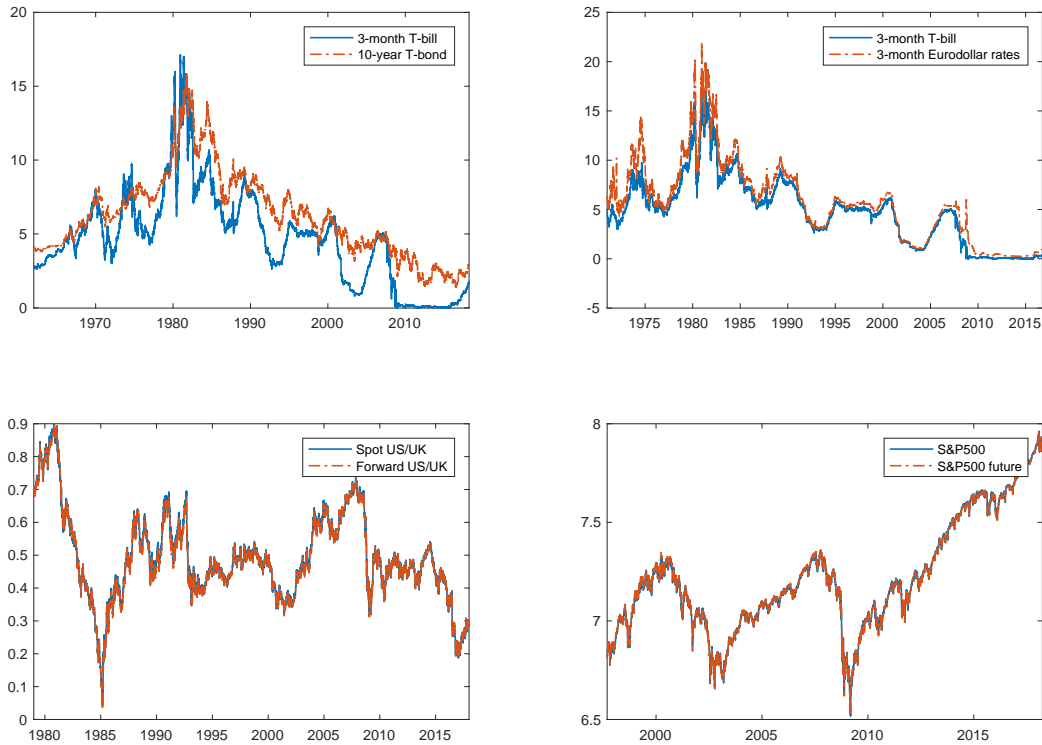
In time series regressions, we often use the robust version of standard tests with a longrun variance estimator in place of the usual variance estimator, to allow for the presence of serial dependence. Our asymptotic theory shows that the robust test may or may not diverge to infinity under the correct null hypothesis, depending upon how we choose the bandwidth parameter in our longrun variance estimator used in the test. If it is chosen appropriately, the test becomes valid at high frequency and the spuriousness at high frequency disappears. However, with a conventional choice of bandwidth made in the discrete time setup, it diverges as the sampling interval decreases like in the case of the standard tests, resulting in the spuriousness at high frequency. The use of data-dependent bandwidth choice also generally has a critical effect on the validity of robust test. For instance, the test becomes valid if the bandwidth selection is made using the procedure by Andrews (1991), while the spuriousness at high frequency arises if the method proposed by Newey and West (1994) is employed. In all of actual regressions we consider in the paper, the tests behave exactly as predicted by our asymptotic theory.

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 setup 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 robust versions of the Wald test statistic defined with longrun variance 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  $y_i = \beta_0 + \beta_1 x_i + u_i$ , where  $\beta_0$  and  $\beta_1$  are respectively the intercept and slope parameters and  $(u_i)$  are the regression errors. For  $(y_i)$  and  $(x_i)$ , we consider the following

Figure 1: Data Plots for Models I–IV



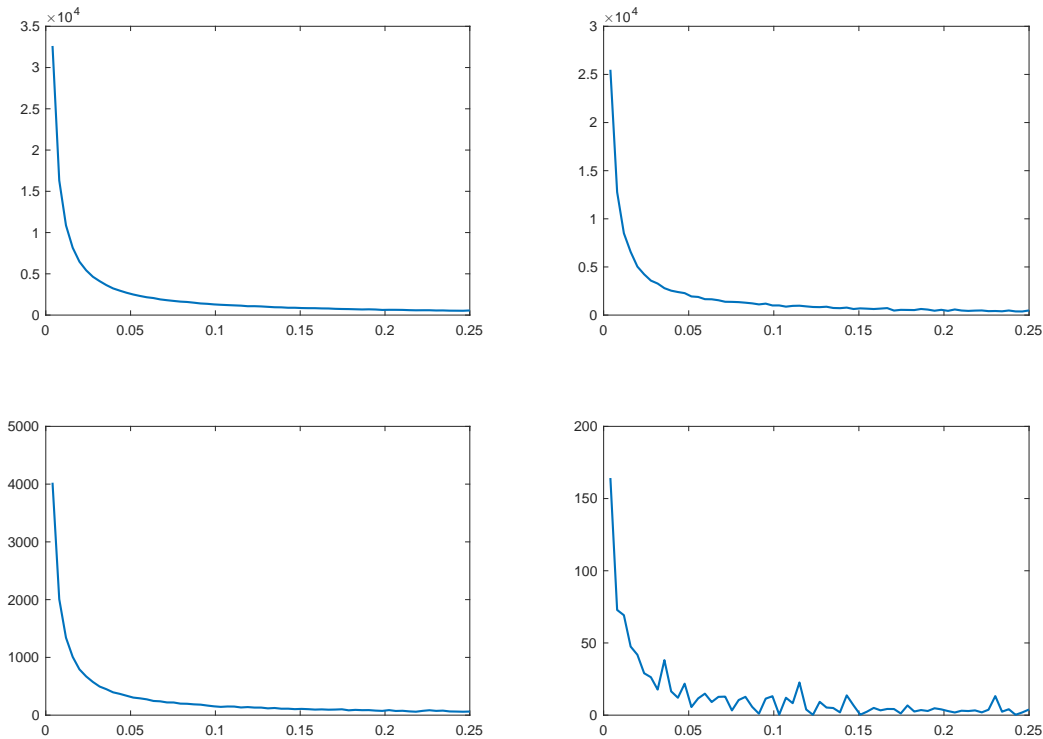
Notes: Presented are daily sample paths of the regressand  $y$  and the regressor  $x$  used for the empirical illustrations of Models I–IV. Top left panel presents the 10-year T-bond rate and 3-month T-bill rate from January 3, 1962 to March 29, 2018. Top right panel presents the 3-month Eurodollar rate and 3-month T-bill rate from January 1, 1971 to October 7, 2016. Bottom left panel presents the log US/UK 3-month forward exchange rate and log US/UK spot exchange rate from January 2, 1979 to December 29, 2017. Bottom right panel presents the log S&P 500 Index future and log S&P 500 Index from September 10, 1997 to March 29, 2018.

four pairs:

Model	$(y_i)$	$(x_i)$
I	10-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 S&P 500 index futures	log S&P 500 index

The plots for  $(y_i)$  and  $(x_i)$  in Models I–IV are given in Figure 1. In all models, possibly except for Model I, two series  $(y_i)$  and  $(x_i)$  move very closely with each other. Therefore, the

Figure 2: Wald Tests in Models I-IV



Notes: Presented are the values of the Wald statistics for the null hypothesis  $H_0 : \beta_0 = 0, \beta_1 = 1$  in Models I-IV plotted against the varying sampling intervals  $\delta$  ranging from  $1/252$  (daily frequency) to  $1/4$  (quarterly frequency).

most natural hypothesis to be tested appears to be  $H_0 : \beta_0 = 0$  and  $\beta_1 = 1$ . The hypothesis may or may not hold. In particular, the null hypothesis  $\beta_0 = 0$  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, in the paper, we do not intend to provide any answers to whether or not the hypothesis should hold in any of the models we consider here. Instead, we simply analyze the dependence of test results on the sampling frequency.

