Nonstationary Nonlinear Heteroskedasticity in Regression[1]

Heetaik Chung
School of Management and Economics
Handong University

and

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

Abstract

This paper considers the regression with errors having nonstationary nonlinear heteroskedasticity. For both the usual stationary regression and the nonstationary cointegrating regression, we develop the asymptotic theories for the least squares methods in the presence of conditional heterogeneity given as a nonlinear function of an integrated process. The conditional heteroskedasticity generated by an integrated process has more fundamental effects on the regression asymptotics than the one generated by a stationary process. In particular, the nonstationarity of volatility in the regression errors may induce spuriousness of the underlying regression. This is true for both the usual stationary regression and the nonstationary cointegrating regression, if excessive nonstationary volatility is present in the errors. Mild nonstationary volatilities do not render the underlying regression spurious. However, their presence makes the least squares estimator asymptotically biased and inefficient and the usual chi-square test invalid. We provide some illustrations to demonstrate the empirical relevancy of the model and theory developed in the paper. For this purpose, examined are US consumption function, EURO/USD forward-spot spreads and capital-asset pricing models for some major NYSE stocks.

This version: August 20, 2005

JEL classification codes: C12, C13, C22.

Key words and phrases: volatility, nonstationary nonlinear heteroskedasticity, regression with heteroskedastic errors, spurious regression

---

[1]We are grateful to a Co-editor and three anonymous referees for many useful comments, which have greatly improved an earlier version of this paper. We also like to thank Yoosoon Chang for many helpful discussions. This paper was written while Chung visited Department of Economics, Rice University and Bauer College of Business, University of Houston during Fall 2002 - Spring 2004. Chung gratefully acknowledges their hospitality. Address correspondence to: Joon Y. Park, Department of Economics - MS 22, Rice University, 6100 Main Street, Houston, TX 77005-1892, Tel: 713-348-2354, Fax: 713-348-5278, Email: jpark@rice.edu.

1. Introduction

Conditional heteroskedasticity in time series has long been routinely modeled as ARCH and GARCH processes. While the ARCH and GARCH type models, due respectively to Engle (1982) and Bollerslev (1986), have been very successful in explaining the observed volatility clustering, they exclude the important possibility that conditional heteroskedasticity may be accounted for by other explanatory variables. Recently, Park (2002) introduced nonstationary nonlinear heteroskedasticity (NNH) to exploit this possibility. The NNH model specifies the conditional heteroskedasticity as a function of some explanatory variables, completely in parallel with the conventional approach. There are, however, two important aspects that are highlighted in the NNH model. First, since conditional variance must be nonnegative and the explanatory variable usually may take negative values, the heterogeneity generating function (HGF) must in general be nonlinear. Secondly, in many potentially interesting applications, the variable affecting the conditional heteroskedasticity is nonstationary and typically follows a random walk.

The nonlinearity and nonstationarity in volatility generates several interesting characteristics for the NNH model. These characteristics are well demonstrated in Park (2002). If the conditional heteroskedasticity is driven by an HGF that is integrable up to a constant shift, the squared process has the asymptotic autocorrelation function that is consistent with stationary long memory or I(d) processes. Unlike the usual ARCH and GARCH models, the NNH model with an integrable HGF has autocorrelation in volatilities that decays at a polynomial rate. If, on the other hand, the conditional heteroskedasticity is generated by a non-constant asymptotically homogeneous HGF, then the asymptotic autocorrelation function of the squared process is given by a random constant, i.e., it takes a constant value across all lags that are given randomly. The autocorrelation pattern of their volatilities are therefore more comparable to integrated ARCH and IGARCH models in this case. Both nonlinearity and nonstationarity are essential for generating all these characteristics of the NNH model, which appear to be frequently observed in many economic and financial time series.

The NNH model is closely related to the volatility model considered by Hansen (1995). However, his model specifies the volatility as a function of normalized (near-integrated) processes and does not have the characteristic features of the NNH model which are mainly due to the presence of stochastic trend in the variable generating volatility. The essential aspects of the NNH model are not to be generated by the volatility driven by the normalized processes, which show no stochastic trends. It is also possible to regard the NNH model as a special stochastic volatility model. See, e.g., Shephard (2005) for an introduction and survey of the stochastic volatility model. The two models are, however, considered in quite different frameworks. In particular, the latter is usually specified and analyzed in continuous time framework, while the former is given intrinsically in discrete time. For this reason, the theories developed for the NNH model are not directly applicable to the general stochastic volatility models. However, it is clear that they would shed lights on the asymptotics for some of stochastic volatility models.

This paper considers the regression models with NNH in the regression errors. There seem to be plenty of examples for such regressions. For the purpose of illustration, we present some compelling cases in Figures 1 and 2. There we explicitly look at the following three regressions. First, the time series regression of consumption function, which specifies consumption expenditure as a linear function of income level. It is natural to expect that the errors in this regression have the conditional heteroskedasticity given as a function of income level. Second, the forward-spot exchange rate model that is often used to test for the unbiasedness of the forward exchange rates. Here we have some strong evidence that forward-spot spread volatility is determined by the levels of the spot exchange rates. Third, the well known regressions for the capital asset pricing model (CAPM), which relate the excess returns from individual stocks to those from the market. Again, we find some convincing evidence that the errors in the CAPM regression typically have the conditional heteroskedasticity that is given as a function of the market levels. In all these three regressions, the presence of NNH in the errors appears to be quite clear.

In the paper, we develop the asymptotics for the regression models with NNH in the regression errors. We consider both the stationary regression and cointegrating regression. Among the three aforementioned illustrative examples, the first two are the cointegrating regression and the third is the stationary regression. Our asymptotics in the paper bring out a number of interesting characteristics of the regressions with NNH in the errors. Most importantly, the presence of NNH in the errors may induce spuriousness of the underlying regression. When the NNH is given by an asymptotically homogeneous HGF that is explosive, the OLS estimator indeed becomes inconsistent. That is, the presence of excessive volatility may render the OLS procedure meaningless and nonsensical. This makes it clear that the nonstationarity in the error volatility may also result in spurious regressions, and extends the results by Granger and Newbold (1974) and Phillips (1986), which found and analyzed the spuriousness caused by the mean nonstationarity in the errors.

If the HGF is an asymptotically homogeneous function that is not excessively explosive or an integrable function, the regression models with NNH in the errors can be estimated consistently by OLS. In other words, we may still run meaningful OLS regression when NNH in the errors are relatively mild. However, the OLS estimator is in general biased and inefficient asymptotically. Furthermore, the standard Wald statistic does not have chi-square limiting null distribution, and the usual chi-square test based on the Wald statistic becomes invalid. The White correction for heteroskedasticity of unknown form is also generally invalid, and works only in an ideal situation that is not likely to happen in practical applications. All our theories are well confirmed by the simulations. When we have mild NNH in the errors, the distribution of the OLS estimator becomes more concentrated around the true value as the sample size increases, though the bias does not diminish. This is not so for the explosive NNH errors, in which case the OLS estimator becomes inconsistent. On the other hand, both the Wald test and the test with the White heteroskedasticity correction yield the actual rejection probabilities that are far from their nominal sizes, regardless of whether the NNH is mild or explosive.

