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

March 21, 2023

Abstract

In this paper, we analyze regressions with observations collected at small time intervals over a 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 setup, we show that the standard Wald statistic diverges to infinity and the regression becomes spurious as long as $\delta \to 0$ sufficiently fast relative to $T \to \infty$. 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 long-run 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, long-run variance estimation

*We would like to thank Torben Andersen, Donald Andrews, Federico Bandi, Barbara Rossi, and Jihyun Kim for helpful discussions, and to the seminar attendants at University of Washington, Michigan, NYU, Bank of Portugal, Carlos III University of Madrid, UPF, and CEMFI for useful discussions.

1 Introduction

A great number of economic and financial time series are now collected and made available at high frequencies, and naturally 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 that 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 for a class of high frequency regressions.[1] 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 large class of regression models, and therefore, our asymptotics are expected to be widely applicable in practical applications.

One of the main findings from our analysis is that both types of regressions eventually become spurious 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 & Newbold (1974) and studied theoretically

[1]As will be explained later, we only consider the class of high frequency regressions consisting of 'stock' variables in the paper. We have other classes of high frequency regressions relying on 'flow' variables, for which our analysis here does not apply.

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 continuous 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 long-run 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 long-run variance estimator used in the test.[2] 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, the test 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 has a critical effect on the validity of robust tests. For instance, the test becomes generally 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 & West (1994) is employed. In all actual regressions we consider in the paper, the tests behave exactly as predicted by our asymptotic theory.

It seems possible to find a discrete time model, which yields a particular feature of our asymptotics based on the continuous time regression. For instance, we may use a near-unit root model to demonstrate how and why a test fails to perform properly in the presence of strong persistence, similarly as we observe in our analysis. This possibility is explored by Müller (2005) for the test of stationarity. However, the near-unit root model cannot generate any other features of high frequency regressions that our framework and analysis reveal. A majority of previous works, which use continuous time models to study discrete time observations, rely on the limit approximation by a continuous time process of a discrete time series model varying with the sample size n like the near-unit root model, see, e.g., Perron (1991a,b). Our framework is totally different: we assume that discrete samples are collected from a given continuous time model. In our analysis, we fix a continuous time model, and study the characteristics of discrete samples as we change the sampling interval δ and the sample span T.

[2]The basic asymptotics of HAC estimation in continuous time is developed in Lu & Park (2019), which was written concurrently with this paper. This paper extends them to allow for data-dependent bandwidth selection procedures, and also establishes general sufficient conditions for the discretization errors being negligible asymptotically, which are widely valid for both stationary and nonstationary continuous time regressions.

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 long-run variance estimators. We also demonstrate that the bandwidth selection is important, and that the modified tests may or may not yield spurious results depending upon the bandwidth choice. An empirical illustration is provided in Section 6, the simulations reported in Section 7 demonstrate the practical relevance of our asymptotic theory. Section 8 concludes the paper, and Appendices provide mathematical proofs, additional simulation results and additional figures.

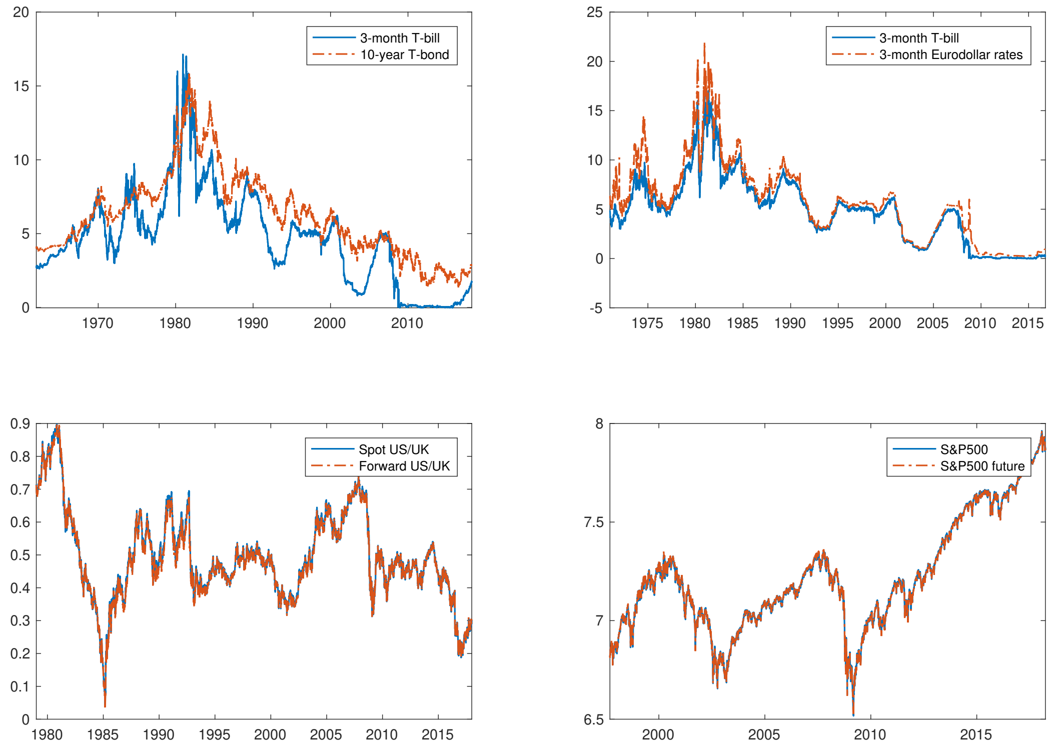
In this paper, the standard notations such as $\to_p$ and $\to_d$ are used to refer to the convergence in probability and in distribution, respectively. We denote by $[U]_T$ the quadratic variation of a continuous time process U over time interval $[0, T]$. Also, $P \sim_p Q$ signifies $P = Q(1 + o_p(1))$ and $P =_d Q$ signifies P and Q follow the same distribution.

2 Background and Motivation

It is widely observed that test results are critically dependent upon the choice of sampling frequency in many time series regressions. To illustrate more explicitly the dependency on the sampling interval of 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 β_0 and β_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 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

Figure 1: Data Plots for Models I–IV



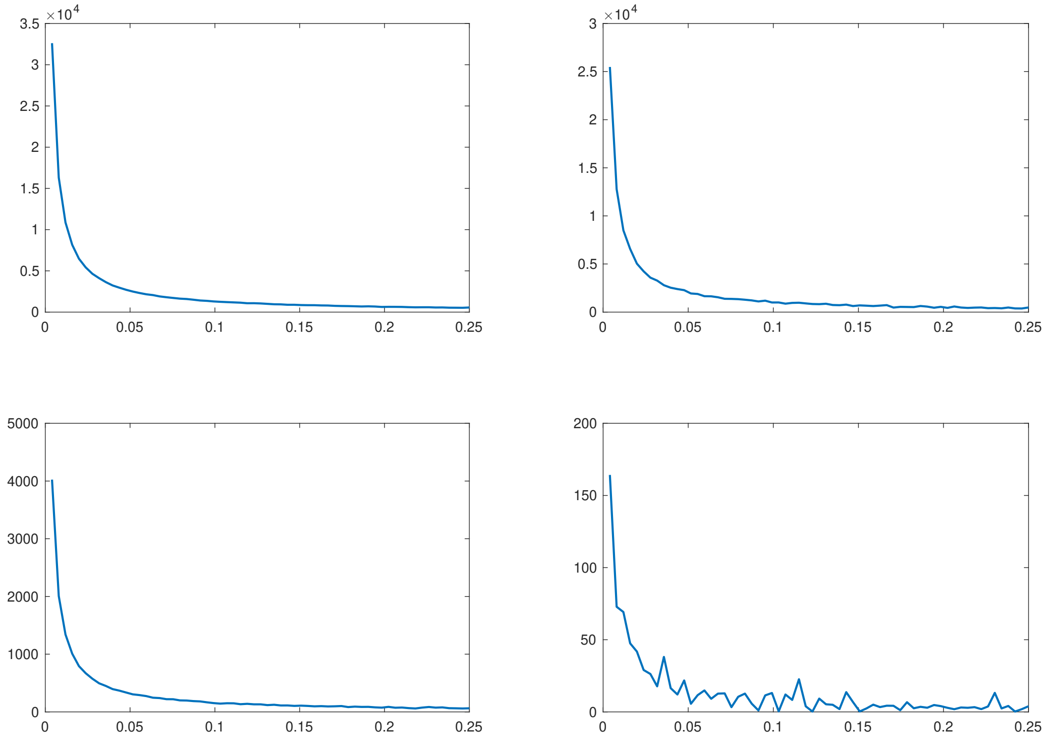
Notes: Presented are daily sample paths of the regressand y and the regressor x used for the empirical illustrations of Models I–IV. Top-left panel presents the 10-year T-bond rate and 3-month T-bill rate from January 3, 1962 to March 29, 2018. Top-right panel presents the 3-month Eurodollar rate and 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.

We consider Models I/II primarily as stationary regressions, and Models III/IV as cointegrating regressions. However, whether or not Models I/II and Models III/IV are truly specified respectively as stationary and cointegrating regressions is not important. Our asymptotics accommodate both types of regressions.

The plots for (y_i) and (x_i) in Models I-IV are given in Figure 1.[3] 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 hypothesis to be tested appears to be $\mathrm{H}_0 : \beta_0 = 0$ and $\beta_1 = 1$. The hypothesis

[3]The S&P index futures used here and elsewhere in the paper are the log prices of the E-mini S&P 500 contracts expiring in the front month (the nearest delivery month) downloaded from Investing.com. The E-mini S&P 500 contract represents one-fifth of the value of the standard-size futures contract, and it is traded electronically on the Chicago Mercantile Exchange (CME). The E-mini contract has been traded much more actively and widely than the standard contract, and it has become the primary futures trading vehicle for the S&P 500 index since the inception its trading on September 9, 1997.

Figure 2: Wald Tests in Models I-IV



Notes: Presented are the values of the Wald statistics for the null hypothesis $H_0 : \beta_0 = 0$ and $\beta_1 = 1$ in Models I-IV (Model I in the upper-left, Model II in the upper-right, Model III in the lower-left, and Model IV in the lower-right panel) plotted against the varying sampling intervals δ ranging from 1/252 (daily frequency) to 1/4 (quarterly frequency).

may or may not hold. In particular, the null hypothesis $\beta_0 = 0$ does not necessarily hold when 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 standard Wald test relying on 5% chi-square critical values 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 dramatically 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 values of the Wald test as δ 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 explode at daily frequency.

Consequently, 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.[4]

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 $\delta \to 0$ as well as $T \to \infty$, it actually becomes clear that the test is expected 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, Setup 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

[4] 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[5], 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 statistic 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 long-run variance of (u_i) based on $(\hat{u}_i)$, or

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

where $\hat{\Omega}$ is a consistent estimator for the long-run variance of $(x_i u_i)$ based on $(x_i \hat{u}_i)$. We consider two different types of regressions 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.[6]

We analyze regression (1), when (y_i) and (x_i) are high frequency observations.[7] 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., we let

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

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 $T = n\delta$. Our asymptotics will be obtained by letting $\delta \to 0$ and $T \to \infty$ jointly. Under our setup, the regression model

[5] Without loss of generality, we assume R is of full row rank and is in *reduced row canonical form*. This means R is in row canonical form with all pivots 1, and each column of R containing a leading 1 has zeros in all its other entries. Any system of non-redundant restrictions can be reduced to its most parsimonious form $R\beta = r$ in which R is a full row rank matrix in reduced row canonical form, by Gaussian-Jordan elimination.

[6] Note that the long-run variance of $(x_i u_i)$ does not exist if (x_i) is nonstationary.

[7] 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.

introduced in (1) can be analyzed using the corresponding continuous time regression

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

for $0 \leq t \leq T$, 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 respectively.