In Figure 2, we present the values of the Wald test in Models I-IV, against various sampling intervals from one quarter with  $\delta = 1/4$  to one day with  $\delta = 1/252$  in yearly unit. As discussed, the values of the Wald test change extensively as the sampling interval  $\delta$  varies. In particular, they tend to increase very rapidly as  $\delta$  gets smaller and becomes near zero. The rate of increase in the values of the Wald test as  $\delta$  approaches zero varies across different models. However, it is common to all models that the values of the Wald

test start to increase sharply as the sampling interval becomes approximately one month or shorter, and eventually at daily frequency, the Wald tests in all models unambiguously and unanimously reject the null hypothesis. In contrast, it appears that the Wald tests yield some mixed results for the same hypothesis as the sampling interval moves away from a neighborhood of zero and further increases.<sup>1</sup>

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 Wald tests is not an anomaly. From our asymptotic analysis relying on  $\delta \rightarrow 0$  as well as  $T \rightarrow \infty$ , it actually becomes clear that the test is expected to diverge up to infinity with probability one as  $\delta$  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, Setup and Preliminaries

Consider the standard regression model

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

for  $i = 1, \dots, n$ , where  $(y_i)$  and  $(x_i)$  are respectively the regressand and regressor,  $\beta$  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  $\beta$ . The general linear hypothesis on  $\beta$ , formulated typically as  $R\beta = r$  with known matrix  $R$

---

<sup>1</sup>Typically, the chi-square critical values are used for the Wald tests we consider here. We expect that the Wald tests relying on chi-square critical values are subject to nonnegligible size distortions, due mainly to the presence of persistency in the regressors and regression errors of our models. Nevertheless, we will not pay attention to this and other related problems of the Wald tests because, as discussed, the focus of this paper is to analyze their frequency dependence, not their performance at any particular frequency.

and vector  $r$  of conformable dimensions, is often tested using the Wald statistic defined by

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

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

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

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

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

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.<sup>2</sup>

In the paper, we analyze regression (1), when  $(y_i)$  and  $(x_i)$  are high frequency observations.<sup>3</sup> 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.,

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

for  $i = 1, \dots, n$  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,

<sup>2</sup>Note that the longrun variance of  $(x_i u_i)$  does not exist if  $(x_i)$  is nonstationary.

<sup>3</sup>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 observations.



it is clear that we may define the continuous time regression

$$Y_t = X_t' \beta + U_t \quad (5)$$

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

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

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

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

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

for some  $\epsilon \geq 0$  and a nonrandom sequence  $(\lambda_T)$  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  $T \rightarrow \infty$  in our setup 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  $i = 1, \dots, n$ , we have*

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

for all small  $\delta$  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.**  $T^{-1} \int_0^T U_t^2 dt \rightarrow_p \sigma^2$  for some  $\sigma^2 > 0$  as  $T \rightarrow \infty$ .

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)  $T^{-1} \int_0^T X_t X_t' dt \rightarrow_p M$  as  $T \rightarrow \infty$  for some nonrandom matrix  $M > 0$ , and
- (b) we have

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

as  $T \rightarrow \infty$ , where  $\Pi = \lim_{T \rightarrow \infty} T^{-1} \mathbb{E} \left( \int_0^T X_t U_t dt \right) \left( \int_0^T X_t U_t dt \right)' > 0$ , 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  $(\Lambda_T)$  of matrices as

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

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

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

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

then  $U^T \rightarrow_d U^\circ$  in  $D[0, 1]$  as  $T \rightarrow \infty$ , where  $U^\circ$  is Brownian motion with variance  $\pi^2 = \lim_{T \rightarrow \infty} T^{-1} \mathbb{E} \left( \int_0^T U_t dt \right)^2 > 0$ , 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 (2017). 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.**  $\Delta_{\delta,T}(U), \Delta_{\delta,T}(\|XX'\|) \rightarrow 0$  and  $\sqrt{T}\Delta_{\delta,T}(\|XU\|) \rightarrow 0$  as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ .

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

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

## 4 Spuriousness of Regression at High Frequency

In this section, we establish the asymptotics of the OLS estimator  $\hat{\beta}$  and analyze the asymptotic behaviors of the standard Wald test based on the test 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*

$$\begin{aligned}\sqrt{T}(\hat{\beta} - \beta) &\rightarrow_d N \\ \delta F(\hat{\beta}) &\rightarrow_d N'R'(RM^{-1}R')^{-1}RN/\sigma^2,\end{aligned}$$

where  $N =_d \mathbb{N}(0, M^{-1}\Pi M^{-1})$ , as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  satisfying Assumption D1.

(b) *Under Assumptions C2, we have*

$$\begin{aligned}\sqrt{T}\Lambda'_T(\hat{\beta} - \beta) &\rightarrow_d P \\ \delta F(\hat{\beta}) &\rightarrow_d P'R'(RQ^{-1}R')^{-1}RP/\sigma^2\end{aligned}$$

where  $P = \left(\int_0^1 X_t^\circ X_t^{\circ'} dt\right)^{-1} \int_0^1 X_t^\circ dU_t^\circ$  and  $Q = \int_0^1 X_t^\circ X_t^{\circ'} dt$ , as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  satisfying Assumption D2.

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

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

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ . 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  $\Delta_{\delta,T}(U) \rightarrow_p 0$  or more strongly  $\sqrt{T}\Delta_{\delta,T}(U) \rightarrow_p 0$  if needed as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ . We may easily deduce that

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

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

However, we have

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

as  $n \rightarrow \infty$  (with  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ ), and consequently, the central limit theory fails to hold for  $(u_i)$ . In fact, in our setup,  $(u_i)$  becomes strongly dependent as  $\delta \rightarrow 0$ , since the correlation between  $u_i$  and  $u_{i-j}$  for any  $i$  and  $j$  becomes unity as  $\delta \rightarrow 0$ . 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 in

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

$$\sup_{0 \leq s, t \leq T} \frac{|A_t - A_s|}{|t - s|} = O_p(p_T) \quad \text{and} \quad \sup_{0 \leq s, t \leq T} \frac{|[M]_t - [M]_s|}{|t - s|} = O_p(q_T),$$

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

The conditions introduced in Assumption E are mild and expected to hold for a wide class of asymptotically stationary error processes. In Part (a), we require that both the bounded variation component and the quadratic variation of the martingale component of the continuous part of the error process  $U$  be Lipschitz continuous and  $\delta$  be small enough to allow their Lipschitz constants to increase with  $T$ . On the other hand, Part (b) holds, for instance, if the number of jump increases at  $T$ -rate and jump size has finite fourth moment, and if the instantaneous variance of  $U$  is asymptotically stationary. Note that  $[U]_T = [U^c]_T + \sum_{0 \leq t \leq T} (\Delta U_t)^2$  and  $[U^c]_T = [M]_T$ .

The asymptotics for the estimated AR coefficient  $\tilde{\rho}$  of  $\rho$  in (7) are given as

**Lemma 4.2.** *Under Assumption E, we have*

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

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

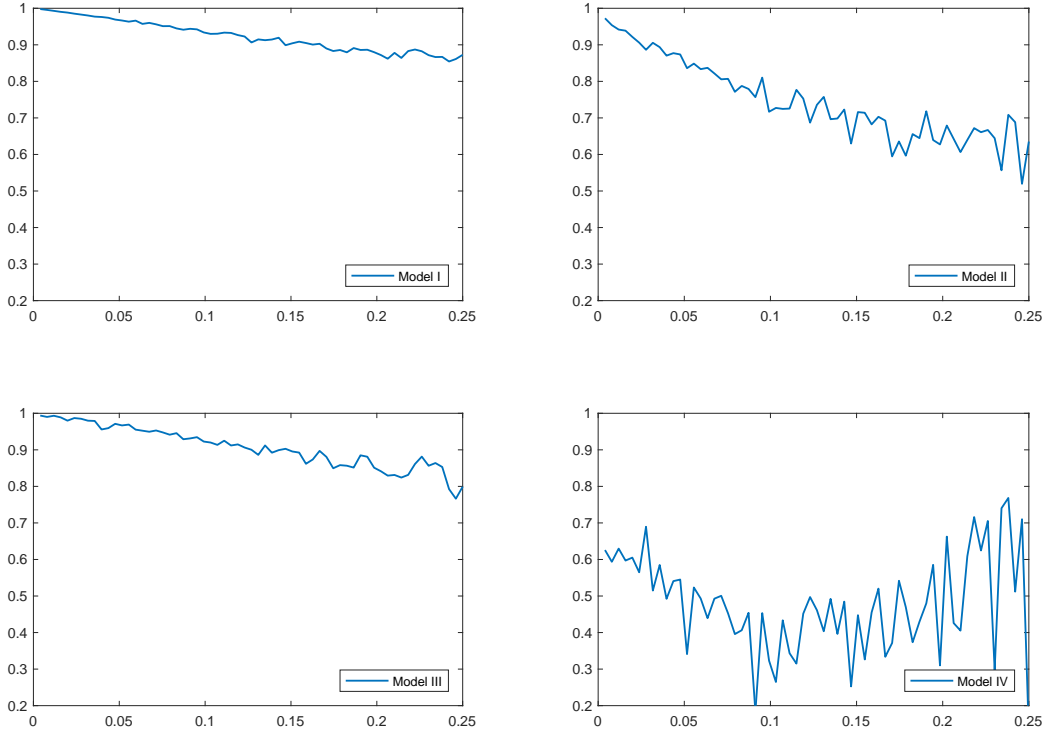
It follows immediately from Lemma 4.2 that

$$\tilde{\rho} \rightarrow_p 1$$

as  $\delta \rightarrow 0$  in particular. Therefore,  $(u_i)$  becomes strongly dependent, and the regression becomes spurious as the sampling interval  $\delta$  approaches to 0. Our regression is completely analogous to the conventional spurious regression, except that we let  $\delta \rightarrow 0$  in contrast with the conventional spurious regression requiring  $n \rightarrow \infty$ . Therefore, the results in Theorem 4.1 may well be expected. Though we let  $T \rightarrow \infty$ , as well as  $\delta \rightarrow 0$ , to get more explicit limit of  $\tilde{\rho}$  as in Lemma 4.2, the condition  $T \rightarrow \infty$  is not essential for the spuriousness in regression (1). This is clear from our proof of Lemma 4.2. If  $T$  is assumed to fixed and set to be  $T = 1$  without loss of generality, we have  $\delta = 1/n$  and  $(u_i)$  asymptotically reduces to a near unit root process with AR coefficient  $1 - c/n$  with  $c = \tau^2/2\sigma^2$ .