The rest of the paper is organized as follows. Section 2 presents the model and assumptions. More specifically, the regression models with NNH in the errors are introduced with a set of regularity conditons. Both the stationary and cointegrating regressions are considered. Some tools necessary to develop the subsequent asymptotics are also introduced. The limit theories of the OLS estimator and the standard Wald statistic are developed in Section

3. For the stationary and cointegrating regressions, the effects of the presence of NNH in the errors are analyzed. Our analysis here covers the NNH models with both integrable and asymptotically homogeneous HGF's. It is shown that the OLS estimator is generally biased and inefficient asymptotically, and the test based on the standard Wald statistic is invalid except for very special cases. In Section 4, we consider the White correction for heteroskedasticity in the errors, and subsequently develop the asymptotics for the test with the White heteroskedasticity correction. Section 5 examines through simulations the finite sample performances of the estimator and tests considered in the paper. Some concluding remarks follow in Section 6, and the proofs are given in Mathematical Appendix.

2. The Model and Assumptions

We consider the regression model given by

$$y_t = x_t'\beta + u_t, \tag{1}$$

where (y_t) and (x_t) are respectively the regressand and regressor to be specified in detail later, and (u_t) is the regression error that is further modelled as

$$u_t = \sigma(z_{t-1})\,\varepsilon_t, \tag{2}$$

where (z_t) is an integrated time series, and for a filtration $(\mathcal{F}_t)$ to which (z_t) is adapted, $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence such that

$$\mathbf{E}\left(\varepsilon_t^2\middle|\,\mathcal{F}_{t-1}\right) = 1. \tag{3}$$

The specifications in (2) and (3) will be maintained throughout the paper.

Under such specifications, we have

$$\mathbf{E}\left(u_t^2\middle|\,\mathcal{F}_{t-1}\right) = \sigma^2(z_{t-1})$$

and our model (1) becomes the regression model with errors having NNH, introduced recently by Park (2002). We will let the regressors (x_t) be either stationary or integrated. If they are stationary, regression (1) becomes the stationary regression with NNH in the errors. If they are integrated, on the other hand, regression (1) reduces to the cointegrating regression with NNH in the errors. Both for the stationary regression and the cointegrating regression, we assume that

$$\mathbf{E}\left(x_t u_t\middle|\,\mathcal{F}_{t-1}\right) = 0,$$

i.e., the conditional orthogonality holds between the regressor and the regression error. For the stationary regression, this only requires to assume $\mathbf{E}(x_t\varepsilon_t|\mathcal{F}_{t-1}) = 0$. However, we need to impose a somewhat stronger condition for the cointegrating regression. In this case, we assume following Park and Phillips (2001) that (x_t) is predetermined with respect to the filtration $(\mathcal{F}_t)$, i.e., (x_t) is $(\mathcal{F}_{t-1})$-measurable.[2] Note that in both cases the regressors are allowed to be correlated with the variable generating heterogeneity.

[2] The reader is referred to Chang and Park (2003) for the effect of endogeneity in nonlinear models with integrated time series.

The function $\sigma : \mathbb{R} \to \mathbb{R}_+$ will be referred to as the *heterogeneity generating function* (HGF) in what follows. Clearly, σ must be a nonlinear function, since it has to be nonnegative. More specifically, we consider two classes of functions: *integrable* and *asymptotically homogeneous functions*. These function classes were first introduced by Park and Phillips (1999) in their study on the asymptotics of nonlinear transformations of integrated time series. They will be denoted respectively by $\mathbb{I}$ and $\mathbb{H}$. More specificially,

Definition 2.1 We let $\sigma \in \mathbb{I}$ if σ is Riemann integrable and $\int_{-\infty}^{\infty} s^2\sigma^2(s)\,ds < \infty$. On the other hand, we write $\sigma \in \mathbb{H}$ if

$$\sigma(\lambda s) = \nu(\lambda)\tau(s) + o(\nu(\lambda))$$

for large λ uniformly in s over any compact interval, where τ is locally Riemann integrable. For $\sigma \in \mathbb{H}$, we call ν and τ, respectively, the *asymptotic order* and *limit homogeneous function* of σ.

The reader is referred to Park and Phillips (1999, 2001) and Park (2003) for more details on these function classes.

The classes $\mathbb{I}$ and $\mathbb{H}$ include a wide class, if not all, of transformations defined on $\mathbb{R}$. The bounded functions with compact supports and more generally all bounded integrable functions with fast enough decaying rates, for instance, belong to the class $\mathbb{I}$. On the other hand, power functions $ax_+^c + bx_-^c$ with nonnegative power $c \geq 0$ belong to the class $\mathbb{H}$ having asymptotic order λ^c and themselves as limit homogeneous functions. Moreover, logistic function $e^x/(1+e^x)$ and all the other distribution function-like functions are also the elements of the class $\mathbb{H}$ with asymptotic order 1 and limit homogeneous function $1\{x \geq 0\}$. Our formulation of the class $\mathbb{H}$ relies on that of Park and Phillips (1999, 2001), and does not include the logarithmic functions or the power functions of negative powers having a pole-type discontinuity at the origin. The extension to allow for such functions is possible and has been done recently by de Jong (2002) and Park (2003). We will not, however, attempt such an extension here, since it seems to have little empirical relevancy in our context.

Of the two classes of functions, $\mathbb{H}$ appears to be more relevant in practical applications. In many cases, the variances of errors in time series regressions are positively correlated with the absolute levels of other covariates. Therefore, if NNH is observed, it is more likely to be generated by an asymptotically homogeneous HGF. In fact, all of our empirical applications have NNH in the errors driven by HGF's that are asymptotically homogeneous and belong to the class $\mathbb{H}$. If NNH in the errors is driven by integrable HGF's, it implies in particular that the error variances decrease with time. We do not expect this to be usual and often observed in the actual applications. However, it is not totally unlikely. For instance, if the HGF is given by $\sigma = f \circ g$ with f locally integrable and g bounded, it generally belongs to $\mathbb{I}$. Such a case arises when NNH is generated by a locally integrable function of a bounded transformation of a random walk. See Miller and Park (2005) for such an example.

In the subsequent development of our theory, we will need to deal with stationary processes satisfying certain regularity conditions. In particular, we let

$$\triangle z_t = c(L)\eta_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}, \tag{4}$$

where (η_t) are a sequence of independent and identically distributed (iid) random variables, and require that $(\triangle z_t)$ be regular in the sense defined in the following definition.

Definition 2.2 We call $(\triangle z_t)$ defined as in (4) a *regular* linear process if (a) $c(1) \neq 0$ and $\sum_{i=0}^{\infty} i|c_i| < \infty$, $\mathbf{E}(\eta_t) = 0$ and $\mathbf{E}|\eta_t|^{2+\delta} < \infty$ for some $\delta > 0$, and if (b) (η_t) has distribution that is absolutely continuous with respect to Lebesgue measure and has characteristic function φ satisfying the condition $\lim_{t\to\infty} t^\delta \varphi(t) = 0$ for some $\delta > 0$.