The underlying continuous time model in (5), which we rewrite as $Y_t = \alpha + \beta X_t + U_t$ in its simplest form for our discussions below, is introduced to study the type of high frequency regressions we consider in Section 2. There are other continuous time models that are more suitable to analyze different types of regressions relying on high frequency observations. Most notably, we may consider a continuous time regression given by

$$dY_t = (\alpha + \beta X_t)dt + dU_t, \tag{6}$$

where $dU = (dU_t)$ is a martingale differential error, which is often specified further as $dU_t = \sigma_t dW_t$ for $t \geq 0$ with instantaneous volatility $\sigma = (\sigma_t)$ and Brownian motion $W = (W_t)$. This type of regression is used in Choi et al. (2016) to investigate stock return predictability. Moreover, we may also look at a continuous time regression of the form

$$dY_t = \alpha dt + \beta dX_t + dU_t, \tag{7}$$

where dU is defined as in (6). See, e.g., Chang et al. (2016), which uses regression in (7) to evaluate factor pricing models.

In continuous time regressions, the differentials dY and dX should be used to represent *flow* variables like stock or portfolio returns over an infinitesimal interval, whereas the levels Y and X are more appropriate to specify *stock* variables such as stock prices, interest rates and predictive ratios. Also note that we need the dt term to include a constant term or to relate any stock variable X to a flow variable dY. In the paper, we only consider the continuous time regression in *levels*, which is useful to analyze the high frequency regression modeling a relationship between a 'stock' variable and other 'stock' covariates. Therefore, we focus on the continuous time regression (5), and will not consider other types of continuous time regressions such as (6) and (7). The asymptotics of these continuous time regressions are quite distinct from our asymptotics in the paper.

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^2, XX' or XU. We have

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

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

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

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

The conditions in Assumption A are very mild and they are satisfied by virtually all stochastic processes used in both theoretical and empirical applications. The first condition is 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 holds trivially for a wide class of stochastic processes. 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 setup and therefore it may be regarded as the global modulus of continuity. Typically, we have

$$\Delta_{\delta,T}(Z) = \delta^{1/2-\epsilon}\lambda_T \tag{8}$$

for some $\epsilon \ge 0$ and a nonrandom sequence (λ_T) of T that is bounded away from zero and, as an example, the condition is clearly satisfied if $\sup_{0\le t\le T} |Z_t| = O_p(T)$,[8] regardless of how we set δ and T as long as $\delta \to 0$ and $T \to \infty$. The condition holds in any case as long as $\delta \to 0$ fast enough relative to $T \to \infty$.[9]

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* $Z = U^2, XX'$ *or* XU *and* $z_i = Z_{i\delta}$ *for*

[8]Note that we have $\sup_{0\le t\le T} |Z_t| = O_p(\sqrt{T})$ if Z is Brownian motion. Therefore, the condition here is very mild and expected to be satisfied even by some explosive processes.

[9]This is a necessary condition to guarantee that discretization errors become negligible asymptotically and our asymptotics are determined solely by those of the underlying continuous time processes. If, for instance, we set $\delta \to 0$ slowly relative to $T \to \infty$, our asymptotics in the paper will not be applicable. Sufficient conditions for the applicability of our asymptotics are dependent on the underlying model and statistical procedure we use, and they will be introduced later.

$i = 1, \ldots, n$, 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 $Z = U^2, XX'$ 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 as

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

on $[0,1]$ with an appropriate nonsingular normalizing sequence (Λ_T) of matrices, 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 $D[0,1]$ jointly with $X^T \to_d X^\circ$ in the product space of $D[0,1]$ 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.

Assumptions C1 and C2 are introduced for the stationary regression and the cointegrating regression, respectively. They are not stringent and expected to hold for a wide class of regressions in continuous time. Assumption C1 is the continuous analog of the standard assumptions for stationary regressions in discrete time. In particular, Assumption C1(a) is a continuous-time law of large numbers for additive functionals, which holds if X_t has finite second moment; Assumption C1(b) is a continuous-time central limit theorem for additive functionals, see e.g. Rozanov (1960), and the sufficient conditions, which include a mixing condition and a moment condition, are similar to those in the discrete-time setting. Assumption C2(a) is satisfied for general null recurrent diffusions and jump diffusions, as shown by Jeong & Park (2011), Jeong & Park (2014) and Kim & Park (2017). Moreover, Assumption C2(b) is the continuous time version of the usual invariance principle. Such an invariance principle holds for a class of functions of continuous-time stationary ergodic Markov process, see e.g. Theorem 2.1 in Bhattacharya (1982) and more discussion in Section 2 of Lu & Park (2019). Note that we assume in Assumption C1 that X and U are uncorrelated as in the standard regression model. However, in Assumption C2, we allow X and U to be dependent with each other in an arbitrary manner.

In parallel with Assumptions C1 and C2, respectively for the stationary and cointegrating regressions, we introduce Assumptions D1 and D2 below.

Assumption D1. $\Delta_{\delta,T}(U^2), \Delta_{\delta,T}(\|XX'\|) \to_p 0$ and $\sqrt{T}\Delta_{\delta,T}(\|XU\|) \to_p 0$ as $\delta \to 0$ and $T \to \infty$.

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

The conditions in Assumptions D1 and D2 hold if $\delta \to 0$ fast enough relative to $T \to \infty$. If the sample paths of U and X are bounded and differentiable with bounded derivatives, then $\Delta_{\delta,T}(\|Z\|) = O_p(\delta)$ for $Z = U^2, XX'$ and XU and $\|\Lambda_T\| = O_p(1)$. In this case, it suffices to have $\delta = o(T^{-1/2})$. Generally, we require $\delta \to 0$ faster relative to $T \to \infty$ if the underlying processes U and X are more volatile locally and more explosive globally. If they behave all like Brownian motion, for instance, then we may easily deduce that $\Delta_{\delta,T}(\|Z\|) = O_p\left(\delta^{1/2-\epsilon}T^{1/2+\epsilon}\right)$ with an arbitrarily small $\epsilon > 0$ for $Z = U^2, XX'$ and XU from, e.g., Kanaya et al. (2018). Therefore, the conditions in Assumption D1 are met if we set $\delta = O\left(T^{-2-\epsilon}\right)$ for any $\epsilon > 0$ arbitrarily small. To satisfy the conditions in Assumption D2, we should have $\delta \to 0$ faster, but so much unless X is overly explosive. If $\|\Lambda_T\| = O(\sqrt{T})$ as in the case of X being given by Brownian motion, we only need to require $\delta = O\left(T^{-3-\epsilon}\right)$ for $\epsilon > 0$ arbitrarily small. In sum, our conditions here are not very stringent.

In our asymptotic analysis, we let $\delta \to 0$ and $T \to \infty$ jointly, under either 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 specified in Assumption D1 or D2. In both of 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 the OLS estimator $\hat{\beta}$ of β in regression (1) and analyze the asymptotic behaviors of the standard Wald test based on the test statistic $F(\hat{\beta})$ in (2) under the null hypothesis $\mathrm{H}_0 : R\beta = r$.

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 always leads us to reject the null hypothesis when it is correct, and the asymptotic size becomes unity, as $\delta \to 0$ and $T \to \infty$. The regressions therefore become spurious in this sense.

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 & 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 long-run relationship diverges to infinity, implying falsely the presence of cointegration. Granger & 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 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[10]

$$u_i = \rho u_{i-1} + \varepsilon_i, \tag{9}$$

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

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

and $p_T \Delta_{\delta,T}(U) \to_p 0$ and $(q_T/\sqrt{T})\Delta_{\delta,T}(U) \to_p 0$ with $\delta = O_p(\Delta^2_{\delta,T}(U))$ 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 $T \to \infty$.

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 δ 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 jumps increases at T-rate and jump size has finite fourth moment, and if the instantaneous variance of U is asymptotically stationary. Note that $[U]_T = [U^c]_T + \sum_{0 \le t \le T} (\Delta U_t)^2$ and $[U^c]_T = [M]_T$.

The asymptotics for the estimated AR coefficient $\tilde{\rho}$ of ρ in (9) 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 $\delta \to 0$ and $T \to \infty$.

[10] We use an AR(1) specification for (u_i) here not as a true model but as a fitted model. Clearly, AR(1) processes cannot be defined meaningfully in a continuous time framework.

It follows immediately from Lemma 4.2 that

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

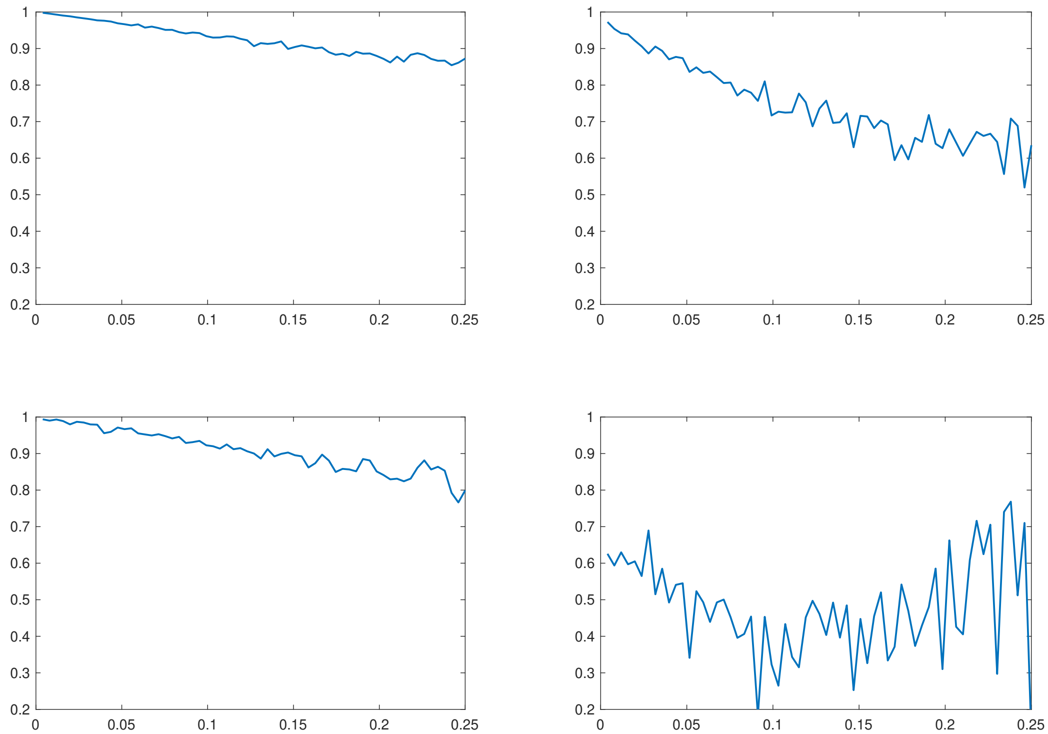
particularly as $\delta \to 0$. Therefore, (u_i) becomes strongly dependent, and the regression becomes spurious as the sampling interval δ approaches 0. Our regression is completely analogous in this regard to the conventional spurious regression, except that we let $\delta \to 0$ in contrast with the conventional spurious regression requiring $n \to \infty$.[11] 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 a 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. If T is assumed to be fixed and set at $T = 1$ without loss of generality, we have $\delta = 1/n$ and (u_i) asymptotically reduces to a near unit root process with AR coefficient $\rho = 1 - c/n$ 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, contrarily to our assumption, 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. In the paper, 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 away from the unity as δ increases depends on the ratio τ^2/σ^2. Roughly, τ^2 measures the mean local variation, while σ^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, the slower $\tilde{\rho}$ converges to the unity as $\delta \to 0$. 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 the Ornstein-Uhlenbeck process given by $dU_t = -\kappa U_t dt + \upsilon dW_t$, we have $\tau^2 = \upsilon^2$ and $\sigma^2 = \upsilon^2/2\kappa$. Therefore, the ratio becomes $\tau^2/\sigma^2 = 2\kappa$, which becomes larger if we have stronger mean reversion. The local-to-global variation ratios of U estimated using the fitted residuals of Models I-IV are 1.08, 14.30, 3.16 and 169.9, respectively. The ratio is particularly large for Model IV, which explains why its estimated residual AR coefficient is not quite close to the unity at the highest frequency, daily.