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  $T^{-1} \int_0^T U_t^2 dt \rightarrow_p \infty$  as  $T \rightarrow \infty$ . Therefore, we expect that our regression becomes spurious as long as  $T$  is large even if  $\delta$  is not small. In the paper, however, we assume  $T^{-1} \int_0^T U_t^2 dt \rightarrow_p \sigma^2$  as  $T \rightarrow \infty$ , and let  $\delta \rightarrow 0$  to analyze the

Figure 3: Estimated Residual AR Coefficients for Models I-IV



Notes: Presented are estimated autoregressive coefficients in the first order autoregression of the residuals from regressions of Models I-IV, plotted against various sampling intervals  $\delta$  ranging from  $1/252$  (daily frequency) to  $1/4$  (quarterly frequency).

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  $\delta$  increases depends on the ratio  $\tau^2/\sigma^2$ . Roughly,  $\tau^2$  measures the mean local variation, while  $\sigma^2$  represents the mean global variation of the error process  $U$ . Therefore, we may refer to it as the *local-to-global variation ratio*. The larger the value of the ratio is,  $\tilde{\rho}$  converges more slowly to the unity as  $\delta \rightarrow 0$ . Note that the ratio becomes large if, in particular, either the existence of excessive volatility makes local variation large, or the presence of strong mean reversion makes global variation small. If  $U$  is Ornstein-Uhlenbeck process given by  $dU_t = -\kappa U_t dt + v dW_t$ , we have  $\tau^2 = v^2$  and  $\sigma^2 = v^2/2\kappa$ . Therefore, the ratio becomes  $\tau^2/\sigma^2 = 2\kappa$ , which becomes larger if we have stronger mean reversion.

The actual estimates of the autoregressive coefficients for the fitted residuals from Models I-IV are plotted in Figure 3 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. This is perhaps due to more irregular and frequent jump activities existing in stock prices than are allowed in our paper. Of course, not all types of jumps are permitted in our paper, though our assumptions on jumps are fairly general and weak. In particular, we assume that the jump intensity is proportional to time span, following most existing literature, which excludes the possibility of having jumps with intensity varying with sampling frequency.

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  $c' X U$  satisfies the same conditions as those we impose on  $U$  above in Assumption E.

## 5 Asymptotics of Robust Wald Tests

In this section, we develop the asymptotics of the robust Wald tests based on the test statistics  $G(\hat{\beta})$  and  $H(\hat{\beta})$  in (3) and (4), which will be referred to the  $G$  and  $H$ -tests for short. Their test statistics involve the longrun variance estimators of  $(u_i)$  and  $(x_i u_i)$ , instead of their usual variance estimators used for the Wald tests, to take into account the presence of serial correlation in  $(u_i)$  and  $(x_i u_i)$ . Our asymptotics in this section show that, in general, the spuriousness in regression at high frequency we observe and analyze in previous sections is expected to appear also in the  $G$  and  $H$ -tests, as well as in the Wald test. Therefore, it is not exclusively due to neglecting the presence of serial correlation in  $(u_i)$  and  $(x_i u_i)$ , which necessarily emerges in high frequency regression. The spuriousness disappears only when we estimate the longrun variances of  $(u_i)$  and  $(x_i u_i)$  adequately using bandwidths chosen suitably.

Under the null hypothesis, we may expect that

$$\begin{aligned} G(\hat{\beta}) &\approx P_T' R' [R Q_T^{-1} R']^{-1} R P_T / (\delta \hat{\omega}^2) \\ H(\hat{\beta}) &\approx P_T' R' [R Q_T^{-1} (\delta \hat{\Omega}) Q_T^{-1} R']^{-1} R P_T \end{aligned} \quad (8)$$

with

$$P_T = \left( \int_0^T X_t X_t' dt \right)^{-1} \int_0^T X_t U_t dt, \quad Q_T = \int_0^T X_t X_t' dt,$$



if  $\delta$  is sufficiently small relative to  $T$ . Therefore, it is clearly seen that the longrun variance estimators,  $\hat{\omega}^2$  and  $\hat{\Omega}$  respectively of  $(u_i)$  and  $(x_i u_i)$ , play very important roles in the asymptotics of  $G$  and  $H$ -tests.

To analyze the longrun variance estimators of  $(u_i)$  and  $(x_i u_i)$ , we set  $v_i = u_i$  or  $x_i u_i$  and  $v_i = V_{i\delta}$ ,  $i = 1, \dots, n$ , with  $V = U$  or  $XU$ , and assume for expositional simplicity that  $V$  is one-dimensional. The commonly used longrun variance estimator  $\omega_n^2$  of  $(v_i)$  is given by

$$\omega_n^2 = \sum_{|j| \leq n} K\left(\frac{j}{b_n}\right) \gamma_n(j), \quad (9)$$

where  $K$  is the kernel function,  $\gamma_n(j) = n^{-1} \sum_i v_i v_{i-j}$  is the sample autocovariance function of  $(v_i)$  and  $b_n$  is the bandwidth parameter. If  $\delta$  is fixed and  $(v_i)$  has a well defined longrun variance  $\omega^2$ , then we would expect  $\omega_n^2 \rightarrow_p \omega^2$  as long as  $b_n \rightarrow \infty$  and  $b_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . If we set  $\delta \rightarrow 0$  as in our setup, however,  $\omega_n^2$  behaves much differently in asymptotics.

The type of longrun variance estimator in (9) defined with high frequency observations is analyzed in Lu and Park (2018). Following them, we relate it to the longrun variance  $\varpi^2$  of the underlying continuous time process  $V$ , which is given by

$$\varpi_T^2 = \int_{|s| \leq T} K\left(\frac{s}{B_T}\right) \Gamma_T(s) ds, \quad (10)$$

where  $K$  is the kernel function,  $\Gamma_T(s) = T^{-1} \int_0^T V_t V_{t-s} dt$  is the sample autocovariance function, and  $B_T$  is the bandwidth parameter. Under very general regularity conditions, they show that

$$\varpi_T^2 \rightarrow_p \varpi^2 \quad (11)$$

as  $T \rightarrow \infty$ , if and only if  $B_T \rightarrow \infty$  and  $B_T/T \rightarrow 0$  as  $T \rightarrow \infty$ .

It is easy to see that

$$\gamma_n(j) = \frac{1}{n} \sum_{i=1}^n v_i v_{i-j} = \frac{1}{T} \sum_{i=1}^n \delta V_{i\delta} V_{(i-j)\delta} \approx \frac{1}{T} \int_0^T V_t V_{t-j\delta} dt = \Gamma_T(j\delta),$$

and therefore, we have

$$\delta \omega_n^2 \approx \sum_{|j| \leq n} \delta K\left(\frac{j\delta}{b_n\delta}\right) \Gamma_T(j\delta) \approx \int_{|s| \leq T} K\left(\frac{s}{B_{n,\delta}}\right) \Gamma_T(s) ds, \quad (12)$$

where we let  $B_{n,\delta} = b_n\delta$  be the continuous time bandwidth corresponding to the discrete time bandwidth  $b_n$ . Consequently, we may readily deduce from (10), (11) and (12) that,

under suitable regularity conditions,

$$\delta\omega_n^2 \rightarrow_p \varpi^2, \quad (13)$$

if and only if

$$B_{n,\delta} \rightarrow \infty \quad \text{and} \quad B_{n,\delta}/T \rightarrow 0 \quad (14)$$

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

If a discrete time bandwidth  $b_n$  yields the corresponding continuous time bandwidth  $B_{n,\delta} = b_n\delta$  satisfying conditions in (14), we say that it is *high-frequency compatible*. Of the two high-frequency compatibility conditions in (14), only the first one  $B_{n,\delta} = b_n\delta \rightarrow \infty$  matters, since the second one  $B_{n,\delta}/T = b_n \rightarrow \infty$  is required for any discrete time bandwidth  $b_n$  as well. For the high-frequency compatibility of a discrete time bandwidth  $b_n$ , we should have  $b_n \rightarrow \infty$  faster than  $\delta \rightarrow 0$ . It is easy and straightforward to find a high-frequency compatible bandwidth  $b_n$ . For instance, if we set  $b_n = cn^a/\delta^{1-a}$  with some  $c > 0$  and  $0 < a < 1$ , then  $B_{n,\delta} = b_n\delta = cT^a \rightarrow \infty$  as long as  $T \rightarrow \infty$ , and therefore,  $b_n$  becomes high-frequency compatible. This bandwidth choice will be referred to as the rule of thumb in continuous time (CRT).