The conditions in (a) of Definition 2.2 for the summability of the coefficients and the moment for the innovation are standard and routinely employed to develop the asymptotics for models involving linear processes. The conditions on the distribution of the innovations in (b) of Definition 2.2 are more restrictive, though they are widely satisfied.

To ease our exposition, we also introduce some additional terminologies and notations here. The standard notations like $\to_p$ and $\to_d$ are used throughout the paper to denote convergence in probability and convergence in distribution, respectively. Moreover, we say that a sequence (v_t) of random vectors satisfies *invariance principle* if the stochastic process B_n defined on the unit interval $[0,1]$ by

$$B_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t,$$

converges weakly to a vector Brownian motion, i.e.,

$$B_n \to_d B$$

as $n \to \infty$, where B is a vector Brownian motion. Here $[s]$ denotes the largest integer which does not exceed s. This type of invariance principle is one of the essential tools to develop asymptotics for the models that involve integrated time series. It is known to hold for a variety of stationary and possibly heterogeneous processes. Throughout the paper, $\mathbb{N}$ signifies the normal law, both for univariate and multivariate distributions. In later sections, some specific class of compound normal distributions will appear. We say that a random vector W has *central normal mixture distribution with mixing variate* Q or $\mathbb{NM}(0,Q)$ if, conditional on Q, W is distributed as $\mathbb{N}(0,Q)$, i.e., normal distribution with mean zero and covariance matrix Q. Finally, we use as usual the notation $\|\cdot\|$ to denote the Euclidean norm.

We assume one of the following two sets of assumptions, depending upon whether the regessor (x_t) in regression (1) is stationary or nonstationary. For the case of stationary (x_t),

Assumption 2.1 We assume that
(a) $(1/n)\sum_{t=1}^n x_t x_t' \to_p M > 0$ as $n \to \infty$,
(b) $(x_t\varepsilon_t, \triangle z_t)$ satisfies an invariance principle with limit Brownian motion (U,V),
(c) $(x_t\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence such that $(1/n)\sum_{t=1}^n \mathbf{E}(\varepsilon_t^2 x_t x_t' | \mathcal{F}_{t-1}) \to_p \Sigma$ and $\sup_{t\geq 1} \mathbf{E}(\|x_t\varepsilon_t\|^4 | \mathcal{F}_{t-1}) < \infty$ a.s.,

(d) $(\triangle z_t)$ is a regular linear process, and

(e) $\sup_{t\geq 1}\mathbf{E}\|x_t\|^4<\infty$.

The conditions in (a) and (b) are fairly weak and satisfied under the usual assumptions imposed routinely on standard regressions. The martingale difference requirement in (c) is needed to develop the asymptotics for our models that include nonstationary nonlinearity. In particular, the regularity condition in (d) is necessary to obtain the limit theory for the models with integrable HGF's. It is not required for the models with asymptotically homogeneous HGF's. The conditions in (c) and (d) are certainly more restrictive than those in (a) and (b). However, they do not seem to be prohibitively so, and indeed satisfied by a variety of models that are used in practical applications. We will discuss more on this issue in a later section. The moment condition in (e) is necessary to validate our method to correct for heteroskedasticity.

There are two special conditions, under which our subsequent asymptotics become significantly simpler. The first condition is that

$$V \text{ is independent of } U, \tag{5}$$

where U and V are the limit Brownian motions appearing in (b) above. As is well known, they become independent when the longrun correlation of $(x_t\varepsilon_t)$ and $(\triangle z_t)$ vanishes. This would have an important consequence on our subsequent asymptotics, as we will see later. The second condition is given by

$$M=\Sigma, \tag{6}$$

where M and Σ are introduced respectively in (a) and (c) above. This also affects our limit theories in an important way. It holds if (x_t) is predetermined with respect to the filtration $(\mathcal{F}_t)$, as we have

$$\mathbf{E}\left(\varepsilon_t^2x_tx_t'\middle|\,\mathcal{F}_{t-1}\right)=x_tx_t' \text{ a.s.}$$

The condition is also very likely to be met, for instance, when (x_tx_t') and (ε_t^2) are conditionally uncorrelated. In this case, we have $\mathbf{E}\left(\varepsilon_t^2x_tx_t'\middle|\,\mathcal{F}_{t-1}\right)=\mathbf{E}\left(x_tx_t'\middle|\,\mathcal{F}_{t-1}\right)$ and, as shown in, e.g., Hall and Heyde (1980, Theorem 2.19),

$$\frac{1}{n}\sum_{t=1}^n x_tx_t'-\frac{1}{n}\sum_{t=1}^n\mathbf{E}\left(x_tx_t'\middle|\,\mathcal{F}_{t-1}\right)\to_p 0$$

whenever there is a bounding random variable for (x_tx_t').

For the case of nonstationary (x_t),

Assumption 2.2 We assume that

(a) $z_{t-1}=\alpha' x_t$,

(b) $(\varepsilon_t,\triangle x_t)$ satisfies an invariance principle with limit vector Brownian motion (U,V) with V nondegenerate,

(c) $(\varepsilon_t,\mathcal{F}_t)$ is a martingale difference sequence such that $\sup_{t\geq 1}\mathbf{E}(|\varepsilon_t|^{2+\delta}|\mathcal{F}_{t-1})<\infty$ a.s. for some $\delta>0$,

(d) (x_t) is adapted to $(\mathcal{F}_{t-1})$, and

(e) $(\triangle z_t)$ is a regular linear process.

Our formulation in (a) of the heterogeneity generating variable is not essential for the development of our theory in the paper. It is just one possible specification, among many others, which seems to be plausible for many practical applications. We may for instance have the heterogeneity generating variable that is totally independent of the regressors. This causes no difficulty and can easily be accommodated in the framework of our subsequent theoretical development. The condition in (b) is standard for cointegrating regressions. The more stringent requirements in (c), (d) and (e) are introduced to deal with nonstationary and nonlinearity in our models. As for the case of stationary (x_t), the regularity condition in (e) is required only for the models with integrable HGF's. For the cointegrating regression, the condition (5) for the limit Brownian motions U and V introduced in (b) implies that the regressor and the regression error are asymptotically independent. This condition entails some important consequences in the asymptotics of the cointegrating regression with NNH in the errors.

As in Park (2002), our subsequent theory for the case of integrable HGF involves the *Brownian local time*. Let B be a scalar Brownian motion.[3] The local time for B is then defined as

$$L(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|B(r) - s| < \varepsilon\}\, dr$$

Roughly speaking, 2ε times $L(t,s)$ measures the actual time spent by B in the ε- neighborhood of s up to time t. The local time yields the *occupation times formula*

$$\int_0^t T(B(r))\, dr = \int_{-\infty}^{\infty} T(s) L(t,s)\, ds$$

for any $T : \mathbb{R} \to \mathbb{R}$ locally integrable. For each t, the occupation times formula allows us to evaluate the time integral of a nonlinear function of Brownian motion by means of the integral of the function itself weighted by the local time.