[11] Clearly, our regression is not directly comparable to the conventional spurious regression. Unlike the classical spurious regression having an error process that is nonstationary and has a stochastic trend, we set the error process U to be essentially stationary in our model.

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 (Model I in the upper-left, Model II in the upper-right, Model III in the lower-left, and Model IV in the lower-right panel), plotted against various sampling intervals δ ranging from 1/252 (daily frequency) to 1/4 (quarterly frequency).

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 of the autoregressive coefficients tend to increase as the sampling interval shrinks. In particular, except for Model IV, the estimates approach unity as the sampling interval decreases. This is exactly what we expect from Lemma 4.2. Model IV is rather exceptional. For Model IV, the estimated autoregressive coefficients do not show any monotonous increasing trend, unlike all other models. In fact, Lemma 4.2 does not seem to apply for Model IV, which alludes that Assumption E does not hold for Model IV. This is perhaps due to more irregular and frequent jump activities present in stock prices than are allowed in Assumption E. 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 of the 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_iu_i), $(c'x_iu_i)$ 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 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 as the G and H-tests for short. These robust test statistics involve the long-run variance estimators of (u_i) and (x_iu_i), instead of the usual variance estimators used for the standard Wald test, to take into account the presence of serial correlation in (u_i) and (x_iu_i). Our asymptotics in this section show that, in general, the spuriousness in regression at high frequency we observe and analyze in the 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_iu_i), which necessarily emerges in high frequency regression. The spuriousness disappears only when we properly estimate the long-run variances of (u_i) and (x_iu_i) using bandwidths chosen suitably.

Under the null hypothesis, we may expect that

$$G(\hat{\beta}) \approx P_T'R'\left[RQ_T^{-1}R'\right]^{-1}RP_T/(\delta\hat{\omega}^2)$$
$$H(\hat{\beta}) \approx P_T'R'\left[RQ_T^{-1}(\delta\hat{\Omega})Q_T^{-1}R'\right]^{-1}RP_T \tag{10}$$

with

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

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

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

$$\pi_n^2 = \sum_{|j|\leq n} K\left(\frac{j}{b_n}\right)\gamma_n(j), \tag{11}$$

where K is the kernel function, $\gamma_n(j) = n^{-1}\sum_i v_iv_{i-j}$ is the sample autocovariance function of (v_i) and b_n is the bandwidth parameter.[12] If δ is fixed and (v_i) has a well defined long-run variance π^2, then we would expect $\pi_n^2 \to_p \pi^2$ as long as $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$.[13] If we set $\delta \to 0$ as in our setup, however, π_n^2 behaves much differently in asymptotics.

The type of long-run variance estimator in (11) defined with high frequency observations is analyzed in Lu & Park (2019). Following them, we relate it to the long-run variance estimator ϖ_T^2 of the underlying continuous time process V, which is given by

$$\varpi_T^2 = \int_{|s|\le T} K\left(\frac{s}{B_T}\right)\Gamma_T(s)ds, \tag{12}$$

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

$$\varpi_T^2 \to_p \varpi^2 \tag{13}$$

as $T \to \infty$, where ϖ^2 is the long-run variance of V, if and only if $B_T \to \infty$ and $B_T/T \to 0$ as $T \to \infty$.

It is easy to see that[14]

$$\gamma_n(j) = \frac{1}{n}\sum_{i=1}^n v_i v_{i-j} = \frac{1}{T}\sum_{i=1}^n \delta V_{i\delta}V_{(i-j)\delta} \approx \frac{1}{T}\int_0^T V_t V_{t-j\delta}dt = \Gamma_T(j\delta), \tag{14}$$

and therefore, we have

$$\delta\pi_n^2 \approx \sum_{|j|\le n} \delta K\left(\frac{j\delta}{b_n\delta}\right)\Gamma_T(j\delta) \approx \int_{|s|\le T} K\left(\frac{s}{B_{n,\delta}}\right)\Gamma_T(s)ds, \tag{15}$$

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 (12), (13) and (15) that, under suitable regularity conditions,

$$\delta\pi_n^2 \to_p \varpi^2, \tag{16}$$

[12] Of course, (u_i) is not observable and should be estimated in practice. However, this is unimportant and it is ignored here.

[13] In the paper, we only consider inferences based on consistent long-run variance estimators. Therefore, in particular, we do not allow the bandwidth parameter to be set as $b_n = bn$ for some constant $b > 0$, which yields the so-called fixed-b asymptotics. See, e.g., Sun (2014) for some recent developments on the subject.

[14] A formal justification of the two approximations in equations (14) and (15) is given in the proof of Theorem 4.1 in Lu & Park (2019, p.263).

if and only if

$$B_{n,\delta} \to \infty \quad \text{and} \quad B_{n,\delta}/T \to 0 \tag{17}$$

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

If a discrete time bandwidth b_n yields the corresponding continuous time bandwidth $B_{n,\delta} = b_n\delta$ satisfying conditions in (17), we say that it is *high-frequency compatible*. Of the two high-frequency compatibility conditions in (17), only the first one $B_{n,\delta} = b_n\delta \to \infty$ matters, since the second one $B_{n,\delta}/T = b_n/n \to 0$ 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 $b_n \to \infty$ faster than $\delta \to 0$. 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 $B_{n,\delta} = b_n\delta = cT^a \to \infty$ as long as $T \to \infty$, 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 $B_{n,\delta} = b_n\delta = c\delta^{1-a}T^a \to 0$ if $\delta \to 0$ fast enough relative to $T \to \infty$. In this case, we have

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

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

$$\delta\pi_n^2 \to_p 0 \tag{18}$$

as $\delta \to 0$ and $T \to \infty$.[15]

Now it is clear that the asymptotics of the G and H-tests rely critically on whether or not the bandwidths used in their long-run variance estimators are high-frequency compatible. We expect them to have well defined limit distributions if they are based on the long-run variances estimated with high-frequency compatible bandwidths, for which (16) holds. However, if the long-run variance estimators rely on the bandwidths that are high-frequency

[15] If we set $b_n = bn$ for some constant $b > 0$ as in the fixed-b asymptotics in discrete time, then it follows that $B_{n,\delta} = b_n\delta = bT$. Indeed, it is not difficult to see that, in our asymptotic setting, using a fixed-smoothing parameter at high frequency regression yields the fixed-b asymptotics in continuous time. Therefore, the fixed-b approach provides self-normalization. We are grateful to a referee for pointing this out.

incompatible and (18) holds, then we have

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

even under the null hypothesis, as $\delta \to 0$ and $T \to \infty$, 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 & 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

$$b_n^* = \left(\frac{r\kappa_r^2 \theta_r^2}{\int K(x)^2 dx} n \right)^{1/(2r+1)}, \tag{19}$$

where r is the characteristic exponent of kernel function K, $\kappa_r = \lim_{x\to 0}(1 - K(x))/|x|^r$ 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 θ_r is unknown and should be estimated. Andrews (1991) assumes that (v_i) is AR(1), in which case we have

$$\theta_1^2 = \frac{4\rho^2}{(1-\rho)^2(1+\rho)^2}, \qquad \theta_2^2 = \frac{4\rho^2}{(1-\rho)^4} \tag{20}$$

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 ρ to obtain an estimate $\tilde{\theta}_r$ of θ_r in (20) for $r = 1, 2$.[16] On the other hand, Newey & West (1994) propose to nonparametrically estimate θ_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 $0 < p < 1$.

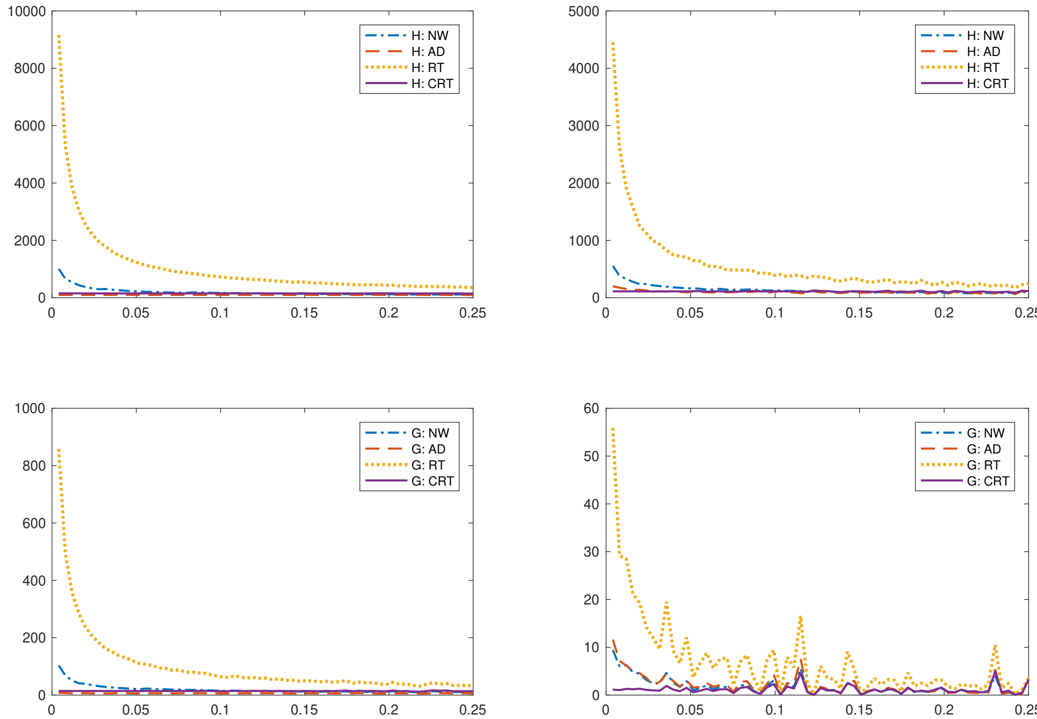
Lemma 5.1. *Let $V = U$ for cointegrating regrssion and $V = XU$ for stationary regression satisfy Assumption E, and assume $\sup_{0\leq t\leq\infty} \mathbb{E}|V|^2 < \infty$. The Andrews bandwidth is high-frequency compatible, while the Newey-West bandwidth is not. For the latter, we have $B_{n,\delta}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

Obviously, the Newey-West bandwidth is not high-frequency compatible because $B_{n,\delta} \to 0$ if $\delta = o\left(T^{-(2pr+1)/(2r(1-p))}\right)$, which holds whenever $\delta \to 0$ fast enough relative to $T \to \infty$. 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

[16] Here we only consider the univariate version of bandwidth in Andrews (1991) to make our discussions simple and focused. However, in our simulation and application, we use its multivariate version with an identity weighting matrix.