However, the usual discrete time bandwidth given by  $b_n = cn^a$  with some  $c > 0$  and  $0 < a < 1$ , the scheme that we will simply refer to as the rule of thumb (RT), is *not* high-frequency compatible, since  $B_{n,\delta} = b_n\delta = c\delta^{1-a}T^a \rightarrow 0$  if  $\delta \rightarrow 0$  fast enough relative to  $T \rightarrow \infty$ . In this case, we have

$$\frac{\delta\omega_n^2}{B_{n,\delta}} \approx \frac{1}{B_{n,\delta}} \int_{|s| \leq T} K\left(\frac{s}{B_{n,\delta}}\right) \Gamma_T(s) ds \rightarrow_p \Gamma(0) \int K(x) dx,$$

where  $\Gamma(s)$  is the autocovariance function of  $V$ , which implies, in particular, that

$$\delta\omega_n^2 \rightarrow_p 0 \quad (15)$$

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

Now it is clear that the asymptotics of the  $G$  and  $H$ -tests rely critically on whether or not the bandwidths used in their longrun variance estimators are high frequency compatible. We expect them to have well defined limit distributions if they are based on the longrun variances estimated with high frequency compatible bandwidths, for which (13) holds. However, if the longrun variances estimators relying on the bandwidths that are

high frequency incompatible and (15), then we have

$$G(\hat{\beta}) \rightarrow_p \infty \quad \text{and} \quad H(\hat{\beta}) \rightarrow_p \infty$$

even under the null hypothesis, as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ , exactly as in the case of the standard Wald test.

In discrete time framework, the bandwidth parameter  $b_n$  is often set to be data-dependent following Andrews (1991) and Newey and West (1994), which we simply refer to as the Andrews bandwidth and Newey-West bandwidth respectively. Specifically, both Andrews and Newey-West bandwidths are based on the infeasible optimal bandwidth given by

$$b_n^* = \left( \frac{r\pi_r^2\theta_r^2}{\int K(x)^2 dx} n \right)^{1/(2r+1)}, \quad (16)$$

where  $r$  is the characteristic exponent of kernel function  $K$ ,  $\pi_r = \lim_{x \rightarrow 0} (1 - K(x))/|x|^r$  and  $\theta_r = \sum_j |j|^r \gamma(j) / \sum_j \gamma(j)$  with  $\gamma(j)$  denoting the autocovariance function of  $(v_i)$ . Note that  $\theta_r$  is unknown and should be estimated. Andrews (1991) assumes that  $(v_i)$  is AR(1), in which case we have

$$\theta_1^2 = \frac{4\rho^2}{(1-\rho)^2(1+\rho)^2}, \quad \theta_2^2 = \frac{4\rho^2}{(1-\rho)^4} \quad (17)$$

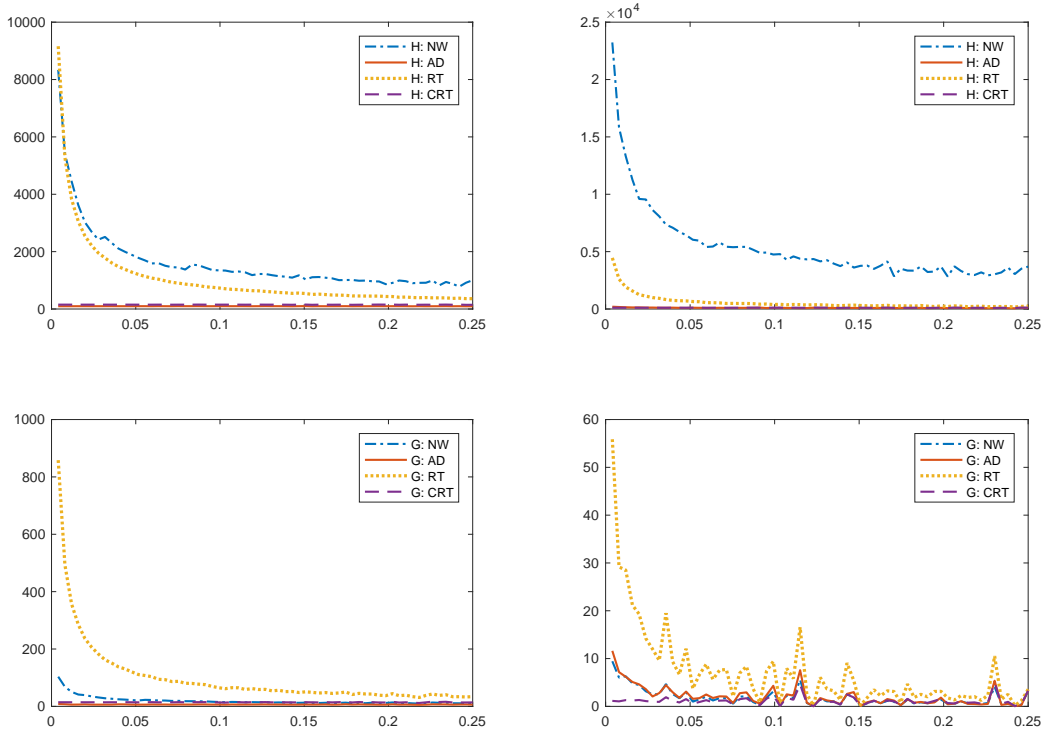
for  $r = 1, 2$ . He suggests to run an AR(1) regression for  $(v_i)$  and use the OLS estimate  $\tilde{\rho}$  of the autoregressive coefficient  $\rho$  to obtain an estimate  $\tilde{\theta}_r$  of  $\theta_r$  in (17) for  $r = 1, 2$ . On the other hand, Newey and West (1994) propose to nonparametrically estimate  $\theta_r$  by  $\hat{\theta}_r = \sum_{|j| \leq a_n} |j|^r \gamma_n(j) / \sum_{|j| \leq a_n} \gamma_n(j)$  with  $a_n = cn^p$  for some constants  $c$  and  $0 < p < 1$ .

**Lemma 5.1.** *Let  $V$  satisfy Assumption E, and assume  $\sup_{0 \leq t \leq \infty} \mathbb{E}|V|^2 < \infty$ . The Andrews bandwidth is high-frequency compatible, while the Newey-West bandwidth is not. For the latter, we have  $B_{n,\delta} \sim_p \delta^{2r(1-p)/(2r+1)} T^{(2pr+1)/(2r+1)}$  as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ .*

Obviously, the Newey-West bandwidth is not high-frequency compatible because  $B_{n,\delta} \rightarrow 0$  if  $\delta = o(T^{-(2pr+1)/(2r(1-p))})$ , which holds whenever  $\delta \rightarrow 0$  fast enough relative to  $T \rightarrow \infty$ . The  $G$  and  $H$ -tests with the Newey-West bandwidth becomes invalid at high frequency. They are expected to be generally valid if the Andrews bandwidth is used.

The actual values of the  $G$  and  $H$ -tests in Models I-IV introduced in Section 2 are presented in Figure 4 for various sampling intervals  $\delta$  ranging from 1/252 to 1/4, which correspond to the daily and quarterly frequencies. We use the  $H$ -test for Models I/II, and the  $G$ -test for Models III/IV. Of course, this implies that we interpret Models I/II as the stationary regression, and Models III/IV as the cointegrating regression. We consider four

Figure 4: Robust Wald Tests in Models I-IV



Notes: Presented are the values of the  $H$  and  $G$ -test statistics for the null hypothesis  $H_0 : \beta_0 = 0, \beta_1 = 1$  in Models I/II and III/IV, respectively, plotted against the varying sampling intervals  $\delta$  ranging from  $1/252$  (daily frequency) to  $1/4$  (quarterly frequency).

schemes for the bandwidth choice: the rule of thumb (RT), the rule of thumb in continuous time (CRT), Andrews (AD) and Newey-West (NW). As shown, CRT and AD are high frequency compatible, whereas RT and NW are not. Our asymptotics show that the  $G$  and  $H$ -tests are valid only when high frequency compatible bandwidths are used. If high frequency incompatible bandwidths are used, the  $G$  and  $H$ -tests are expected to diverge even if the null hypothesis is true.