3. Asymptotic Theory for Regressions with NNH

In this section, we establish the asymptotic theory for our model given by (1) – (3). In particular, we investigate how the usual inferential procedure based on the least squares would be affected by the presence of NNH in the errors for both the usual stationary regressions and the nonstationary cointegrating regressions. In what follows, we denote by $\hat{\beta}$ the OLS estimator of β in regression (1), for which we have

$$\hat{\beta} - \beta = \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \sum_{t=1}^{n} \sigma(z_{t-1}) x_t \varepsilon_t \tag{7}$$

[3]The local time of a vector Brownian motion does not exist, at least in the sense used here.

due to (1) and (2). Moreover, we let $\hat{\sigma}_u^2$ be the usual error variance estimator from regression (1), which is given by

$$\hat{\sigma}_u^2 = \sum_{t=1}^n \sigma^2(z_{t-1})\varepsilon_t^2 - \left(\sum_{t=1}^n \sigma(z_{t-1})x_t'\varepsilon_t\right)\left(\sum_{t=1}^n x_t x_t'\right)^{-1}\left(\sum_{t=1}^n \sigma(z_{t-1})x_t\varepsilon_t\right). \tag{8}$$

The general linear null hypothesis on β is routinely formulated as

$$\mathrm{H}_0 : R\beta = r$$

with a pair of the restriction matrix and vector (R, r), and tested against the alternative hypothesis H_1: $R\beta \neq r$ using the Wald statistic defined by

$$F(\hat{\beta}) = \left(R\hat{\beta} - r\right)'\left(\hat{\sigma}_u^2 R\left(\sum_{t=1}^n x_t x_t'\right)^{-1} R'\right)^{-1}\left(R\hat{\beta} - r\right), \tag{9}$$

where $\hat{\beta}$ and $\hat{\sigma}_u^2$ are respectively the estimated regression coefficient and error variance introduced above for regression (1). It is assumed that the matrix R has q-linearly independent rows. Below we present the limiting distributions of $\hat{\beta}$ and $F(\hat{\beta})$ for the stationary and cointegrating regressions in sequel.

3.1 Stationary Regression

Now we suppose that (x_t) is a stationary process satisfying Assumption 2.1. First, we look at the case where the HGF is asymptotically homogeneous.

Theorem 3.1 *Let $\sigma \in \mathbb{H}$, and let* Assumption 2.1 (a)–(c) *hold. We have*

$$\sqrt{n}\nu(\sqrt{n})^{-1}(\hat{\beta} - \beta) \to_d M^{-1}N$$

and, under H_0,

$$F(\hat{\beta}) \to_d \frac{N'M^{-1}R'\left(RM^{-1}R'\right)^{-1}RM^{-1}N}{S},$$

where

$$N = \int_0^1 \tau(V(r))\,dU(r) \quad \text{and} \quad S = \int_0^1 \tau^2(V(r))\,dr.$$

If $\nu(\sqrt{n})/\sqrt{n} \to \infty$, $F(\hat{\beta})$ has the same limiting distribution also under H_1, while we have $F(\hat{\beta}) \to_p \infty$ under H_1 if $\nu(\sqrt{n})/\sqrt{n} \to 0$.

We may readily derive the results in Theorem 3.1 from (7), (8) and (9) using in particular the nonstationary nonlinear asymptotics

$$n^{-1/2}\nu(\sqrt{n})^{-1}\sum_{t=1}^n \sigma(z_{t-1})x_t\varepsilon_t \approx n^{-1/2}\sum_{t=1}^n \tau\left(\frac{z_{t-1}}{\sqrt{n}}\right)x_t\varepsilon_t \to_d \int_0^1 \tau(V(r))\,dU(r)$$

and

$$n^{-1}\nu(\sqrt{n})^{-2}\sum_{t=1}^{n}\sigma^2(z_{t-1})\varepsilon_t^2 \approx n^{-1}\sum_{t=1}^{n}\tau^2\left(\frac{z_{t-1}}{\sqrt{n}}\right) \to_d \int_0^1 \tau^2(V(r))\,dr$$

obtained by Park and Phillips (1999, 2001).

It is interesting to note that the presence of NNH in the errors affects the convergence rate, as well as the limiting distribution, of the OLS estimator. This is in sharp constrast with the case where the heterogeneity is given as a function of stationary variates, which in general changes only the distributional results for the OLS estimator leaving the convergence rate unaffected. In particular, the OLS estimator becomes consistent, only when $\nu(\lambda)/\lambda \to 0$ as $\lambda \to \infty$. If $\nu(\lambda)/\lambda \nrightarrow 0$ as $\lambda \to \infty$, the underlying regression is not consistently estimable and becomes spurious. It is now clear that the presence of NNH in the errors may also induce spuriousness of the OLS regression, if the heteroskedasticity is strong and explosive. This extends the results by Granger and Newbold (1974) and Phillips (1986), who observed and analyzed the spurious regressions caused by the mean nonstationarity in the errors. The OLS estimator is consistent, as long as $\nu(\lambda)/\lambda \to 0$ as $\lambda \to \infty$ and we only have mild heteroskedasticity. The convergence rate, however, is reduced whenever $\nu(\lambda) \to \infty$ as $\lambda \to \infty$.

If $\nu(\lambda) = \lambda^\kappa$, for instance, then the convergence rate for the OLS estimator becomes $n^{(1-\kappa)/2}$. Consequently, if $\kappa \geq 1$, the OLS estimator becomes inconsistent and the underlying regression reduces to a spurious regression. Roughly, this occurs when the heteroskedasticity becomes explosive and strong, and the HGF induces volatilities in the errors at least as big as those that can be generated by the linear function. On the other hand, the OLS estimator remains to be consistent as long as $\kappa < 1$, i.e., the HGF generates less volatilities in the errors than the linear function and the induced heteroskedasticity is relatively mild. In general, the limiting distribution of the OLS estimator is nonnormal and not centered around zero. In particular, it is biased unless condition (5) holds, and it is generally inefficient as we will show later in more detail.

The test based on the Wald statistic $F(\hat{\beta})$ becomes inconsistent, if $\nu(\lambda)/\lambda \to \infty$ as $\lambda \to \infty$. This is well predicted, since under this condition the heteroskedasticity is excessive and the OLS estimator becomes inconsistent. In fact, in this case, the Wald statistic has the same limiting distribution under both the null and alternative hypotheses. On the other hand, the test becomes consistent if the heterogeneity is mild and $\nu(\lambda)/\lambda \to 0$ as $\lambda \to \infty$. Even in this case, however, the test becomes invalid if the usual chi-square critical values are used. The limiting null distribution of the statistic $F(\hat{\beta})$ is nonstandard, and depends upon various nuisance parameters. The nuisance parameter dependency is in general quite complicated. Nevertheless, there is a special case where the limit distribution of $F(\hat{\beta})$ becomes chi-square. This indeed happens when both conditions (5) and (6) hold. Note that we have

$$N =_d S^{1/2}\mathbb{N}(0, M)$$

in this case. See the paragraph following Assumption 2.1 for more discussions on these conditions.