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 $\mathrm{H}_0 : \beta_0 = 0$ and $\beta_1 = 1$ in Models I/II (upper panel) and III/IV (lower panel), respectively, plotted against the varying sampling intervals δ ranging from 1/252 (daily frequency) to 1/4 (quarterly frequency).

presented in Figure 4 for various sampling intervals δ 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 schemes for the bandwidth choice: the rule of thumb (RT), the rule of thumb in continuous time (CRT), Andrews (AD) and Newey-West (NW).[17] 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 high-frequency compatible CRT, the G and H-tests yield values that are not sensitive to the sampling frequency. They remain stable more or less across the entire range of the sampling intervals we consider. The G and H-tests with AD, which is

[17] Here and elsewhere in the paper, we use the Parzen kernel. All our results are not sensitive to the choice of kernel function.

also high-frequency compatible in our setup, behave as expected only in Models I-III. In Model IV, they become unreliable at high frequency. This is mainly because the estimated autoregressive coefficient for Model IV does not converge to unity, as demonstrated in Figure 3. In sharp contrast, the values of the G and H-tests become heavily dependent upon the sampling frequency if high-frequency incompatible RT and NW are used. In this case, the overall pattern 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 dramatically as the sampling frequency varies. 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 from the standard Wald test cannot be simply dealt with by using its robust version that allows 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 $V = U$ for cointegrating regrssion and $V = XU$ for stationary regression satisfy Assumptions 2.2(b), 2.3 and 4.1, and K satisfy Assumptions 2.1, 2.2(a) and 4.2 in Lu & Park (2019). Moreover, we assume that b_n used in $\hat{\omega}^2$ and $\hat{\Omega}$ is high-frequency compatible, and $b_n \delta \Delta_{\delta,T}(V) \to 0$ as $\delta \to 0$ and $T \to \infty$.

Assumption F specifies the regularity conditions on the processes U and XU, and the kernel function K used in the long-run variance estimation. See Lu & Park (2019) 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 χ_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

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

as $\delta \to 0$ and $T \to \infty$.
(b) Under Assumptions C2 and D2, we have

$$G(\hat{\beta}) \to_d P^{*\prime} R' \left(R Q^{-1} R' \right)^{-1} R P^*$$

as $\delta \to 0$ and $T \to \infty$, where $P^* = \left(\int_0^1 X_t^\circ X_t^{\circ\prime} dt \right)^{-1} \int_0^1 X_t^\circ dU_t^*$ with standard Brownian motion U^* defined as $U^* = U^\circ / \pi$ and $Q = \int_0^1 X_t^\circ X_t^{\circ\prime} dt$ using the notations in Theorem 4.1.

Both 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 long-run 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° and U° are independent, then its limit null distribution reduces to chi-square distribution. When there is no evidence to show independence of X° and U° in practice, one may use resampling methods such as bootstrap or subsampling to obtain the critical values of the nonstandard null distribution of $G(\hat{\beta})$. A similar issue arises when the residual-based cointegration test is conducted with high-frequency data, as discussed in Jiang et al. (2020, Section, 4.2), and they propose to use subsampling approach to conduct inference.

6 An Empirical Illustration

In this section, we investigate the extent to which the interest rates of securities with time varying maturities move together using the tests introduced in the previous section. To eliminate default probabilities as a confounding factor, we consider U.S. treasury bonds and study specifically interest rate co-movement along the U.S. treasury yield curve. Our question can then be posed alternatively as to whether or not the slope of the yield curve remains constant in response to an economic shock, for example an unexpected change in monetary policy. We will say if interest rates along the yield curve move together that the data supports the idea of a "parallel shift" in the yield curve. A parallel shift occurs when, for example, a twenty-five basis point increase in the federal funds rate is associated with a twenty-five basis point increase in the ten-year yield. We are agnostic as to whether a parallel yield curve shift is associated with the state of the economy. This parallel shift may therefore happen in normal times, characterized with a typical upward sloping yield curve and also in bad times, when a flat or inverted yield curve is more common.

The consistent co-movement of long and short rates, should we establish this as a stylized fact, would have important monetary policy implications. Such a finding would imply that the U.S. Federal Reserve System (FED) could effectively control long rates simply by manipulating the short-term policy rate, the federal funds rate (FFR), through traditional open market operations. The targeting of rates had indeed been one of the conventional monetary policy tools employed by the FED over the past four decades to execute its dual mandate.[18] The FED's efforts to counter the negative shocks associated with the

[18] The other monetary policy tool used is the targeting of monetary aggregates such as M2. The FED Chairman Alan Greenspan noted in July 1993 testimony to Congress that this policy tool was no longer

financial market meltdown in 2008-9 were challenged by the fact that nominal rates had fallen quite dramatically in the decade preceding the crisis, due in large part to persistently low inflation expectations (Bernanke (2020)). In an environment with the policy rate at or near the zero lower bound (ZLB), the FED was compelled to consider implementing a set of non-conventional policy tools, including most notably quantitative easing (QE), forward guidance, large scale asset purchases (LSAP), repurchase agreement (REPO), reverse REPO, negative interest rates and yield curve control. These non-conventional tools would allow for the continued provision of an accommodative monetary policy despite the constraints imposed by the ZLB.

The FED actually implemented many of these non-conventional monetary policy tools during and after the 2008-9 financial crisis. With these non-conventional monetary policies, the FED hoped to directly affect long rates in spite of its inability to further lower nominal short term rates - as the latter had been either at or near the zero lower bound since the crisis.[19] Long rates, of course, may change independently of the FED due to portfolio reallocation strategies by investors that reflect state specific risk and expected return forecasts by asset class over the course of the business cycle. Nevertheless, the unconventional policies, QE and forward guidance in particular, turned out to be effective for the FED to achieve its dual mandate of achieving maximum employment and stable inflation not only during the crisis but also for many years thereafter. Naturally the uncertainties about the effectiveness of these non-conventional policies have been diminished, and these policy tools seem no longer considered as non-conventional. Rather, they are now considered to be effective policy tools that provide policy makers with even more operating flexibility to support the economy and financial markets in times of great distress (Bernanke (2020)).

Now we test empirically if the FED has indeed controlled long rates by manipulating the short term policy rate until the 2008-9 financial crisis. Specifically, we consider the simple continuous time regression of long-term interest rate Y_t on short-term interest rate X_t

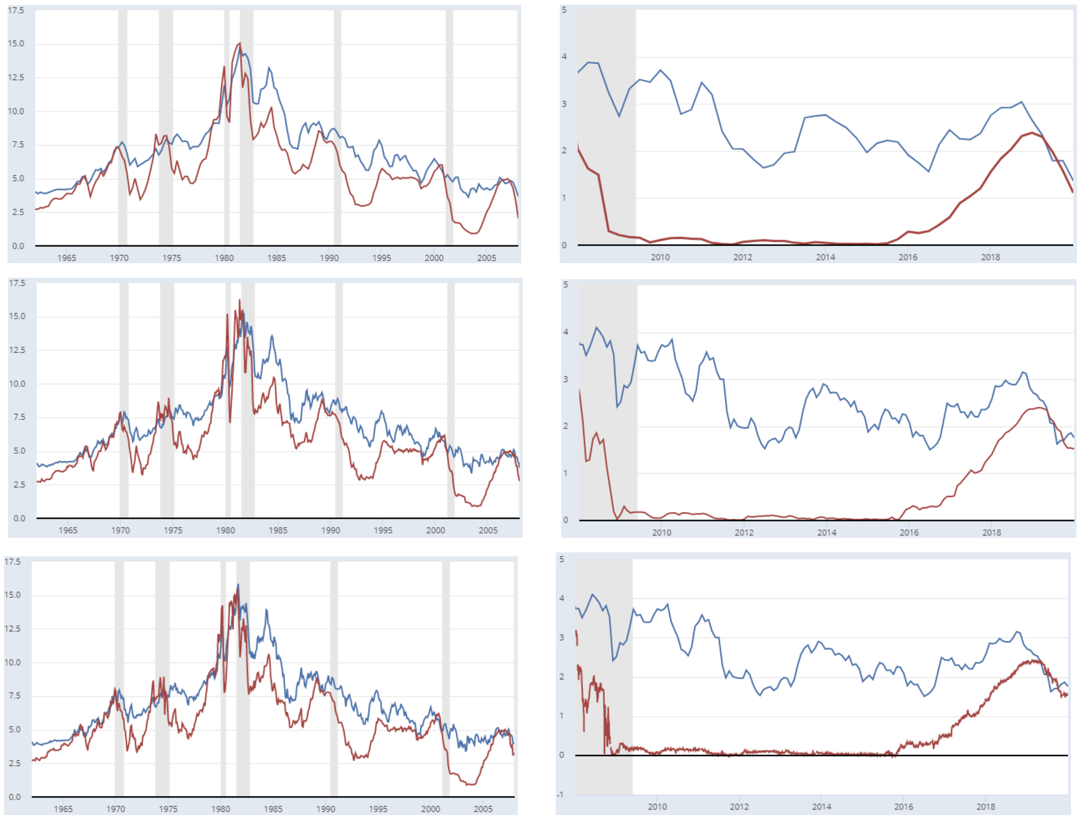
$$Y_t = \alpha + \beta X_t + U_t \tag{21}$$

using observations $(Y_{i\delta})$ and $(X_{i\delta})$, $i = 1, \ldots, n$, collected at discrete time intervals of length δ, corresponding to quarterly, monthly and daily frequencies, where α represents the historic average of term premium, β the parameter revealing the existence of the aforementioned parallel shift of yield curve, and U_t the regression error. When the intercept α is zero, there is no term premium. When the parameter β is unity, the long and short rates increase or

useful.

[19] Interest rates hit the zero lower bound in September 2008. The three-month secondary market rate was 0.03% on September 17, 2008, and the FED made its first purchase of financial assets in November 2008.

Figure 5: Short and Long Rates Before/After 2008-9 Financial Crisis at Quarterly, Monthly and Daily Frequencies



Notes: The panels on the left present short and long rates before the crisis, 1962-2007, and those on the right plot the same series after the crisis, 2008-2019, and the top, middle and bottom panels present short and long rates at quarterly, monthly and daily frequencies, respectively.

decrease in the same direction and by the same amount (in basis points), implying that the parallel shift phenomena to the yield curve described above is a valid characterization. As long as $\beta = 1$, we may still perfectly predict the movement of the long rate by that of the short rate conditional on the term premium. Therefore non-rejection of $\beta = 1$ may provide an empirical justification for FED's efforts to influence long rates through its usual open market operations.

We use the interest rate on the ten-year U.S. Treasury bond for the long-term interest rate ($Y_{i\delta}$), and the three-month U.S. Treasury bill secondary market rate for the short-term rate ($X_{i\delta}$). Both data series are obtained at quarterly, monthly and daily frequencies from the Federal Reserve Economic Data (FRED) of St. Louis FED. The two interest rates at quarterly, monthly and daily frequencies are plotted in Figure 5 where the panels on the left and right are respectively for the sample periods before and after the 2008-9 financial crisis, 1962-2007 and 2008-2019. The shaded areas indicate the NBER recessions. We start from 1962 when the 10 year rate became available at daily frequency.

We test if $\alpha = 0$ and $\beta = 1$ both separately and jointly under the null hypotheses $H_0^{\alpha} : \alpha = 0$, $H_0^{\beta} : \beta = 1$, and $H_0^{\alpha,\beta} : (\alpha, \beta) = (0, 1)$ for the pre- and post-crisis sample periods using the four robust H-tests, AD, CRT, NW and RT, as well as the non-robust Wald test introduced in the previous section. All tests are valid asymptotically for low frequency data. The non-robust Wald test is valid only for the regressions with observations collected at low frequency. For the regressions with observations collected at high frequency, however, only two robust H-tests, AD and CRT, are valid according to our theoretical results and simulation evidence. The other two robust tests, NW and RT, are valid only in low frequency regressions. For easy reference, we call the AD and CRT as high-frequency-valid tests and NW and RT as high-frequency-invalid tests.

To empirically demonstrate this point, we test our hypotheses using both low frequency (quarterly) and high frequency (daily) data. Monthly data are also considered for comparison. In fact, they can be considered as either low or high frequency depending on the sample size. Recall that our asymptotics are applicable only when $\delta \to 0$ sufficiently fast relative to $T \to \infty$. High-frequency-invalid tests may behave well at monthly frequency if their divergence rates are not too fast. As we show in our simulations, the high-frequency-invalid NW test diverges more slowly as the sampling interval decreases than the other high-frequency-invalid RT test, and thus it may perform better than RT at monthly frequency.