When we use CRT and AD, which are high frequency compatible, the  $G$  and  $H$ -tests yield values that are not sensitive to the sampling frequency. They remain stable more or less across all ranges of the sampling interval we consider. In sharp contrast, the values of  $G$  and  $H$ -tests become heavily dependent upon the sampling frequency. In this case, the overall feature of frequency dependence of the  $G$  and  $H$ -tests is exactly identical to that of the standard Wald test we explore in Section 2. The test values change extensively as the sampling frequency varies. In particular, they increase rapidly as the sampling interval

becomes smaller than one month, and explode as the sampling interval approaches one day. This shows in particular that the spuriousness we observe for the standard Wald test cannot be simply dealt with the use of its robust version allowing for the presence of serial correlation.

In the rest of the section, we introduce a set of regularity conditions and formally establish the asymptotics for the  $G$  and  $H$ -tests. To ensure that the  $G$  and  $H$ -tests have well defined limit null distributions, we need to assume

**Assumption F.** Let  $V$  satisfy Assumptions 2.2(b), 2.3 and 4.1, and  $K$  satisfy Assumptions 2.1, 2.2(a) and 4.2, in Lu and Park (2018). Moreover, we assume that  $b_n$  used in  $\hat{\omega}^2$  and  $\hat{\Omega}$  is high frequency compatible, and  $b_n \delta \Delta_{\delta, T}(V) \rightarrow 0$  as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ .

Assumption F specifies the regularity conditions on the process  $V$  and the kernel function  $K$  used in the long-run variance estimation. See Lu and Park (2018) for more discussions on these conditions.

The limit null distributions of the  $G$  and  $H$ -tests are presented below. We let  $q$  be the number of restrictions, and  $\chi_q^2$  denote the chi-square distribution with  $q$  degrees of freedom.

**Theorem 5.2.** *Assume  $R\beta = r$  and let Assumptions A and F hold.*

(a) *Under Assumptions C1 and D1, we have*

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

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

(b) *Under Assumptions C2 and D2, we have*

$$G(\hat{\beta}) \rightarrow_d P^{*'} R' (RQ^{-1}R')^{-1} RP^*$$

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ , where  $P^* = \left( \int_0^1 X_t^\circ X_t^{\circ'} dt \right)^{-1} \int_0^1 X_t^\circ dU_t^*$  with standard Brownian motion  $U^*$  defined as  $U^* = U^\circ / \pi$  and  $Q = \int_0^1 X_t^\circ X_t^{\circ'} dt$  using the notations in Theorem 4.1.

Both of the  $G$  and  $H$ -tests have well defined limit null distributions respectively for general stationary and nonstationary regressions, if in particular we use high-frequency compatible bandwidths for their longrun variance estimates. The  $H$ -test has the standard chi-square limit null distribution for stationary regressions. On the other hand, the limit null distribution of the  $G$ -test is generally nonnormal and nonstandard. If, however, the limit processes  $X^\circ$  and  $U^\circ$  are independent, then its limit null distribution reduces to chi-square distribution.

## 6 Simulation

Our simulation shows that the asymptotics developed in the paper are relevant and useful to analyze regressions using high-frequency observations. In our simulation, we consider two types of regressions, the stationary and cointegration regressions, based on the continuous time regression  $Y_t = \beta_0 + \beta_1 X_t + U_t$  for  $0 \leq t \leq T$ , where  $X = (X_t)$  and  $U = (U_t)$  are given by

$$dX_t = -\kappa_x X_t dt + \sigma_x dV_t \quad \text{and} \quad dU_t = -\kappa_u U_t dt + \sigma_u dU_t,$$

where  $V = (V_t)$  and  $U = (U_t)$  are two independent Brownian motions. We let  $\kappa_u > 0$  for both types of regressions, and let  $\kappa_x > 0$  and  $\kappa_x = 0$  for the stationary and cointegrating regressions, respectively.

For the stationary regression, we set the parameter values  $(\kappa_x, \sigma_x) = (0.1020, 1.5514)$  and  $(\kappa_u, \sigma_u) = (6.9011, 2.7566)$ , which are obtained from  $(x_i)$  and  $(\hat{u}_i)$  in Model II fitted to our simulation model. Both  $X$  and  $U$  are therefore specified as stationary Ornstein-Uhlenbeck (OU) processes in our stationary regression. For the cointegrating regression, the parameter values are given as  $(\kappa_x, \sigma_x) = (0, 0.0998)$  and  $(\kappa_u, \sigma_u) = (1.5718, 0.0097)$ , which are identical to those from  $(x_i)$  and  $(\hat{u}_i)$  in Model III fitted with restriction  $\kappa_x = 0$  to our simulation model. In our cointegrating regression,  $X$  therefore becomes a Brownian motion, while  $U$  is a stationary OU process. We consider the test of the null hypothesis  $H_0 : \beta_0 = 0$  and  $\beta_1 = 1$  using the  $H$  and  $G$ -tests respectively for the stationary and cointegrating regressions.

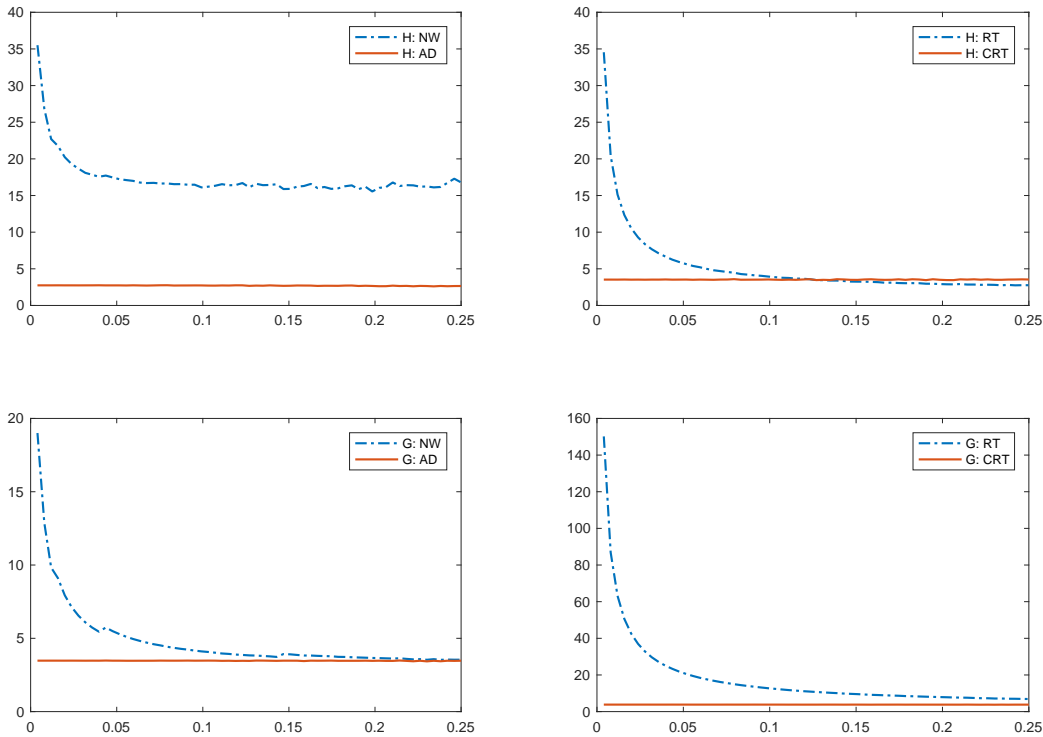
In the simulation, the exact transition densities of Brownian motion and OU process are used to generate daily sample paths of  $X$  and  $U$  with 3,000 iterations. We collect discrete samples at varying frequencies ranging from daily to quarterly levels, which correspond to  $\delta = 1/252$  and  $1/4$  respectively, and compute the 5%  $H$  and  $G$ -tests at each sampling frequency. The long-run variance estimates in the test statistics are obtained using Parzen kernel and four different bandwidth choices including RT, CRT, AD and NW, which are introduced earlier in Section 5. As discussed, CRT and AD are high frequency compatible, while RT and NW are not.

Figure 5 presents the means of the  $H$ -test for the stationary regression and the  $G$ -test for the cointegrating regression computed under the null hypothesis using 5% chi-square critical value, relying on 30 years of simulated data sampled at frequencies varying from daily and quarterly frequencies corresponding to  $\delta = 1/252$  and  $1/4$  respectively.<sup>4</sup> The 5% chi-square critical values are used in the simulation. The  $G$  and  $H$ -tests behave quite similarly on

---

<sup>4</sup>Our simulation with 50 years data shows exactly the same patterns as those in Figure 5.