Theorem 3.2 *Let $\sigma \in \mathbb{I}$, and let* Assumption 2.1 (a)–(d) *hold. We have*

$$n^{3/4}(\hat{\beta} - \beta) \to_d M^{-1}N,$$

where N is distributed as central normal mixture with mixing variate given by

$$\left(L(1,0) \int_{-\infty}^{\infty} \sigma^2(s)\, ds \right) \Sigma.$$

Moreover, we have

$$F(\hat{\beta}) \to_d Z'M^{-1}R' \left(RM^{-1}R' \right)^{-1} RM^{-1}Z$$

under H_0, where Z is distributed as $\mathbb{N}(0,\Sigma)$, and $F(\hat{\beta}) \to_p \infty$ under H_1.

The results in Theorem 3.2 can also be deduced from (7), (8) and (9), due to the nonstationary nonlinear asymptotics

$$n^{-1/4} \sum_{t=1}^{n} \sigma(z_{t-1}) x_t \varepsilon_t \to_d \mathbb{NM} \left(0, \left(L(1,0) \int_{-\infty}^{\infty} \sigma^2(s)\, ds \right) \Sigma \right)$$

and

$$n^{-1/2} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \varepsilon_t^2 \approx n^{-1/2} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \to_d L(1,0) \int_{-\infty}^{\infty} \sigma^2(s)\, ds$$

established in Park and Phillips (1999, 2001).

If the regression errors have NNH with an integrable HGF, then the convergence rate for the OLS estimator becomes faster. There is no possibility that NNH in the errors induces the spurious regression. This is contrary to the models with the errors having NNH generated by an asymptotically homogeneous HGF. As shown above, the rate is $n^{3/4}$ and an order of magnitude faster than the rate in the stationary regression without NNH. The limit distribution of the OLS estimator is mixed normal, with the mixing variate given by a Brownian local time. It is therefore asymptotically unbiased, though not efficient as we will show in the next section. The Wald statistic, albeit consistent, in general does not have a limiting chi-square distribution. The Wald test thus becomes invalid if it is based on the usual chi-square critical values. Its limiting distribution depends upon various nuisance parameters. There is, however, a special case where it has the usual chi-square limiting distribution. This case arises when (6) holds. In this case, it is indeed straightforward to show that $F(\hat{\beta}) \to_d \chi_q^2$. For the validity of the Wald test, we do not require condition (5) to hold here. This is in contrast with the model having NNH given by an asymptotically homogeneous function, for which both conditions (5) and (6) are necessary to have the chi-square limiting distribution for the Wald statistic.

3.2 Cointegrating Regression

We now assume that the conditions in Assumption 2.2 hold and (1) becomes the cointegrating regression. Similarly as before, we separately consider the two cases where the HGF is asymptotically homogeneous and integrable.

Theorem 3.3 *Let $\sigma \in \mathbb{H}$, and let* Assumption 2.2 (a)–(d) *hold. We have*

$$n\nu(\sqrt{n})^{-1}(\hat{\beta} - \beta) \to_d W^{-1}Z$$

and, under H_0,

$$F(\hat{\beta}) \to_d \frac{Z'W^{-1}R'\left(RW^{-1}R'\right)^{-1}RW^{-1}Z}{S},$$

where

$$W = \int_0^1 V(r)V(r)'dr, \quad Z = \int_0^1 \tau(\alpha'V(r))V(r)\,dU(r), \quad S = \int_0^1 \tau^2(\alpha'V(r))\,dr.$$

If $\nu(\sqrt{n})/n \to \infty$, $F(\hat{\beta})$ has the same limiting distribution also under H_1, while we have $F(\hat{\beta}) \to_p \infty$ under H_1 if $\nu(\sqrt{n})/n \to 0$.

The results in Theorem 3.3 follow immediately from (7), (8) and (9) with $z_{t-1} = \alpha' x_t$, and the nonstationary nonlinear asymptotics

$$n^{-1}\nu(\sqrt{n})^{-1}\sum_{t=1}^{n}\sigma(z_{t-1})x_t\varepsilon_t \approx n^{-1/2}\sum_{t=1}^{n}\tau\left(\frac{\alpha' x_t}{\sqrt{n}}\right)\frac{x_t}{\sqrt{n}}\varepsilon_t \to_d \int_0^1 \tau(\alpha'V(r))V(r)\,dU(r)$$

and

$$n^{-1}\nu(\sqrt{n})^{-2}\sum_{t=1}^{n}\sigma^2(z_{t-1})\varepsilon_t^2 \approx n^{-1}\sum_{t=1}^{n}\tau^2\left(\frac{\alpha' x_t}{\sqrt{n}}\right) \to_d \int_0^1 \tau^2(\alpha'V(r))\,dr$$

in Park and Phillips (1999, 2001).

Just as in the case of the stationary regression, the presence of NNH in the errors affects the convergence rate of the OLS estimator in the cointegrating regression. If the HGF σ is asymptotically homogeneous and has the asymptotic order ν such that $\nu(\lambda) \to \infty$ as $\lambda \to \infty$, then the convergence rate is reduced by an order of magnitude to $n/\nu(\sqrt{n})$. If $\nu(\lambda) = \lambda^\kappa$ with some $\kappa \geq 2$, the OLS estimator becomes inconsistent. The excessive volatility would thus render the underlying regression spurious, precisely as in the case of the stationary regression. Note, however, that the spuriousness appears in the cointegrating regression when the errors are more volatile, compared with the stationary regression. This is because the OLS estimator is super-consistent, converging faster to its true value, for the cointegrating regression. The OLS estimator is asymptotically biased unless condition (5) is met, and generally inefficient. As long as the OLS estimator is consistent, so is the Wald test. On the other hand, if there is explosive and strong heteroskedasticity and the OLS estimator is inconsistent, then the Wald test becomes inconsistent as well. In general, the Wald statistic has nonstandard limiting distribution, as in the case of the stationary regression. Moreover, it depends upon various nuisance parameters. Once again, the test relying on the chi-square critical values is invalid. The invalidity of the Wald test is still true under condition (5), due to the presence of heteroskedasticity.

To present the limit theory for the cointegrating regression having NNH in the errors with an integrable HGF, we first introduce an $(m \times m)$-dimensional orthogonal matrix

$H = (h_1, H_2)$. Without loss of generality, we let $\|\alpha\| = 1$ and let $h_1 = \alpha$ by redefining, if necessary, the HGF σ. The components h_1 and H_2 of H are of dimensions $m \times 1$ and $m \times (m-1)$, respectively. Then we define

$$V_1 = h_1'V \quad \text{and} \quad V_2 = H_2'V$$

Note that V_1 is a scalar, and V_2 is an $(m-1)$-dimensional vector. By convention, we assume that all terms transformed by H_2 disappear if (x_t) is indeed a scalar and $m = 1$. Moreover, we use subscripts "1" and "2" to denote the subvectors and submatrices that are partitioned conformably with V_1 and V_2.