The test results from the quarterly, monthly and daily regressions are presented in Tables 1, 2 and 3, respectively. The results from the pre- and post-crisis samples are presented in the upper and lower panels in each table. The test values along with the 1% and 5% critical values for the five tests considered are presented. For the robust tests on H_0^{β} for pre-crisis samples, the p-values are also provided. The test values and the associated p-values are marked with one star (two stars) when the null hypothesis is rejected at 5% (1%) significance level. Non-rejection of a null hypothesis at 5% significance level provide stronger support for the null than the non-rejection at 1% level.[20] In the following discussions, we regard only the rejection at 1% level as being significant, since we believe that the actual rejection probabilities for our tests are larger than their nominal sizes, as shown in our simulations. This is due to the persistency of the covariate in our regression. It is widely known that the actual rejection probabilities of the chi-square tests in regressions with persistent covariates are generally larger, often considerably, than their nominal sizes.

Most interestingly, we find strong support for the hypothesis $\beta = 1$ in the pre-crisis

[20]Though we do not report to save space, the results obtained using the data from the entire sample period 1962-2019 are qualitatively similar to those obtained using the post-crisis data.

Table 1: Testing Results from Quarterly Regression

	Pre-2007 sample (1962–2007)				
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	102.24**	18.82**	26.83**	21.96**	55.47**
H_0^{β}	11.82**	1.89	2.55	2.12	5.22*
		[0.1692]	[0.1103]	[0.1454]	[0.0223*]
$H_0^{\alpha,\beta}$	257.42**	40.12**	66.30**	51.59**	157.14**
	Post-2007 sample (2008–2019)				
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	259.81**	81.72**	100.38**	90.30**	180.94**
H_0^{β}	27.20**	29.92**	15.97**	16.12**	26.48**
$H_0^{\alpha,\beta}$	287.50**	87.99**	108.60**	95.10**	203.04**

Notes: Presented are test statistics (and p-values in the square brackets) of the quarterly regression. The critical values of the 5% and 1% level χ^2 tests for the single hypotheses H_0^{α} and H_0^{β} are, respectively, 3.84 and 6.63. The critical values of the 5% and 1% level χ^2 test for the joint hypothesis $H_0^{\alpha,\beta}$ are, respectively, 5.99 and 9.21. Rejection at the 5% level is signified by $*$ and at the 1% level signified by $**$.

Table 2: Testing Results from Monthly Regression

	Pre-2007 sample (1962–2007)				
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	320.90**	19.20**	30.03**	36.44**	131.99**
H_0^{β}	44.10**	2.38	3.41	4.09*	14.37**
		[0.1229]	[0.0648]	[0.0431*]	[0.0002**]
$H_0^{\alpha,\beta}$	752.60**	37.61**	67.26**	83.58**	361.49**
	Post-2007 sample (2008–2019)				
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	960.40**	114.70**	147.93**	195.56**	575.63**
H_0^{β}	100.20**	33.42**	18.83**	21.70**	65.29**
$H_0^{\alpha,\beta}$	1077.50**	134.91**	153.16**	207.05**	634.23**

Notes: Presented are test statistics (and p-values in the square brackets) of the monthly regression. The critical values of the 5% and 1% level χ^2 tests for the single hypotheses H_0^{α} and H_0^{β} are, respectively, 3.84 and 6.63. The critical values of the 5% and 1% level χ^2 test for the joint hypothesis $H_0^{\alpha,\beta}$ are, respectively, 5.99 and 9.21. Rejection at the 5% level is signified by $*$ and at the 1% level signified by $**$.

Table 3: Testing Results from Daily Regression

Pre-2007 sample (1962–2007)					
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	6608**	55.97**	31.04**	175.90**	1357.90**
H_0^{β}	832**	5.35*	3.10	15.80**	124.60**
		[0.0207*]	[0.0783]	[0.0001**]	[0.0000**]
$H_0^{\alpha,\beta}$	14814**	133.64**	68.76**	480.00**	4086.20**
Post-2007 sample (2008–2019)					
	Nonrobust test	Robust H-test			
		AD	CRT	NW	RT
H_0^{α}	16893**	222.82**	141.24**	1062.80**	6402.30**
H_0^{β}	1619**	23.12**	17.31**	102.90**	686.10**
$H_0^{\alpha,\beta}$	18960**	235.20**	146.82**	1140.40**	7051.40**

Notes: Presented are test statistics (and p-values in the square brackets) of the daily regression. The critical values of the 5% and 1% level χ^2 tests for the single hypotheses H_0^{α} and H_0^{β} are, respectively, 3.84 and 6.63. The critical values of the 5% and 1% level χ^2 test for the joint hypothesis $H_0^{\alpha,\beta}$ are, respectively, 5.99 and 9.21. Rejection at the 5% level is signified by $*$ and at the 1% level signified by $**$.

sample at all three data frequencies when the high-frequency-valid tests are used. In the post-crisis sample, however, it is clearly rejected by all tests at all data frequencies. This is as expected and consistent with the changes in the FED's policy behaviors after the 2008-9 financial crisis discussed above. Our results therefore provide empirical validity to using the short rate to control long rates in the conventional monetary policy regime that prevailed until the financial crisis.

We also find that term premium does exist in both pre- and post-crisis samples at all data frequencies, indicating that the U.S. Treasury yield curve has been on the average upward sloping. The no term premium hypothesis $\alpha = 0$ is clearly rejected by all five tests regardless of whether it is tested separately under $H_0^{\alpha} : \alpha = 0$ or jointly under $H_0^{\alpha,\beta} : (\alpha, \beta) = (0, 1)$ during both pre- and post-crisis sample periods in the regressions with observations collected at all three data frequencies.

Our empirical findings from the pre-crisis sample period corroborate nicely with our theoretical results and simulation evidence presented in the previous sections on the five tests we consider. As expected, however, the data in the post-crisis sample strongly disagree with the hypotheses $\alpha = 0$ and $\beta = 1$ whether they are tested separately or jointly, and as a result all five tests reject the null hypotheses clearly at all data frequencies. The information from the pre-crisis sample are relatively weaker and consequently the five tests behave differently when applied to regressions with observations collected at different frequencies. We will therefore use the results from the pre-crisis sample to evaluate and compare the performance of the five tests we consider.

In the pre-crisis sample, we find that the high-frequency-valid AD and CRT tests provide consistent results for all three data frequencies we consider as our theory predicts. Moreover, they provide more significant results with higher p-values than the high-frequency-invalid NW and RT tests. As expected, the NW and RT tests perform well only for low frequency, quarterly regression, and their performance deteriorates as we increase data frequency to monthly. At the monthly regressions, the high-frequency-invalid test RT no longer support the null, but the NW test performs better than RT for the reasons we discussed above, and continues to support the null. Only the high-frequency-valid AD and CRT tests support the null in the high frequency regression as expected. Our findings clearly demonstrate the importance of using a valid test so as to draw correct inference from our sample set. Moreover, it is particularly important to be able to perform the correct analyses with high frequency data since doing so with lower frequency data (when data are available at high frequency) implies huge information loss.

7 Simulation

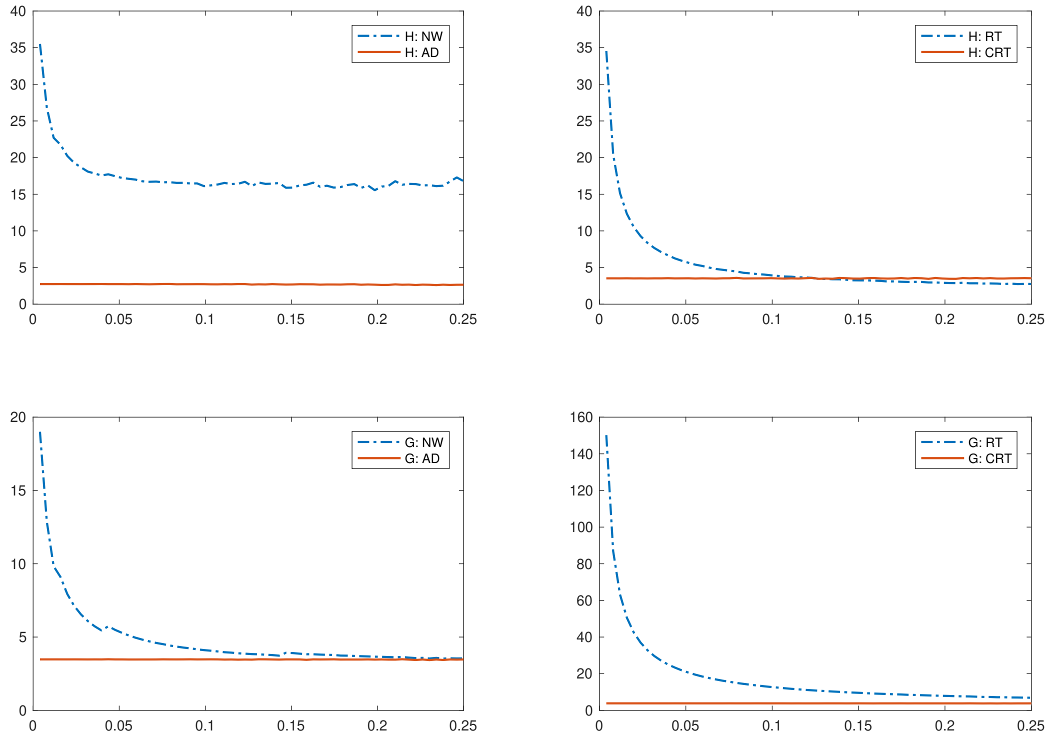
Our simulation shows that the asymptotics developed in the paper are relevant and useful to analyze regressions using high-frequency observations. In our simulation, we consider two types of regressions, the stationary and cointegration regressions, based on the continuous time regression (21) for $0 \leq t \leq T$, where $X = (X_t)$ and $U = (U_t)$ are given by

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

where $W = (W_t)$ 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 and $\hat{U}$ 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 the estimates from X and $\hat{U}$ 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

Figure 6: Simulated Means of Robust Wald Tests



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

a stationary OU process. We consider the test of the null hypothesis $\mathrm{H}_0 : \alpha = 0$ and $\beta = 1$ 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. The long-run variance estimates in the test statistics are obtained using Parzen kernel and the four different bandwidth choices, which are introduced earlier in Section 5 and referred to as RT, CRT, AD and NW, respectively. As discussed, CRT and AD are high-frequency compatible, while RT and NW are not.

Figure 6 presents the means of the H-test for the stationary regression and the G-test for the cointegrating regression computed under the null hypothesis, 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.[21] The simulated means of the G and H-tests behave quite similarly as we change the sampling intervals. Their behaviors are critically dependent upon the high frequency compatibility of the bandwidth used in estimating the required long-run 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 constructed 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.

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 the 5% chi-square critical value for $T = 30$ and 50. They are presented in Figure 7 in Appendix B. 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 below one month.[22] In contrast, the rejection probabilities of the tests with high-frequency compatible bandwidths are remarkably stable across all sampling frequencies, and they tend closer to the nominal size of the tests as T increases.

To see more clearly the effect of our double asymptotics relying on $\delta \to 0$ and $T \to \infty$ jointly, we set $\delta = (1/3)T^{-2}$ so that T changes simultaneously along with δ. Under this setup, we simulate the means of the H and G-tests for the stationary and cointegrating regressions under the null hypothesis, which is reported in Figure 8 in Appendix B. The overall pattern of the frequency dependency remains the same as what we observe in Figure 6, which is obtained by changing only δ with T fixed. The only notable difference is that the simulated means of the tests with high-frequency compatible bandwidths, CRT and AD, now show downward trends as the sampling interval gets smaller. As discussed, no such trends appear when T is fixed. Therefore, we may conclude that the trends here are generated not directly by varying δ but by T changing with δ. The downward trends emerging under our double asymptotics just indicate that the tests have positive biases in our simulation model when T is small. The simulated means of the tests with high-frequency incompatible bandwidths RT and NW behave similarly, regardless of T being fixed or varying with δ.