Figure 5: Simulated Means of Robust Wald Tests



Notes: Presented are the simulated means of the  $H$ -test for the stationary regression (in the upper panel) and the  $G$ -test for the cointegrating regression (in the lower panel) computed under the null hypothesis using 5% chi-square critical value. The bandwidth choices are given by RT, CRT, AD and NW, and the sampling interval  $\delta$  ranges from 1/252 (daily frequency) to 1/4 (quarterly frequency).

average as we change the sampling intervals. Their behaviors are critically dependent upon the high frequency compatibility of the bandwidth used in estimating the required longrun variance. If high frequency compatible bandwidths are used, both tests are insensitive to the sampling intervals and they yield stable test values across all different sampling frequencies. On the other hand, the tests with high frequency incompatible bandwidths are extremely unreliable and they provide test values changing radically as sampling intervals and frequencies vary. In particular, the averages of their test values increase very rapidly as the sampling frequency exceeds the monthly level and approaches the daily level, exactly as we observe for the  $G$  and  $H$ -tests applied to the actual data in Figure 4.

In our simulation, we compute the rejection probabilities, as well as the means, of the  $G$  and  $H$ -tests for the stationary and cointegrating regressions under the null hypothesis using 5% chi-square critical value for  $T = 30$  and 50. They are provided in Figure 6. If

high frequency incompatible bandwidths are used, not only the means but also the rejection probabilities of the tests become highly sensitive to the sampling frequency and increase rapidly as the sampling interval decreases down beyond one month.<sup>5</sup> The rejection probabilities of the tests with high frequency compatible bandwidths are pretty stable across all sampling frequencies, and their overall rejection probabilities tend to get closer to the nominal size as  $T$  increase.

To more directly illustrate our double asymptotics relying on  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  jointly, we also set  $\delta = (1/3)T^{-2}$  so that  $T$  changes along with  $\delta$ . Under this setup, we simulate the means of the  $H$  and  $G$ -tests for the stationary and cointegrating regressions under the null hypothesis using the 5% chi-square critical value, which is reported in Figure 7. The overall pattern in the frequency dependence of the tests is the same as what we observe in Figure 5, where we fix  $T$  and only change  $\delta$ . The only difference is that downward trends exist in the simulated means of the tests with high frequency compatible bandwidths CRT and AD. No such trends show up in our simulation with fixed  $T$ . The trends appear under double asymptotics just because the test values on average get close to the nominal chi-square critical value as  $T$  increases. The simulated means of the tests with high frequency incompatible bandwidths RT and NW behave similarly, regardless of  $T$  being fixed or varying with  $\delta$ .

## 7 Conclusion

The Wald test is widely to test for restrictions in regressions. However, the test is extremely sensitive to the sampling frequency, and it is very likely that the test result depends on the frequency of the samples we use for the test. This, of course, is highly undesirable, since in most cases the sampling frequency has no bearing on the hypothesis to be tested. The dependence of the Wald test on the sampling frequency stands out so vividly in time series regressions. In particular, when observations are used at high frequency, the Wald test almost explode and we are always led to reject the null hypothesis no matter whether it is true or not. The problem, however, appears to be either overlooked or neglected in the literature for long. Certainly, it has now become more important as more economic and financial time series are collected and made available at high frequency these days.

In the paper, we develop a new continuous time framework and develop relevant asymptotics, relying on sampling interval  $\delta$  and time span  $T$  jointly, to analyze the Wald test in regressions with high frequency observations over long span. Our framework accommodates

---

<sup>5</sup>The bandwidths given by RT and NW yield large size distortions even at relatively low frequencies respectively for our stationary and cointegrating simulation models. However, they also have the tendency we describe here.



both stationary and cointegrating regressions, and our asymptotics hold under mild regularity conditions. According to our asymptotic theory, the Wald test is expected to diverge if  $\delta \rightarrow 0$  fast enough relative to  $T \rightarrow \infty$ . This is exactly what we observe in practice. The robust Wald tests, which use the longrun variance estimates in place of the usual variance estimates, are expected to behave similarly unless the bandwidths for their longrun variance estimates are chosen to be high frequency compatible. The high frequency compatibility of a bandwidth ensures that the resulting longrun variance estimate captures the existing serial correlation properly at high frequency. Only when high frequency compatible bandwidths are used, the robust Wald tests become valid and behave as expected under the null hypothesis.

## 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 (2017): “Asymptotics for Recurrent Diffusions with Application to High Frequency Regression,” *Journal of Econometrics*, 196, 37–54.
- LU, Y. AND J. Y. PARK (2018): “Estimation of Longrun Variance of Continuous Time Stochastic Process Using Discrete Sample, forthcoming in *Journal of Econometrics*,” .
- NEWBY, 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

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

from which it follows that

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

However, we have

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

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

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

and

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

and we may deduce the stated result immediately.  $\square$

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

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

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

For the proof of part (a), we write

$$\sqrt{T}(\hat{\beta} - \beta) = \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^n x_i u_i, \quad (18)$$

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

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n x_i x_i' &= \frac{1}{T} \int_0^T X_t X_t' dt + o_p(1) \rightarrow_p M > 0 \\ \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^n x_i u_i &= \frac{1}{\sqrt{T}} \int_0^T X_t U_t dt + o_p(1) \rightarrow_d \mathbb{N}(0, \Pi) \end{aligned}$$

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  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

$$\sqrt{T} \Lambda_T' (\hat{\beta} - \beta) = \left( \frac{1}{n} \sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1'} \right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^n \Lambda_T^{-1} x_i u_i, \quad (19)$$

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

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

and

$$\begin{aligned} \frac{\sqrt{\delta}}{\sqrt{n}} \sum_{i=1}^n \Lambda_T^{-1} x_i u_i &= \frac{1}{\sqrt{T}} \int_0^T \Lambda_T^{-1} X_t U_t dt + o_p(1) \\ &= \int_0^1 X_t^T dU_t^T + o_p(1) \rightarrow_d \int_0^1 X_t^\circ dU_t^\circ \end{aligned}$$

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

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

from which the stated results readily follow.  $\square$

**Proof of Lemma 4.2** Write

$$\tilde{\rho} - 1 = \frac{\sum_{i=1}^n u_{i-1} (u_i - u_{i-1})}{\sum_{i=1}^n u_{i-1}^2}. \quad (20)$$

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

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

and therefore, we may deduce that

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

which will be further analyzed subsequently.

We have

$$\begin{aligned} (U_{i\delta} - U_{(i-1)\delta})^2 &= 2 \int_{(i-1)\delta}^{i\delta} (U_t - U_{(i-1)\delta}) dU_t^c + ([U^c]_{i\delta} - [U^c]_{(i-1)\delta}) \\ &\quad + \sum_{(i-1)\delta < t \leq i\delta} \Delta(U_t - U_{(i-1)\delta})^2, \end{aligned} \quad (21)$$

and

$$\begin{aligned}
\Delta(U_t - U_{(i-1)\delta})^2 &= (U_t - U_{(i-1)\delta})^2 - (U_{t-} - U_{(i-1)\delta})^2 \\
&= (U_t - U_{t-})(U_t + U_{t-} - 2U_{(i-1)\delta}) \\
&= (U_t - U_{t-}) \left[ (U_t - U_{t-}) + 2(U_{t-} - U_{(i-1)\delta}) \right] \\
&= 2(U_{t-} - U_{(i-1)\delta})\Delta U_t + (\Delta U_t)^2
\end{aligned} \tag{22}$$

for  $i = 1, \dots, n$ . Therefore, it follows from (21) and (22) that

$$\sum_{i=1}^n (U_{i\delta-} - U_{(i-1)\delta})^2 = [U]_T + 2Z_T, \tag{23}$$

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

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

Note that

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

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

$$U_{t-} - U_{(i-1)\delta} = (U_t^c - U_{(i-1)\delta}^c) + \sum_{(i-1)\delta < s < t} \Delta U_s, \tag{24}$$

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

To consider  $Z^c$ , we write

$$Z^c = Z^a + Z^b, \tag{25}$$

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

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

and

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

from which it follows that

$$Z_T^a = O_p(p_T T \Delta_{\delta,T}(U)) + O_p(p_T T \delta) = O_p(p_T T \Delta_{\delta,T}(U)), \quad (26)$$

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

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

Furthermore, it follows that

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

and that

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

Therefore, we may deduce that

$$Z_T^b = O_p \left( \sqrt{q_T T} \Delta_{\delta, T}(U) \right) + O_p \left( \sqrt{q_T T \delta} \right) = O_p \left( \sqrt{q_T T} \Delta_{\delta, T}(U) \right), \quad (27)$$

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

$$\begin{aligned} Z_T^c &= O_p(p_T T \Delta_{\delta, T}(U)) + O_p \left( \sqrt{q_T T} \Delta_{\delta, T}(U) \right) \\ &= T \left[ O_p(p_T \Delta_{\delta, T}(U)) + O_p \left( \sqrt{q_T / T} \Delta_{\delta, T}(U) \right) \right] \end{aligned} \quad (28)$$

from (26) and (27).