Theorem 3.4 *Let $\sigma \in \mathbb{I}$, and let* Assumptions 2.2 (a)–(e) *hold. Moreover, we define $D_n = diag\,(n^{7/4}, n^{5/4}I_{m-1})$ and $W_n = (1/n^2)\sum_{t=1}^n x_t x_t'$. We have*

$$D_n H' W_n(\hat\beta - \beta) \to_d Z \quad \text{and} \quad W_n \to_d W$$

jointly, where $W = \int_0^1 V(r)V(r)'\,dr$ and Z is distributed as central normal mixture with mixing variate

$$\int_{-\infty}^{\infty}\int_0^1 \begin{pmatrix} \sigma(s)s \\ \sigma(s)V_2(r) \end{pmatrix} \begin{pmatrix} \sigma(s)s \\ \sigma(s)V_2(r) \end{pmatrix}' L_1(dr, 0)\,ds.$$

If $m = 1$, then $F(\hat\beta) \to_p 0$ and $F(\hat\beta) \to_p \infty$, respectively under H_0 and H_1. If $m \geq 2$, then we have under H_0

$$F(\hat\beta) \to_d \frac{N'W^{-1}R'\left(RW^{-1}R'\right)^{-1}RW^{-1}N}{S},$$

where $N = H_2 Z_2$ and $S = L_1(1,0)\int_{-\infty}^{\infty}\sigma^2(s)\,ds$, while $F(\hat\beta) \to_p \infty$ under H_1.

To derive the results in Theorem 3.4, we note that

$$n^{-1/2}\sum_{t=1}^n \sigma^2(z_{t-1})\varepsilon_t^2 \approx n^{-1/2}\sum_{t=1}^n \sigma^2(\alpha' x_t) \to_d L_1(1,0)\int_{-\infty}^{\infty}\sigma^2(s)\,ds.$$

Moreover, we have

$$\sum_{t=1}^n \sigma(z_{t-1})x_t\varepsilon_t = \sum_{t=1}^n \sigma(\alpha' x_t)x_t\varepsilon_t,$$

and it is easy to see that $(\sigma(\alpha' x_t)x_t)$ yields an integrable transformation of an integrated variable in the direction of α, while it reduces to a product of an integrable and homogeneous transformations of two integrated variables in all the other directions. The asymptotics for such mixtures of integrated variables are given in Chang and Park (2003).

The asymptotic properties of the OLS estimator in the cointegrating regression, at least qualitatively, are similar to those for the stationary regression that we considered earlier. The OLS estimator converges faster to its true value than in the model without NNH in the errors, and has mixed normal limiting distribution. Note that

$$\hat\beta - \beta \approx_d W^{-1}\left(n^{-7/4}h_1 Z_1 + n^{-5/4}H_2 Z_2\right).$$

Therefore, if $m = 1$, the convergence rate is $n^{7/4}$. If, however, $m \geq 2$, then it is reduced to $n^{5/4}$, in which case we have $n^{5/4}(\hat{\beta} - \beta) \to_d W^{-1}H_2Z_2$ as can be easily deduced. These differing rates of convergence render the asymptotic behavior of the Wald statistic to be dependent upon whether $m = 1$ or $m \geq 2$. In the former case, the Wald statistic converges under the null in probability to zero, which implies that the asymptotic size of the Wald test is zero regardless of the critical values used for the test. In the latter case, it has a well defined limiting null distribution, though it is not chi-square. Under the alternative, the Wald statistic diverges in both cases. The Wald test is therefore consistent for all values of m. In particular, we may say that it is "super-consistent" for the case of $m = 1$, since in this case it has zero asymptotic size under the null as well as unit asymptotic power under the alternative.

4. White Correction for Heteroskedasticity

As we have seen in the previous section, the OLS estimator and the Wald statistic have nonstandard limit theories in the regressions with NNH. In general the OLS estimator is nonnormal, and it is asymptotically biased. Likewise, the Wald statistic does not have limiting chi-square distribution except for some special cases, and consequently, it cannot be used as a basis for standard chi-square test. Its limiting distribution generally depends on various nuisance parameters.

In the presence of NNH in the errors, we are naturally led to consider the White correction for heteroskedastic errors. Therefore, we define the heteroskedasticity-corrected (HC) Wald statistic to be given by

$$G(\hat{\beta}) = \left(R\hat{\beta} - r\right)' \left(R \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \left(\sum_{t=1}^{n} \hat{u}_t^2 x_t x_t' \right) \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} R' \right)^{-1} \left(R\hat{\beta} - r\right)$$

where $\hat{\beta}$ is the OLS estimator for β and $(\hat{u}_t)$ are the usual OLS residuals in regression (1). The motivation for the HC Wald statistic is precisely the same as in the classical regression model: The HC Wald statistic is formulated with the White estimator for the asymptotic variance of $\hat{\beta}$, which is valid for the regressions with heteroskedasticity of unknown form. As is well known, the usual asymptotic covariance matrix of $\hat{\beta}$ is generally not applicable if the regression errors are heteroskedastic.

The limiting distribution of $G(\hat{\beta})$ can be obtained similarly as for our results in the previous section.

Theorem 4.1 *We have*

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

if one of the following conditions holds:
(i) if $\sigma \in \mathbb{I}$ and either Assumptions 2.1 (a)–(e) *or* Assumptions 2.2 (a)–(e) *hold, or*
(ii) if $\sigma \in \mathbb{H}$ and either Assumptions 2.1 (a)–(c) *and* (e) *hold with independent U and V in* (b) *and the asymptotic order of σ satisfies $\nu(\sqrt{n})/\sqrt{n} \to 0$, or* Assumptions 2.2 (a)–(d) *hold with independent U and V in* (b) *and the asymptotic order of σ satisfies $\nu(\sqrt{n})/n \to 0$.*

Under appropriate conditions, the HC Wald statistic indeed effectively corrects for heterogeneity in our regression models with NNH in the errors. Therefore, the White correction basically works for our models with the NNH errors, as well as for the classical regressions.

Correcting for heterogeneity is sufficient for the resulting test statistic to have chi-square limit distribution if the HGF is given by an integrable function. Theorem 4.1 shows in this case that the HC Wald statistic $G(\hat{\beta})$ has chi-square limiting distribution, and therefore, the usual chi-square test can be based on $G(\hat{\beta})$. No further condition is necessary. This is true for both the stationary and cointegrating regressions. The correction for heterogeneity, however, is not generally sufficient if the HGF is given by an asymptotically homogeneous function. Indeed, Theorem 4.1 makes it clear that we need an extra condition to ensure that the chi-square test using $G(\hat{\beta})$ is applicable: The independence of two limit Brownian motions U and V, i.e., the condition that we introduced earlier in (5).

5. Simulation Results

A set of simulations are conducted to investigate the finite sample performances of the estimator and test statistics considered in the paper. Both the stationary and cointegrating regressions with NNH are examined. For the HGF's, we choose the following four different types of functions:

(a) $\sigma(z) = \dfrac{e^{\alpha z}}{1 + e^{\alpha z}}$

(b) $\sigma(z) = |z|^{\alpha}$

(c) $\sigma(z) = e^{-\alpha z^2}$

(d) $\sigma(z) = e^{z}$

where $\alpha > 0$. For both the stationary and cointegrating regressions, we set the values of α's in (a) and (c) to be $0, 1$ and 0.02, respectively. For α's in (b), we choose the values $1/4, 1/2$ and 1 for the stationary regression, and $1/2, 1$ and 2 for the cointegrating regression.