[21] Our simulation with 50 years data shows exactly the same patterns as those in Figure 6.

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

We conduct some additional simulations based on the models that can capture more realistic features of the economic and financial series used in our empirical illustrations presented in the previous section. Details of the additional simulation models are described and the results obtained from these models are reported in Appendix C.

8 Conclusion

The Wald test is widely used 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 is manifested vividly in many time series regressions. In particular, when observations are used at high frequency, the standard Wald test almost explodes and we are always led to reject the null hypothesis no matter whether it is true or not. The problem, however, appears to have long been either overlooked or neglected in the literature. Certainly, it has now become increasingly more important as more economic and financial time series are collected and made available at high frequencies.

In the paper, we develop a new continuous time framework and develop relevant asymptotics, relying on sampling interval δ and time span T jointly, to analyze various versions of the Wald test in regressions with high frequency observations over long sample span. Our framework accommodates both stationary and cointegrating regressions, and our asymptotics hold under mild regularity conditions. According to our asymptotic theory, the standard Wald test is expected to diverge if $\delta \to 0$ fast enough relative to $T \to \infty$. This is exactly what we observe in practice. The robust Wald tests, which use the long-run variance estimates in place of the usual variance estimates, are expected to behave similarly unless the bandwidths for their long-run variance estimates are appropriately chosen to be high-frequency compatible. The high frequency compatibility of a bandwidth ensures that the resulting long-run variance estimate properly captures the clear and present serial correlation at high frequency. Only when high-frequency compatible bandwidths are used do the robust Wald tests become valid and behave as expected under the null hypothesis.

References

Aït-Sahalia, Y., & Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, *83*(2), 413–452.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*, 817–858.

Bernanke, B. S. (2020). The new tools of monetary policy. *American Economic Review*, *110*, 943–983.

Bhattacharya, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *60*(2), 185–201.

Chang, Y., Choi, Y., Kim, H., & Park, J. Y. (2016). Evaluating factor pricing models using high frequency panels. *Quantitative Economics*, *7*, 889-933.

Choi, Y., Jacewitz, S., & Park, J. Y. (2016). A reexamination of stock return predictability. *Journal of Econometrics*, *192*, 168-189.

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*, 111-120.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343.

Jeong, M., & Park, J. Y. (2011). *Asymptotic theory of maximum likelihood estimator for diffusion model.* (Working Paper, Indiana University)

Jeong, M., & Park, J. Y. (2014). *Asymptotic theory of maximum likelihood estimator for jump diffusion model.* (Working Paper, Indiana University)

Jiang, B., Lu, Y., & Park, J. Y. (2020). Testing for stationoarity at high frequency. *Journal of Econometrics*, *215*, 341-374.

Kanaya, S., Kim, J., & Park, J. Y. (2018). *On the modulus of continuity of brownian motion.* (Working Paper, Indiana University)

Kim, J., & Park, J. Y. (2017). Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics*, *196*, 37-54.

Lu, Y., & Park, J. Y. (2019). Estimation of longrun variance of continuous time stochastic process using discrete sample. *Journal of Econometrics*, *210*, 236-267.

Müller, U. K. (2005). Size and power of tests of stationarity in highly autocorrelated time series. *Journal of Econometrics*, *128*, 195-213.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, *61*, 631-653.

Perron, P. (1991a). A continuous-time approximation to the stationary first-order autoregressive model. *Econometric Theory*, *7*, 236-252.

Perron, P. (1991b). A continuous-time approximation to the unstable first-order autoregressive model: The case without an intercept. *Econometrica*, *59*, 211-236.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, *33*, 311-340.

Rozanov, Y. A. (1960). A central limit theorem for additive random functions. *Theory of Probability & Its Applications*, *5*(2), 221–223.

Sun, Y. (2014). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, *178*, 659-677.

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

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

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 $i = 1, \ldots, n$ 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

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

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

$$\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i, \tag{22}$$

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

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

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 analogous. We write

$$\sqrt{T}\Lambda_T(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime}\right)^{-1} \frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} \Lambda_T^{-1} x_i u_i, \tag{23}$$

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

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \Lambda_T^{-1} x_i x_i' \Lambda_T^{-1\prime} &= \frac{1}{T}\int_0^T \Lambda_T^{-1} X_t X_t' \Lambda^{-1\prime} dt + o_p(1) \\
&= \int_0^1 X_t^T X_t^{T\prime} dt + o_p(1) \to_d \int_0^1 X_t^\circ X_t^{\circ\prime} dt := Q
\end{aligned} \tag{24}$$

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. Therefore, it follows that

$$\sqrt{T}\Lambda_T(\hat{\beta} - \beta) \to_d \left(\int_0^1 X_t^\circ X_t^{\circ\prime} dt\right)^{-1} \int_0^1 X_t^\circ dU_t^\circ := P. \tag{25}$$

To derive the asymptotic null distribution of $F(\hat{\beta})$, we first note that the middle part of $F(\hat{\beta})$ satisfies

$$nR\left(\sum_{i=1}^{n} x_ix_i'\right)^{-1} R' \sim_p (R\Lambda_T^{-1})Q^{-1}(R\Lambda_T^{-1})', \tag{26}$$

by (24). Moreover, based on (25), the term in $F(\hat{\beta})$ on the sides satisfies

$$\sqrt{T}(R\hat{\beta} - r) = \sqrt{T}R(\hat{\beta} - \beta) = R\Lambda_T^{-1}\sqrt{T}\Lambda_T(\hat{\beta} - \beta) \sim_p R\Lambda_T^{-1}P. \tag{27}$$

Lastly, the denominator of $F(\hat{\beta})$ satisfies

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \frac{1}{T}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n} u_ix_i'\Lambda_T^{-1\prime}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda_T^{-1}x_ix_i'\Lambda_T^{1\prime}\right)^{-1}\left(\frac{\sqrt{\delta}}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_T^{-1}x_iu_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 + O_p(T^{-1}) \to_p \sigma^2. \tag{28}$$

Note that $R\Lambda_T^{-1}$ in (26) and (27) is a $q\times p$ matrix where q is the number of restrictions and p is the number of regressors including the constant term. We denote $\Lambda_T = \text{diag}(\lambda_{1T}, \ldots, \lambda_{pT})$, and assume without loss of generality that $\lambda_{1T} \preceq \cdots \preceq \lambda_{pT}$. Since R is assumed to be of full row rank and in reduced row canonical form (see Footnote 5), we can find $\iota_j \in \{1, \ldots, p\}$ be the column index of the leading 1 in the jth row of R, for $j = 1, \ldots, q$. Then, we define a $q \times q$ diagonal matrix

$$\Lambda_T^* = \text{diag}(\lambda_{\iota_1 T}, \ldots, \lambda_{\iota_q T}).$$

It then follows that, as $T \to \infty$,

$$\Lambda_T^* R\Lambda_T^{-1} \to R^*, \tag{29}$$

where R^* is a $q \times p$ matrix with full row rank and (j,k) entry given by

$$R^*_{jk} = \begin{cases} R_{jk}, & k = 0, \dots, \iota_j \\ R_{jk} \lim_{T\to\infty} \lambda_{\iota_j T}/\lambda_{kT}, & k = \iota_{j+1}, \dots, p \end{cases}, \quad j = 1, \dots, q.$$

Note that $\lim_{T\to\infty} \lambda_{\iota_j T}/\lambda_{kT}$ takes value either 0 or 1 depending on the relative level of nonstationarity of the ι_jth and the kth regressors.

Then it follows from (26), (27) and (29) that

$$n\Lambda^*_T R \left(\sum_{i=1}^n x_i x_i' \right)^{-1} R' \Lambda^*_T \to_d R^* Q^{-1} R^{*\prime}, \tag{30}$$

$$\sqrt{T}\Lambda^*_T (R\hat{\beta} - r) \to_d R^* P, \tag{31}$$

from which and (28) we can deduce that

$$\begin{aligned} \delta F(\hat{\beta}) &= \sqrt{T}(R\hat{\beta} - r)'\Lambda^*_T \left[n\Lambda^*_T R \left(\sum_{i=1}^n x_i x_i' \right)^{-1} R' \Lambda^*_T \right]^{-1} \sqrt{T}\Lambda^*_T (R\hat{\beta} - r)/\hat{\sigma}^2 \\ &\to_d P^{*\prime} R^{*\prime} \left(R^* Q^{-1} R^{*\prime} \right)^{-1} R^* P^*/\sigma^2, \end{aligned}$$

as to be shown. □

Proof of Lemma 4.2 Write

$$\tilde{\rho} - 1 = \frac{\sum_{i=1}^n u_{i-1}(u_i - u_{i-1})}{\sum_{i=1}^n u_{i-1}^2}. \tag{32}$$

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

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

which will be further analyzed subsequently.

We have

$$(U_{i\delta} - U_{(i-1)\delta})^2 = 2\int_{(i-1)\delta}^{i\delta} (U_{t-} - U_{(i-1)\delta}) dU_t^c + \Big([U^c]_{i\delta} - [U^c]_{(i-1)\delta}\Big) + \sum_{(i-1)\delta < t \le i\delta} \Delta(U_t - U_{(i-1)\delta})^2, \tag{33}$$

and

$$\begin{aligned}\Delta(U_t - U_{(i-1)\delta})^2 &= (U_t - U_{(i-1)\delta})^2 - (U_{t-} - U_{(i-1)\delta})^2 \\ &= (U_t - U_{t-})(U_t + U_{t-} - 2U_{(i-1)\delta}) \\ &= (U_t - U_{t-})\Big[(U_t - U_{t-}) + 2(U_{t-} - U_{(i-1)\delta})\Big] \\ &= 2(U_{t-} - U_{(i-1)\delta})\Delta U_t + (\Delta U_t)^2 \end{aligned} \tag{34}$$

for $i = 1, \dots, n$. Therefore, it follows from (33) and (34) that

$$\sum_{i=1}^n (U_{i\delta-} - U_{(i-1)\delta})^2 = [U]_T + 2Z_T, \tag{35}$$

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

$$U_{t-} - U_{(i-1)\delta} = (U_t^c - U_{(i-1)\delta}^c) + \sum_{(i-1)\delta < s < t} \Delta U_s, \tag{36}$$

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

To consider Z^c, we write

$$Z^c = Z^a + Z^b, \tag{37}$$

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

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

and

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

from which it follows that

$$
Z_T^a = O_p(p_T T \Delta_{\delta,T}(U)) + O_p(p_T T \delta) = O_p(p_T T \Delta_{\delta,T}(U)), \tag{38}
$$

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

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

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

$$Z_T^b = O_p\left(\sqrt{q_T T}\Delta_{\delta,T}(U)\right)+O_p\left(\sqrt{q_T T\delta}\right)=O_p\left(\sqrt{q_T T}\Delta_{\delta,T}(U)\right), \tag{39}$$

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

$$\begin{aligned}
Z_T^c &= O_p\left(p_T T\Delta_{\delta,T}(U)\right)+O_p\left(\sqrt{q_T T}\Delta_{\delta,T}(U)\right)\\
&= T\left[O_p\left(p_T\Delta_{\delta,T}(U)\right)+O_p\left(\sqrt{q_T/T}\Delta_{\delta,T}(U)\right)\right]
\end{aligned}\tag{40}$$

from (38) and (39).