To analyze  $Z^d$ , we let

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

We have

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} (U_t^c - U_{(i-1)\delta}^c) \Delta U_t \right]^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \mathbb{E}(\Delta U_t)^2 \\
&\leq \left[ \max_{1 \leq i \leq n} \sup_{(i-1)\delta < t \leq i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \right] \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} \mathbb{E}(\Delta U_t)^2 \\
&= \left[ \max_{1 \leq i \leq n} \sup_{(i-1)\delta < t \leq i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \right] \sum_{0 < t \leq T} \mathbb{E}(\Delta U_t)^2 = O(T \Delta_{\delta, T}^2(U)). \quad (29)
\end{aligned}$$

Moreover, we may easily deduce that

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} \left( \sum_{(i-1)\delta < s < t} \Delta U_s \right) \Delta U_t \right]^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} \mathbb{E} \left( \sum_{(i-1)\delta < s < t} \Delta U_s \right)^2 \mathbb{E}(\Delta U_t)^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \leq i\delta} \left[ \sum_{(i-1)\delta < s < t} \mathbb{E}(\Delta U_s)^2 \right] \mathbb{E}(\Delta U_t)^2 \\
&\leq \sum_{i=1}^n \sum_{(i-1)\delta < s, t \leq i\delta} \mathbb{E}(\Delta U_s)^2 \mathbb{E}(\Delta U_t)^2 \\
&\leq \sum_{i=1}^n \sum_{(i-1)\delta < s, t \leq i\delta} \mathbb{E}(\Delta U_t)^4 = \sum_{0 < t \leq T} \mathbb{E}(\Delta U_t)^4 = O(T). \quad (30)
\end{aligned}$$

Therefore, it follows from (29) and (30) that

$$Z_T^d = O_p(\sqrt{T} \Delta_{\delta, T}(U)) + O_p(\sqrt{T}) = O_p(\sqrt{T}) = o_p(T) \quad (31)$$

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



Now we have, due in particular to (28) and (31),

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

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ . Consequently, the stated result follows immediately from (20), and the proof is complete.  $\square$

**Proof of Lemma 5.1** For the Andrews bandwidth, we may readily deduce from Lemma 4.2 that

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

and in particular  $\tilde{\theta}_r^2 = O_p(\delta^{-2r})$  for  $r = 1, 2$  as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  satisfying Assumption E. With these choices of  $\theta_r^2$ ,  $r = 1, 2$ , the feasible optimal bandwidth  $\tilde{b}_n^*$  in (16) becomes

$$\tilde{b}_n^* = \left( \frac{\tau^4\pi_1^2}{\sigma^4 \int K(x)^2 dx} n \right)^{1/3} \delta^{-2/3} (1 + o_p(1)) \quad \text{or} \quad \left( \frac{8\tau^8\pi_2^2}{\sigma^8 \int K(x)^2 dx} n \right)^{1/5} \delta^{-4/5} (1 + o_p(1)),$$

and therefore,

$$B_{n,\delta} = \tilde{b}_n^* \delta = \left( \frac{\tau^4\pi_1^2}{\sigma^4 \int K(x)^2 dx} T \right)^{1/3} (1 + o_p(1)) \quad \text{or} \quad \left( \frac{8\tau^8\pi_2^2}{\sigma^8 \int K(x)^2 dx} T \right)^{1/5} (1 + o_p(1))$$

depending upon whether  $r = 1$  or  $2$ , as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  satisfying Assumption E. Consequently, we have  $B_{n,\delta} \rightarrow_p \infty$  and  $B_{n,\delta}/T \rightarrow_p 0$ , which implies that the Andrews bandwidth is high-frequency compatible.

For the Newey-West bandwidth, note that for  $r = 1, 2$ ,

$$\delta^r \hat{\theta}_r = \frac{\delta^{1+r} \sum_{|j| \leq a_n} |j|^r \gamma_n(j)}{\delta \sum_{|j| \leq a_n} \gamma_n(j)}, \quad (32)$$

and we will show in the following that as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ ,

$$\delta^{1+r} \sum_{|j| \leq a_n} |j|^r \gamma_n(j) \sim_p \int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds \quad (33)$$

for  $r = 0, 1, 2$ , where  $A_{n,\delta} = a_n\delta$ . To see this, we write

$$\int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds - \delta^{1+r} \sum_{|j| \leq a_n} |j|^r \gamma_n(j) = P_{n,\delta} + Q_{n,\delta} + R_{n,\delta},$$

where  $P_{n,\delta}, Q_{n,\delta}$  and  $R_{n,\delta}$  are the same as those in the proof of Proposition 4.2 in Lu and Park (2018) with  $A_T$  replaced by  $A_{n,\delta}$ , and the analyses of their stochastic orders are also analogous. In particular, given  $\sup_{0 \leq t \leq \infty} \mathbb{E}|V_t|^2 < \infty$ , Lemma A.2 in Lu and Park (2018) holds, and we have  $P_{n,\delta} = O(\Delta_{\delta,T}(V)A_{n,\delta}^{1+r})$ ,  $Q_{n,\delta} = O_p(\Delta_{\delta,T}(V)A_{n,\delta}^{1+r} + \delta A_{n,\delta}^r)$  and  $R_{n,\delta} = O(\delta A_{n,\delta}^r)$ . Therefore,

$$\int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds - \delta^{1+r} \sum_{|j| \leq a_n} |j|^r \gamma_n(j) = O(\Delta_{\delta,T}A_{n,\delta}^{1+r} + \delta A_{n,\delta}^r). \quad (34)$$

Note that as  $A_{n,\delta} = cn^p\delta = c\delta^{1-p}T^p \rightarrow 0$  as  $\delta \rightarrow 0$  sufficiently fast relative to  $T \rightarrow \infty$ , so we have

$$A_{n,\delta}^{-(1+r)} \int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds = \int_{|s| \leq 1} |s|^r \left( \frac{1}{T} \int_0^T V_t V_{t-sA_{n,\delta}} dt \right) ds \rightarrow_p \frac{2\sigma^2}{1+r}$$

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$ . This implies  $A_{n,\delta}^{-(1+r)} [\delta^{1+r} \sum_{|j| \leq a_n} |j|^r \gamma_n(j) - \int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds] = O(\Delta_{\delta,T}(V) + \delta/A_{n,\delta}) = O(\Delta_{\delta,T}(V) + \delta^p T^{-p}) = o_p(1)$  which proves (33). Now we can readily deduce from (32), (33) and (34) that

$$A_{n,\delta}^{-r} \delta^r \hat{\theta}_r \sim_p A_{n,\delta}^{-r} \frac{\int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds}{\int_{|s| \leq A_{n,\delta}} \Gamma_T(s) ds} = \frac{A_{n,\delta}^{-(1+r)} \int_{|s| \leq A_{n,\delta}} |s|^r \Gamma_T(s) ds}{A_{n,\delta}^{-1} \int_{|s| \leq A_{n,\delta}} \Gamma_T(s) ds} \rightarrow_p \frac{1}{1+r},$$

and therefore for  $B_{n,\delta} = \hat{b}_n^* \delta = \left( r\pi_r \hat{\theta}_r^2 n / \int K(x)^2 dx \right)^{1/(2r+1)} \delta$ , we have

$$A_{n,\delta}^{-2r/(2r+1)} T^{-1/(2r+1)} B_{n,\delta} \rightarrow_p \left( \frac{r\pi_r^2}{(1+r)^2 \int K(x)^2 dx} \right)^{1/(2r+1)}$$

from which it follows that

$$\delta^{-2r(1-p)/(2r+1)} T^{-(2pr+1)/(2r+1)} B_{n,\delta} \rightarrow_p \left( \frac{rc^{2r}\pi_r^2}{(1+r)^2 \int K(x)^2 dx} \right)^{1/(2r+1)}$$

given  $A_{n,\delta} = c\delta^{1-p}T^p$ . Therefore,  $B_{n,\delta} \rightarrow_p 0$  if  $\delta \rightarrow 0$  fast enough relative to  $T \rightarrow \infty$  so that  $\delta = o(T^{-(2pr+1)/(2r(1-p))})$ , and hence Newey-West bandwidth is not high-frequency compatible.  $\square$

**Proof of Theorem 5.2** To show part (a), we follow the proof in part (a) of Theorem 4.1 and note that  $\delta \hat{\Omega} \rightarrow_p \Pi$  under Assumption F, due to Theorem 4.1 in Lu and Park (2018).