In (a) and (b), we consider the asymptotically homogeneous HGF's. The logistic function in (a) has the asymptotic order $\nu(\lambda) = 1$ and the limit homogeneous function $\tau(s) = 1\{s \geq 0\}$. In particular, we have $\nu(\lambda)/\lambda,\ \nu(\lambda)/\lambda^2 \to 0$ as $\lambda \to \infty$, and the consistency condition holds for both the stationary and cointegrating regressions. Therefore the OLS estimator is consistent for both regressions. The parameter α determines how close the logistic function is to its limit homogeneous function, i.e., the two become closer as α gets large. The power function in (b) is homogeneous (and hence, asymptotically homogeneous) with asymptotic order $\nu(\lambda) = \lambda^{\alpha}$. For the stationary regression, the consistency condition $\nu(\lambda)/\lambda \to 0$ is therefore satisfied if $\alpha < 1$. For the cointegrating regression, the consistency condition is given by $\nu(\lambda)/\lambda^2 \to 0$, which is met as long as $\alpha < 2$. In contrast to (a) and (b), the Gaussian HGF in (c) is integrable. Roughly, the parameter α in the Gaussian function specifies the integrability of the function. As α gets large, the function becomes more integrable and vice versa. In our previous theory, we did not consider the exponential function in (d). It is, however, considered in our simulations, since it is used in most of the stochastic volatility literature.

For the stationary regression, $(\varepsilon_t), (x_t)$ and (z_t) are generated as

$$\begin{aligned} \varepsilon_t &= e_{1t} + e_{2t} \\ x_t &= e_{1t} - e_{2t} \\ \triangle z_t &= e_{1t}^2 - 1 \end{aligned}$$

where (e_{1t}) and (e_{2t}) are iid standard normals, which are independent of each other. It can be readily verified that $(\varepsilon_t), (x_t)$ and (z_t) specified as above satisfy all the conditions in Assumption 2.1. More specifically, we have $M = \mathbf{E}x_t^2 = 2$ in (a), and U and V in (b) are given by Brownian motions with variances 4 and 2 respectively with covariance 2. Furthermore, $\Sigma = 4$. Neither (5) nor (6) is satisfied for our specification here. Therefore, both heterogeneity and endogeneity are present in our simulation model for the stationary regression. Consequently the OLS estimator becomes asymptotically inefficient and biased, and the standard Wald test is invalid for all HGF's considered in the simulations.

For the cointegration regression, we generate $(\varepsilon_t), (x_t)$ and (z_t) as

$$\begin{aligned} \varepsilon_t &= e_{1t} + e_{2t} \\ \triangle x_t &= e_{1,t-1} \\ \triangle z_t &= e_{1t} \end{aligned}$$

where, just as above, (e_{1t}) and (e_{2t}) are independent iid standard normals. We may easily show that all the conditions in Assumption 2.2 hold for the specification introduced here. More precisely, we have $z_{t-1} = x_t$ in (a), and the limit Brownian motions U and V respectively have variances 2 and 1 with covariance 1. In particular, U and V are not independent. Therefore, endogeneity as well as heterogeneity is present in the cointegrating regression used in the simulations.

The OLS estimator, Wald tests and HC Wald tests are considered in the simulations. In the simulations, we look at the samples of sizes 100, 200 and 500. This is to consider both the samples of moderate and relatively large sizes, and also to investigate how the performances of the estimators and test statistics change as the sample size increases. The performances of the estimator and the tests are evaluated based on 100,000 iterations. The simulation results are summarized in Figure 3 and Tables 1 and 2. The figure presents the estimated densities of the OLS estimator, while the tables report the actual rejection probabilities for the Wald and HC Wald tests. Our simulation results, both for the estimator and tests, are largely consistent with our asymptotic theories developed in the paper.

As we have shown in Section 3, the OLS estimator has nonzero asymptotic bias, if the HGF is asymptotically homogeneous and if there is endogeneity. For the model with integrable HGF's, in contrast, the OLS estimator is asymptotically unbised and therefore the finite sample bias should disappear in the limit. Our simulations indeed show that this is the case at least for the cointegrating regression, as is illustrated in Figure 3. The OLS estimator shows some significant biases for the models with HGF's in (a) and (b), which do not vanish as the sample size gets large. The bias for the OLS estimator for the model with HGF (c), however, diminishes as the sample size increases. For the stationary regression, the OLS has the largest bias for HGF (a). For the model with HGF in (d), the OLS estimator is clearly seen to be inconsistent, having increasing variances as the sample size grows. Given that the exponential HGF generates highly explosive volatilities, the inconsistency of the OLS estimator is well expected from our results in Section 3.

The standard Wald statistic does not have the usual chi-square limiting null distribution in all cases we consider here. Our simulation models have both heteroskedasticity and endogeneity, in which case the standard Wald test has nonstandard asymptotics. This was shown earlier in Section 3. The test based on the standard Wald statistic is therefore expected to yield the rejection probabilities that are different from the nominal sizes. Our simulation results demonstrate that this discrepancy can be very large. For the models with the HGF's in (a) and (b), the standard Wald test over- or under-reject the null hypothesis substantially. The actual rejection rates are extremely large, when the HGF is given by (b) with explosive powers in the cointegrating regression. They are, however, much smaller than the nominal rates under similar situations for the stationary regression. If the HGF is given by (c), the behavior of the Wald statistic differs substantially depending upon the underlying model. For the stationary regression, the Wald test tends to over-reject as for the models with the HGF's (a) and (b). However, for the cointegrating regression, the Wald test severely under-rejects the null hypothesis, and the problem deteriorates as the sample size increases. This quite well coroborates our theory, which predicts that the Wald statistic would converge in probability zero in this case.

The HC Wald test corrects for heteroskedasticity. When the HGF is given by (c), only the correction for heteroskedasticity is required for the Wald statistic, and the HC Wald statistic indeed has asymptotic chi-square distribution. This is shown clearly in our simulations, especially for the cointegrating regression. When the HGF is given by either (a) or (b), we also need the correction for endogeneity for the chi-square limit theory, both in the stationary and cointegrating regressions. Therefore, the HC Wald statistic in this case does not have chi-squre limiting distribution. Yet, the HC Wald test sometimes effectively downsizes the the actual rejection probabilities, when the over-rejections of the Wald test are rather drastic as in the cointegrating regression with HGF (b).

6. Conclusions

In this paper, we consider the regression models with NNH in the errors and develop their asymptotic theories. The class of models analyzed in the paper is quite broad: In particular, it includes both the stationary and cointegrating regressions with NNH generated by integrable and asymptotically homogeneous HGF's. Though we present as illustrations only three different regression models that fit well into our framework in the paper, there seems to be plenty of examples for such regressions. Indeed we feel that many of the time series regressions commonly run by the researchers and practioners in the fields of economics and finance have at least some NNH features in them. In other words, they appear to have conditional heteroskedasticity, at least some part of which is affected by one or more nonstationary integrated processes. The theories developed in the paper would thus be widely relevant for many practical applications in empirical economics and finance.