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

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} (U_t^c - U_{(i-1)\delta}^c) \Delta U_t \right]^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \mathbb{E}(\Delta U_t)^2 \\
&\le \left[\max_{1 \le i \le n} \sup_{(i-1)\delta < t \le i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \right] \sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} \mathbb{E}(\Delta U_t)^2 \\
&= \left[\max_{1 \le i \le n} \sup_{(i-1)\delta < t \le i\delta} \mathbb{E}(U_t^c - U_{(i-1)\delta}^c)^2 \right] \sum_{0 < t \le T} \mathbb{E}(\Delta U_t)^2 = O\left(T \Delta_{\delta,T}^2(U) \right). \qquad (41)
\end{aligned}
$$

Moreover, we may easily deduce that

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} \left(\sum_{(i-1)\delta < s < t} \Delta U_s \right) \Delta U_t \right]^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} \mathbb{E}\left(\sum_{(i-1)\delta < s < t} \Delta U_s \right)^2 \mathbb{E}(\Delta U_t)^2 \\
&= \sum_{i=1}^n \sum_{(i-1)\delta < t \le i\delta} \left[\sum_{(i-1)\delta < s < t} \mathbb{E}(\Delta U_s)^2 \right] \mathbb{E}(\Delta U_t)^2 \\
&\le \sum_{i=1}^n \sum_{(i-1)\delta < s,t \le i\delta} \mathbb{E}(\Delta U_s)^2 \mathbb{E}(\Delta U_t)^2 \\
&\le \sum_{i=1}^n \sum_{(i-1)\delta < s,t \le i\delta} \mathbb{E}(\Delta U_t)^4 = \sum_{0 < t \le T} \mathbb{E}(\Delta U_t)^4 = O(T). \qquad (42)
\end{aligned}
$$

Therefore, it follows from (41) and (42) that

$$
Z_T^d = O_p\left(\sqrt{T} \Delta_{\delta,T}(U) \right) + O_p(\sqrt{T}) = O_p(\sqrt{T}) = o_p(T) \qquad (43)
$$

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

Now we have, due in particular to (40) and (43),

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

as $\delta \to 0$ and $T \to \infty$. Consequently, the stated result follows immediately from (32), and the proof is complete. $\square$

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

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

and in particular $\tilde{\theta}_r^2 = O_p(\delta^{-2r})$ for $r = 1, 2$ as $\delta \to 0$ and $T \to \infty$ satisfying Assumption E. With these choices of θ_r^2, $r = 1, 2$, the feasible optimal bandwidth $\tilde{b}_n^*$ in (19) becomes

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

and therefore,

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

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

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

$$\delta^r\hat{\theta}_r = \frac{\delta^{1+r}\sum_{|j|\le a_n}|j|^r\gamma_n(j)}{\delta\sum_{|j|\le a_n}\gamma_n(j)}, \tag{44}$$

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

$$\delta^{1+r}\sum_{|j|\le a_n}|j|^r\gamma_n(j) \sim_p \int_{|s|\le A_{n,\delta}}|s|^r\Gamma_T(s)ds \tag{45}$$

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

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

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 & Park (2019) with A_T replaced by $A_{n,\delta}$, and the analyses of their stochastic orders are also analogous. In particular, given $\sup_{0\le t\le\infty} \mathbb{E}|V_t|^2 < \infty$, Lemma A.2 in Lu & Park (2019) 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,

$$\int_{|s|\le A_{n,\delta}} |s|^r \Gamma_T(s)ds - \delta^{1+r} \sum_{|j|\le a_n} |j|^r \gamma_n(j) = O(\Delta_{\delta,T} A_{n,\delta}^{1+r} + \delta A_{n,\delta}^r). \tag{46}$$

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

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

as $\delta \to 0$ and $T \to \infty$. This implies $A_{n,\delta}^{-(1+r)} \left[\delta^{1+r} \sum_{|j|\le a_n} |j|^r \gamma_n(j) - \int_{|s|\le A_{n,\delta}} |s|^r \Gamma_T(s)ds\right] = O(\Delta_{\delta,T}(V) + \delta/A_{n,\delta}) = O(\Delta_{\delta,T}(V) + \delta^p T^{-p}) = o_p(1)$ which proves (45). Now we can readily deduce from (44), (45) and (46) that

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

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

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

from which it follows that

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

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

Proof of Theorem 5.2 To show Part (a), we follow the proof in part (a) of Theorem 4.1 and note that $\delta\hat{\Omega} \to_p \Pi$ under Assumption F, due to Theorem 4.1 in Lu & Park (2019).

Then we can show that, under $H_0 : R\beta = r$,

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

where

$$\begin{aligned} R\sqrt{T}(\hat{\beta} - \beta) &\to_d RM^{-1}\mathbb{N}(0, \Pi) =_d \mathbb{N}(0, RM^{-1}\Pi M^{-1}R'), \\ R \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \delta\hat{\Omega} \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} R' &\to_p RM^{-1}\Pi M^{-1}R', \end{aligned}$$

as $\delta \to 0$ and $T \to \infty$ under Assumption A, C1 and D1. Therefore, it follows immediately that $H(\hat{\beta}) \to_p \chi_q^2$ as $\delta \to 0$ and $T \to \infty$.

The proof of Part (b) follows from that of Part (b) of Theorem 4.1, as the only difference between the test statistics $G(\hat{\beta})$ and $F(\hat{\beta})$ is in the denominator. Note that the denominator of $G(\hat{\boldsymbol{\beta}})$ satisfies $\delta\hat{\omega}^2 \to_p \pi^2$ where π^2 is the LRV of U_t defined in Assumption C2. It then follows from the definition of $G(\hat{\beta})$ and (30)–(31) that

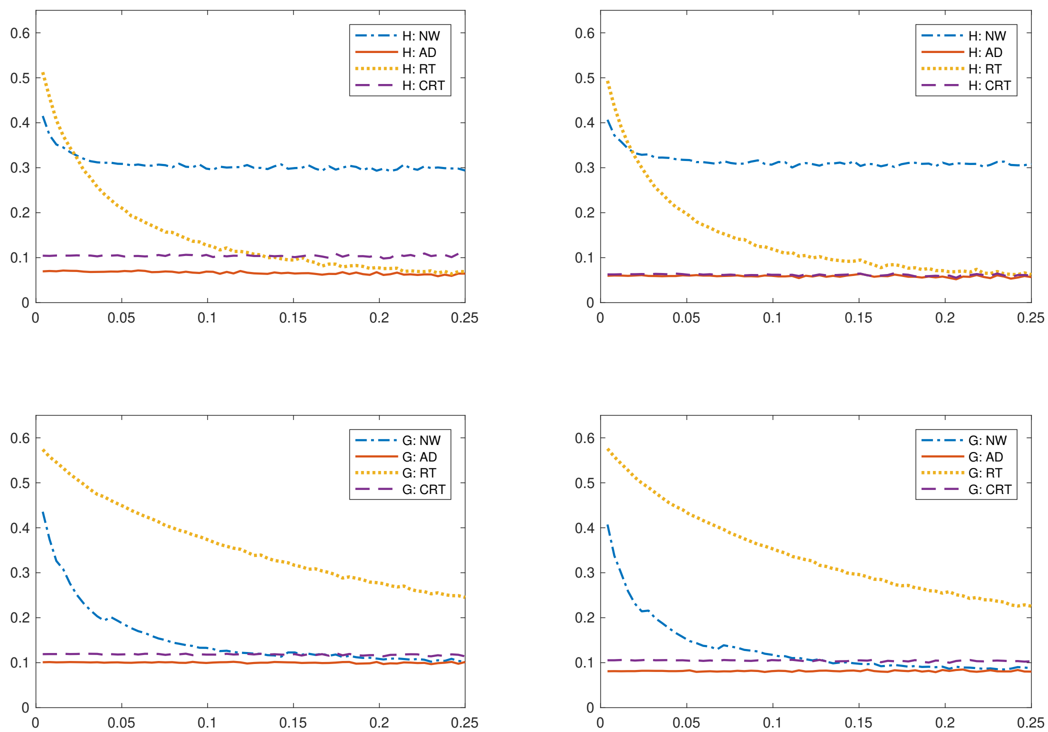
$$\begin{aligned} G(\hat{\beta}) &= \sqrt{T}(R\hat{\beta} - r)'\Lambda_T^* \left[n\Lambda_T^* R \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} R'\Lambda_T^* \right]^{-1} \sqrt{T}\Lambda_T^*(R\hat{\beta} - r)/(\delta\hat{\omega}^2) \\ &\to_d P^{*'} R^{*'} \left(R^* Q^{-1} R^{*'} \right)^{-1} R^* P^* \end{aligned}$$

where $P^* = \left(\int_0^1 X_t^\circ X_t^{\circ\prime} dt \right)^{-1} \int_0^1 X_t^\circ dU_t^*$ where $U^* = U^\circ/\pi$. This complete the proof of the theorem. $\square$

B Additional Figures

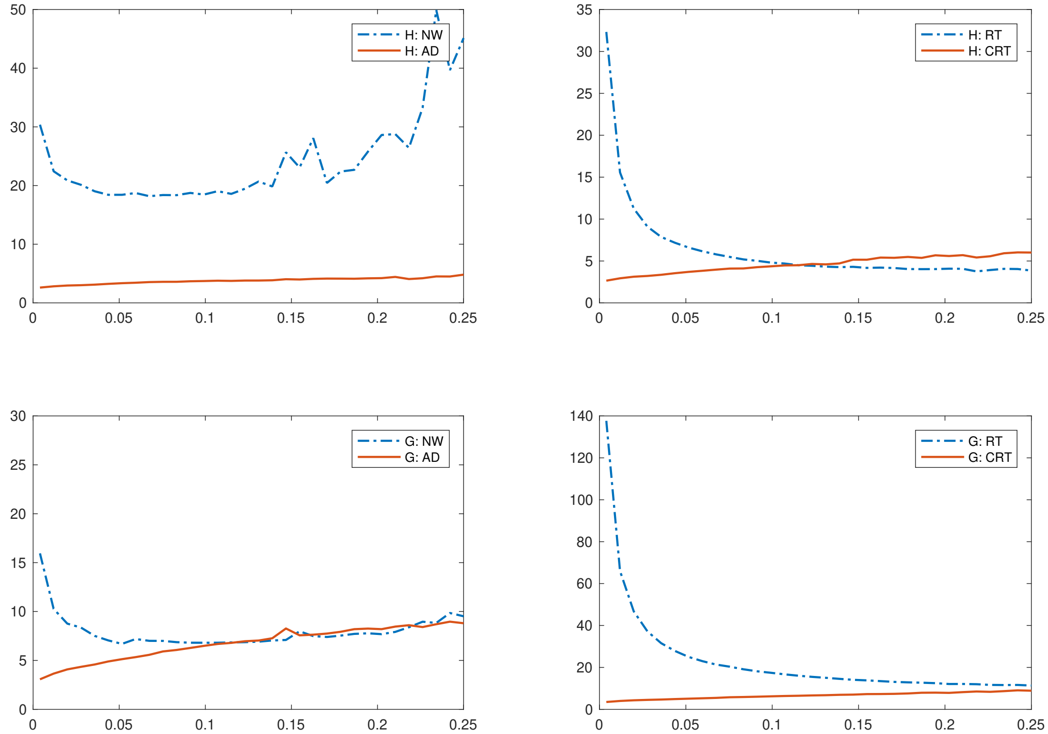
In this section, we present some additional figures.