Then it readily follows that

$$\begin{aligned}
H(\hat{\beta}) &= \sqrt{T}(\hat{\beta} - \beta)' R' \left[ R \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \delta \hat{\Omega} \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} R' \right]^{-1} R \sqrt{T}(\hat{\beta} - \beta) \\
&= \sqrt{T}(\hat{\beta} - \beta)' \left[ \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \delta \hat{\Omega} \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \right]^{-1} \sqrt{T}(\hat{\beta} - \beta) \\
&\sim_p \left( \frac{1}{\sqrt{T}} \int_0^T X_t U_t dt \right)' \Pi^{-1} \left( \frac{1}{\sqrt{T}} \int_0^T X_t U_t dt \right) \rightarrow_d \chi_q^2
\end{aligned}$$

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  under Assumption A, C1 and D1. The proof of part (b) is analogous to the proof of part (b) of Theorem 4.1. Specifically, we first write (19) and note that  $\delta \hat{\omega}^2 \rightarrow_p \varpi^2$  under Assumption F. Therefore,

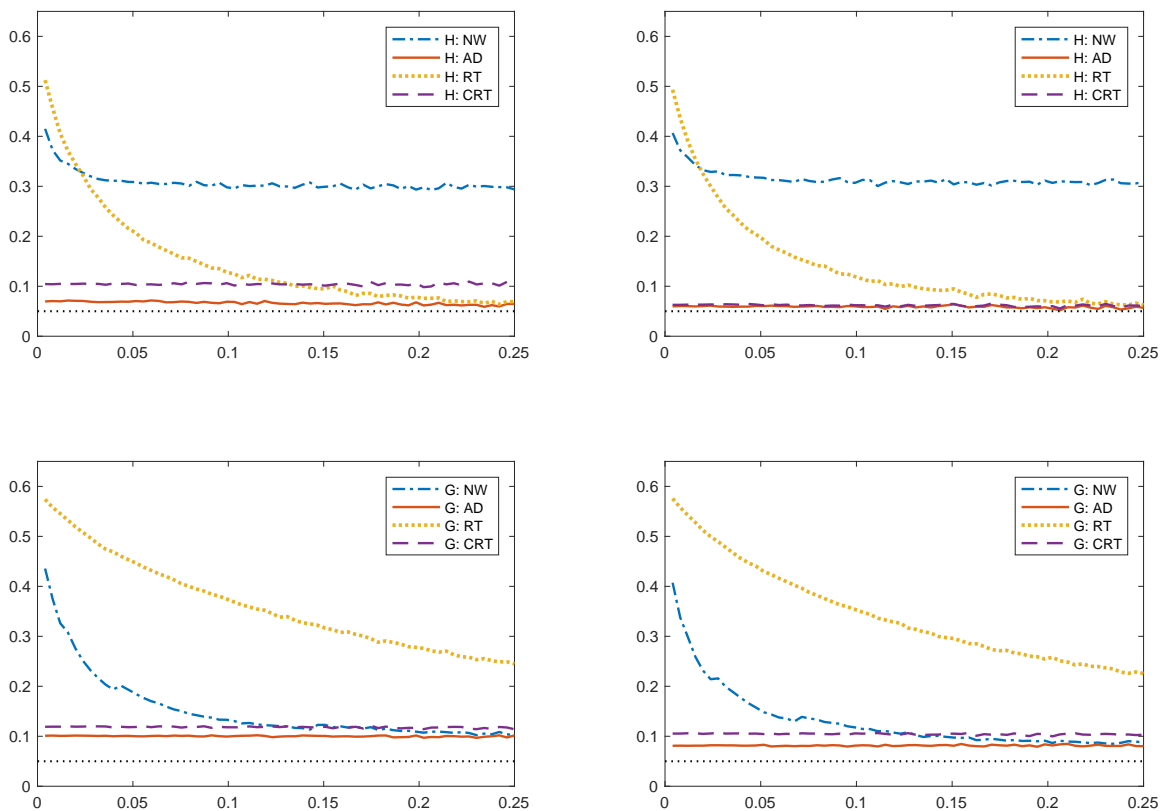
$$\begin{aligned}
G(\hat{\beta}) &= \sqrt{T}(\hat{\beta} - \beta)' R' \left[ R \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} R' \right]^{-1} R \sqrt{T}(\hat{\beta} - \beta) / \delta \hat{\omega}^2 \\
&= \sqrt{T}(\hat{\beta} - \beta)' \Lambda_T \left( \frac{1}{n} \sum_{i=1}^n \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1'} \right) \sqrt{T} \Lambda_T' (\hat{\beta} - \beta) / \delta \hat{\omega}^2 \\
&\sim_p \left( \int_0^1 X_t^T dU_t^T \right)' \left( \int_0^1 X_t^T X_t^{T'} dt \right)^{-1} \left( \int_0^1 X_t^T dU_t^T \right) / \varpi^2
\end{aligned}$$

as  $\delta \rightarrow 0$  and  $T \rightarrow \infty$  under Assumption A and D2, from which the desired result follows immediately from Assumption C2.  $\square$

## B Additional Figures

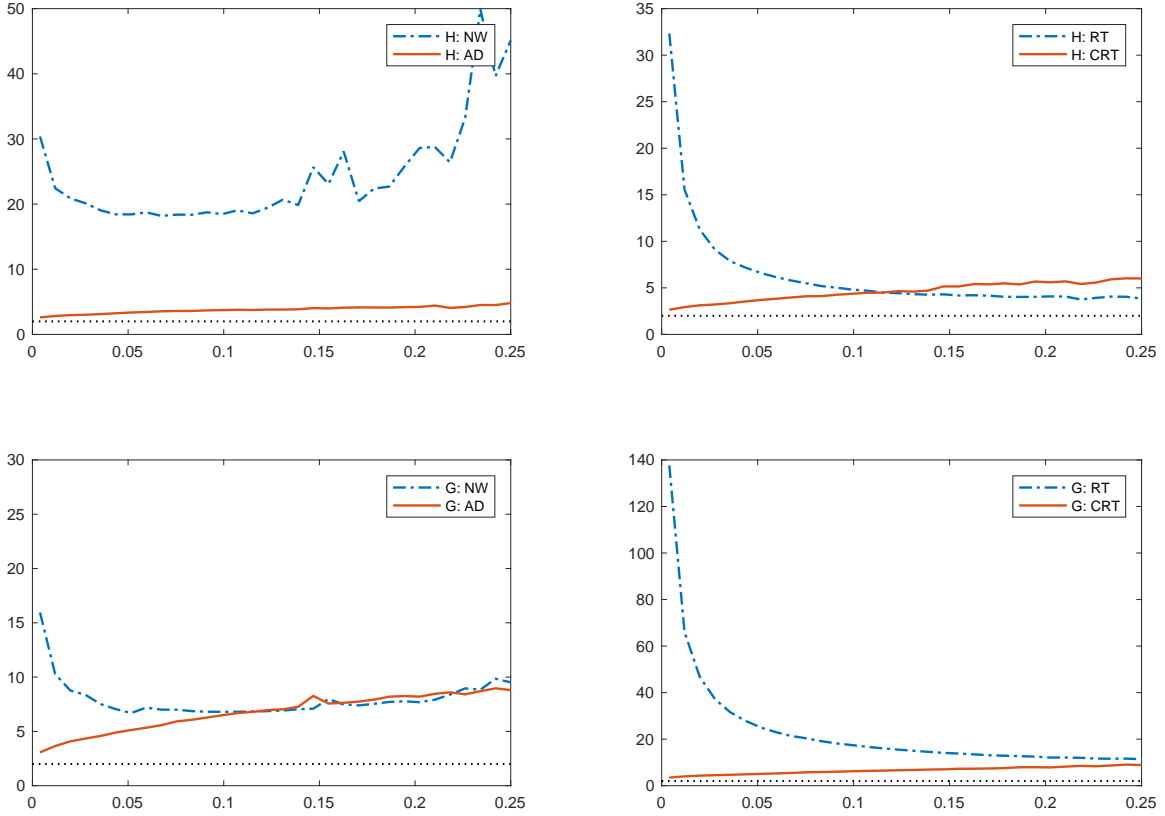
In this section, we present some additional figures. On the other hand, Figure ?? and ?? presents the values of the  $G$ -statistics at different frequencies for Models I and II. As discussed, the underlying stochastic processes generating regressands and regressors for Models I and II appear to be at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

Figure 6: Simulated Rejection Probabilities



Notes: Presented are the simulated rejection probabilities of the  $H$ -test (in the upper panel) for the stationary regression and the  $G$ -test (in the lower panel) for the cointegrating regression with  $T = 30$  (in the left panel) and  $T = 50$  (in the right panel) computed under the null hypothesis using 5% chi-square critical value represented by the dotted line. The bandwidth choices are given by RT, CRT, AD and NW, and the sampling interval  $\delta$  ranges from 1/252 (daily frequency) to 1/4 (quarterly frequency).

Figure 7: Simulated Means of Robust Wald Tests Under Double Asymptotics



Notes: Presented are the simulated means of the  $H$ -test (in the upper panel) for the stationary regression and the  $G$ -test (in the lower panel) for the cointegrating regression computed under the null hypothesis using 5% chi-square critical value represented by the dotted line. The time span  $T$  is set to be a function of the sampling intervals  $\delta$  as  $T = 3\delta^{-1/2}$ . The bandwidth choices are given by RT, CRT, AD and NW, and the sampling interval  $\delta$  ranges from  $1/252$  (daily frequency) to  $1/4$  (quarterly frequency).