Figure 7: Simulated Rejection Probabilities



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

Figure 8: Simulated Means of Robust Wald Tests Under Joint Asymptotics



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

C Additional Simulation Results

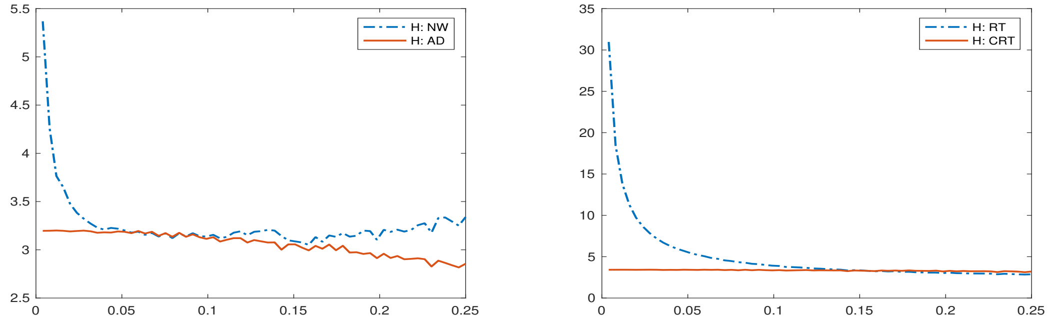
Here we present some additional simulation results based on the models that can capture more realistic features of the economic and financial series used in our empirical illustrations. In particular, we introduce stochastic volatilities in $X = (X_t)$ and $U = (U_t)$, and allow them to interact with each other. Again, we consider two types of regressions, the stationary and cointegration regressions, based on the continuous time regression (21), and test the null hypothesis $H_0 : \alpha = 0$ and $\beta = 1$ using the H and G-tests, respectively. Except for using more realistic models, all other simulation setups are the same as those in Section 7.

For the stationary regression, we let

$$dX_t = \kappa_x(\mu_x - X_t)dt + \sigma_x\sqrt{X_t}dW_{1t} \tag{47}$$
$$dU_t = -\kappa_u U_t dt + \sigma_u dW_{2t}, \tag{48}$$

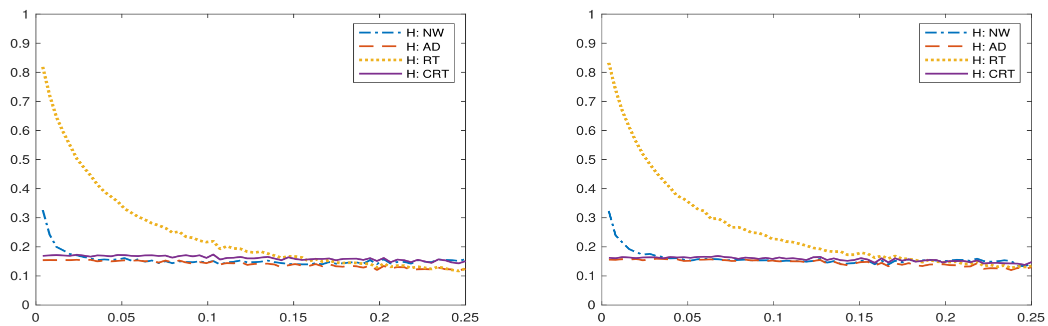
where $W_1 = (W_{1t})$ and $W_2 = (W_{2t})$ are two standard Browian motions with $d[W_1, W_2] = \upsilon dt$, $-1 \leq \upsilon \leq 1$, and set their parameter values to satisfy stationarity conditions so

Figure 9: Simulated Means of Robust Wald Tests of Stationary Regression



Notes: Presented are the simulated means of the H-test for the stationary regression computed under the null hypothesis for $T = 30$. The bandwidth choices are given by RT, CRT, AD and NW, and the sampling interval δ ranges from 1/252 (daily frequency) to 1/4 (quarterly frequency).

Figure 10: Rejection Probabilities of Robust Wald Tests of Stationary Regression



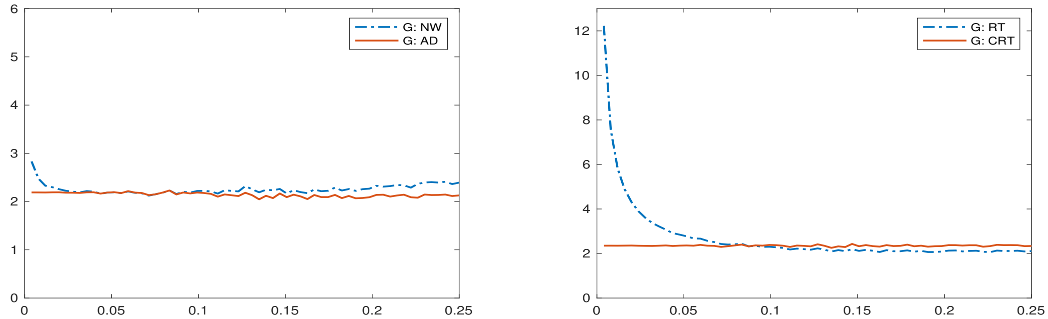
Notes: Presented are the simulated rejection probabilities of the H-test for the stationary regression with $T = 30$ (in the left panel) and $T = 50$ (in the right panel) computed under the null hypothesis using the 5% chi-square critical value. The bandwidth choices are given by RT, CRT, AD and NW,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

that X and U are stationary Feller's square root process and Ornstein-Uhlenckbeck process, respectively.[23] Comparing to our previous simulation model for stationary regression where X has constant volatility and is independent of U, this model is more realistic in the sense that it allows for time varying volatility in X and interaction between X and U. The parameters in the model are set to be $(\kappa_x, \mu_x, \sigma_x) = (0.1060, 5.4949, 0.6000)$, $(\kappa_u, \sigma_u) = (9.4196, 3.1080)$ and $\upsilon = -0.2413$, which are obtained by fitting Model II using daily observations.[24] The parameters $\kappa_x, \mu_x, \sigma_x, \kappa_u$ and σ_u are estimated by the maximum likelihood method, and the estimate for υ is obtained by the sample analogue principle using $\upsilon = [X, U]_T \Big/ \left(\sigma_x \sigma_u \int_0^T X_t dt\right)$ subsequently.

[23]In particular, this requires $2\kappa_x \mu_x \geq \sigma_x^2$, which ensures the process X is bounded above from zero which is suitable for modelling the interest rates in normal times. This condition is imposed for all Feller's square root processes used in our simulations.

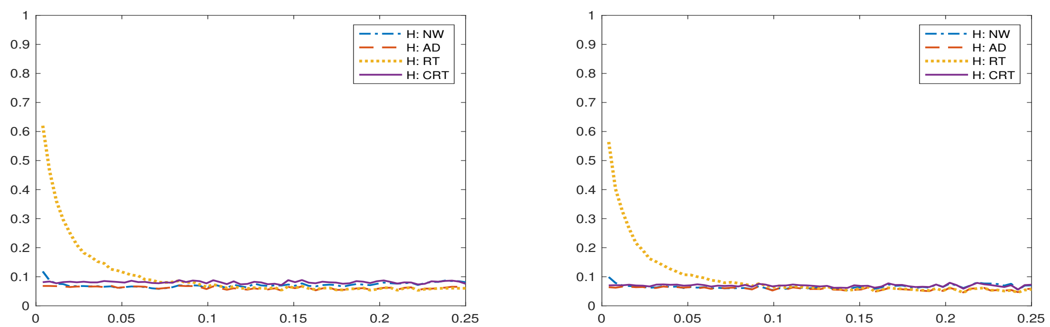
[24]The 3-month T-bill rates hit the zero lower bound on some days after the year 2007. To ensure sensible estimation results, we exclude the post-2007 period of the data to fit the simulation model.

Figure 11: Simulated Means of Robust Wald Tests of Cointegration Regression



Notes: Presented are the simulated means of the G-test for the cointegration regression computed under the null hypothesis for $T = 30$. The bandwidth choices are given by RT, CRT, AD and NW, and the sampling interval δ ranges from 1/252 (daily frequency) to 1/4 (quarterly frequency).

Figure 12: Rejection Probabilities of Robust Wald Tests of Cointegration Regression



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

The simulated means of the test statistics for $T = 30$ are presented in Figure 9 and the simulated rejection probabilities of the 5% test for both $T = 30$ and $T = 50$ are presented in Figure 10, which correspond to the upper panels in Figures 6 and 7, respectively. The simulation results are again in line with our theory: The test statistics employing the NW and RT bandwidths diverge as sampling frequency increases, while the test statistics employing the AD and CRT bandwidths behave much more stably across different sampling frequencies. Our theory is therefore fully demonstrated even when more realistic features are added to the simple baseline models. For more realistic models we consider here, the actual rejection probabilities are all significantly above the nominal 5% level for all bandwidth choices. The overall levels of the actual rejection probabilities appear to be generally higher for realistic models than for simple models, although comparisons across the tests with different bandwidth choices cannot be clearly made in any meaningful manners.

Now we consider the cointegrating regression. To generate X, we use the Heston (1993) model (see also in Aït-Sahalia & Kimmel (2007)), which is often used for modelling the log stock index price process X jointly with the spot variance process V^x as

$$d\begin{bmatrix} X_t \\ V_t^x \end{bmatrix} = \begin{bmatrix} (r-d)+bV_t^x \\ \kappa_x(\mu_x - V_t^x) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\upsilon^2)V_t^x} & \upsilon\sqrt{V_t^x} \\ 0 & \sigma_x\sqrt{V_t^x} \end{bmatrix} d\begin{bmatrix} W_{1t}^x \\ W_{2t}^x \end{bmatrix}, \tag{49}$$

where r is the risk-free interest rate, d is the dividend yield of the stock index, and $W^x = (W_1^x, W_2^x)$ is standard bivariate Brownian motion. For simplicity, we let r and d be constant. Note that the spot variance process V^x is assumed to follow a stationary Feller's square root process. Moreover, $-1 < \upsilon < 0$ represents the leverage effect, i.e., the negative correlation between the shocks to the stock price and its volatility. The parameter values used in our simulations are obtained from X in Model IV fitted to the model in (49) using the method proposed by Aït-Sahalia & Kimmel (2007). In particular, we set $\kappa_x = 3, \mu_x = 0.1, \sigma_x = 0.25, \upsilon = -0.8$ and $b = 6.06$. The values of r and d are set 2% and 1.5%, respectively.

For U, we also use a stochastic volatility model. We specify U jointly with its spot variance process V^u as

$$d\begin{bmatrix} U_t \\ V_t^u \end{bmatrix} = \begin{bmatrix} -\kappa U_t \\ \kappa_u(\mu_u - V_t^u) \end{bmatrix} dt + \begin{bmatrix} \sqrt{V_t^u} & 0 \\ 0 & \sigma_u\sqrt{V_t^u} \end{bmatrix} d\begin{bmatrix} W_{1t}^u \\ W_{2t}^u \end{bmatrix}, \tag{50}$$

where $W^u = (W_1^u, W_2^u)$ is standard bivariate Brownian motion. Note that U is specified as a zero-mean stationary process having stochastic volatility generated as the Feller's square root process V^u. The parameter values used in our simulations are from U in Model IV, estimated and fitted to the model in (50). In particular, we set $\kappa = 20, \kappa_u = 2, \mu_u = 0.0002$ and $\sigma_u = 0.008$. Each simulated data series is initialized with the instantaneous volatilities V_0^x and V_0^u at their unconditional means, the log stock index price X_0 at $\ln(100)$, and the error process U_0 at 0. The initial 500 observations are discarded to remove any potential initialization effects.

The simulated means of the test statistics for $T = 30$ are presented in Figure 11, and the simulated rejection probabilities of the 5% test for both $T = 30$ and $T = 50$ are presented in Figure 12. They correspond to the lower panels in Figures 6 and 7. Once again, our simulation results are consistent with our theory. They clearly demonstrate that the tests employing the NW and RT bandwidths tend to diverge as sampling frequency increases, whereas the tests employing the AD and CRT bandwidths do not show such a tendency. Such contrast in the limit behaviors of the tests is much more conspicuous between the tests with the RT and CRT bandwidths, than between the tests with the NW and AD bandwidths. We do not attempt to further interpret the details of our simulation results here. There are many aspects of our simulation models, including their parameter values, that could affect the performances of the tests with different choices of bandwidths in different ways, and we do not believe that we can unravel the individual effect of any one of those aspects on the test with a particular choice of bandwidth.