Strictly speaking, our analysis of the regression models with NNH in the errors is not quite complete. We have only diagnosed the problems in such regressions. Our main diagnosis is simple and clear. The regressions become spurious if NNH in the errors is explosive, and even if it is mild, the OLS estimator is inefficient and the Wald test is not valid in general. In the former case, there is nothing we can do. The relationship represented by the regression is irrecoverably contaminated due to the presence of excessive NNH in the errors. The problems in the latter case, however, are rectifiable. Indeed, we may develop a new methodology to do more efficient and robust inference in regressions with mild NNH. The inefficiency of the OLS estimator and invalidity of the Wald test in the regressions with NNH are mostly due to the presence of heterogeneity and endogeneity. Therefore, we may follow Phillips and Hansen (1990) or Park (1992) to correct for heterogeneity and endogeneity by modifying the OLS estimator and the Wald statistic, and obtain the efficient estimator and valid chi-square test. The research along this line is now underway, and will be reported in a subsequent work by the authors.

Mathematical Appendix

In what follows, we let $M_n = n^{-1}\sum_{t=1}^n x_t x_t'$ for the stationary regressor (x_t) and $W_n = n^{-2}\sum_{t=1}^n x_t x_t'$ for the integrated regressor (x_t). These notations will be used repeatedly in the subsequent proofs.

Proof of Theorem 3.1 To prove the first part, we let

$$\frac{\sqrt{n}}{\nu(\sqrt{n})}\left(\hat{\beta}-\beta\right) = M_n^{-1}\left(\frac{1}{\sqrt{n}\,\nu(\sqrt{n})}\sum_{t=1}^n \sigma(z_{t-1})x_t\varepsilon_t\right).$$

The first term converges to M^{-1} by the Assumption 2.1(a). For the second term, we may use the Cramer-Wald device and extend Theorem 3.3 of Park and Phillips(2001) to deduce that

$$\frac{1}{\sqrt{n}\,\nu(\sqrt{n})}\sum_{t=1}^n \sigma(z_{t-1})x_t\varepsilon_t \to_d \int_0^1 \tau(V(r))dU(r)$$

as was to be shown.

We now prove the result in the second part. Under H_0, we may write

$$F(\hat{\beta}) = Q_n/P_n, \tag{10}$$

where

$$\begin{aligned} P_n &= \frac{1}{n\,\nu^2(\sqrt{n})}\sum_{t=1}^n u_t^2 - \frac{1}{n}\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right]' M_n^{-1}\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right], \\ Q_n &= \left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right]' R'\left(R\,M_n^{-1}\,R'\right)^{-1} R\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right]. \end{aligned}$$

It follows directly from the first part and the continuous mapping theorem that

$$Q_n \to_d N' M^{-1} R' \left(R M^{-1} R' \right)^{-1} R M^{-1} N. \tag{11}$$

Moreover, we have

$$\begin{aligned} P_n &= \frac{1}{n\,\nu^2(\sqrt{n})} \sum_{t=1}^{n} u_t^2 + O_p(n^{-1}) \\ &= \frac{1}{n\,\nu^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1})\,\varepsilon_t^2 + O_p(n^{-1}) \\ &= \frac{1}{n\,\nu^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1})\,\sigma_\varepsilon^2 + O_p(n^{-1/2}) \\ &\to_d \int_0^1 \tau^2(V(r))dr. \end{aligned} \tag{12}$$

Note that $\sigma^2(s)$ is asymptotically homogeneous with asymptotic order $\nu^2(\lambda)$ and limit homogeneous function $\tau^2(s)$. Therefore, we may deduce from Park and Phillips (2001, Theorem 3.3) that

$$\frac{1}{n\,\nu^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1})(\varepsilon_t^2 - \sigma_\varepsilon^2) = O_p(n^{-1/2}),$$

since $(\varepsilon_t^2 - \sigma_\varepsilon^2)$ is a martingale difference sequence and the fourth order moment of (ε_t) is bounded by Assumption 2.1(e). Furthermore,

$$\frac{1}{n\,\nu^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \to_d \int_0^1 \tau^2(V(r))\,dr,$$

due to Park and Phillips(2001, Theorem 3.3). The second part is now immediate from (11) and (12).

To prove the third part, we need to obtain the asymptotics for $F(\hat{\beta})$ under H_1. Let $R\beta \neq r$, and write Q_n in (10) as

$$\begin{aligned} Q_n &= \frac{\sqrt{n}}{\nu(\sqrt{n})} \Big(R(\hat{\beta} - \beta) + (R\beta - r) \Big)' \left(R\, M_n^{-1}\, R' \right)^{-1} \frac{\sqrt{n}}{\nu(\sqrt{n})} \Big(R(\hat{\beta} - \beta) + (R\beta - r) \Big) \\ &= \left[\frac{\sqrt{n}}{\nu(\sqrt{n})} (\hat{\beta} - \beta) \right]' R' \left(R\, M_n^{-1}\, R' \right)^{-1} R \left[\frac{\sqrt{n}}{\nu(\sqrt{n})} (\hat{\beta} - \beta) \right] \\ &\quad + 2 \frac{\sqrt{n}}{\nu(\sqrt{n})} (R\beta - r)' \left(R\, M_n^{-1}\, R' \right)^{-1} R \left[\frac{\sqrt{n}}{\nu(\sqrt{n})} (\hat{\beta} - \beta) \right] \\ &\quad + \frac{\sqrt{n}}{\nu(\sqrt{n})} (R\beta - r)' \left(R\, M_n^{-1}\, R' \right)^{-1} (R\beta - r) \frac{\sqrt{n}}{\nu(\sqrt{n})}. \end{aligned}$$

The second and third terms are of orders $O_p(\sqrt{n}/\nu(\sqrt{n}))$ and $O_p(n/\nu^2(\sqrt{n}))$, respectively, and both terms vanish if $\nu(\sqrt{n})/\sqrt{n} \to \infty$. Therefore, $F(\hat{\beta})$ has the same limiting distribution as under H_0 in this case. If $\nu(\sqrt{n})/\sqrt{n} \to 0$, both terms go to infinity, and $F(\hat{\beta}) \to_p \infty$. This completes the proof.

Proof of Theorem 3.2 To prove the first part,we let

$$n^{3/4}\left(\hat{\beta}-\beta\right)=M_n^{-1}\left(\frac{1}{n^{1/4}}\sum_{t=1}^{n}\sigma(z_{t-1})x_t'\varepsilon_t\right).$$

Using the Cramer-Wold device, we may easily extend Park and Phillips(2001, Theorem 3.2) to obtain

$$\frac{1}{n^{1/4}}\sum_{t=1}^{n}\sigma(z_{t-1})x_t\varepsilon_t\to_d S^{1/2}Z,$$

where

$$S=L(1,0)\int_{-\infty}^{\infty}\sigma^2(s)\,ds$$

and Z is distributed as $\mathbb{N}(0,\Sigma)$ independently of $L(1,0)$. The stated result therefore follows immediately if we let $N=S^{1/2}Z$.

We now prove the second part. Under H_0, we write $F(\hat{\beta})$ as in (10), where

$$\begin{aligned}P_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t^2-\frac{1}{n}\left[n^{3/4}(\hat{\beta}-\beta)\right]'M_n\left[n^{3/4}(\hat{\beta}-\beta)\right],\\ Q_n &= \left[n^{3/4}(\hat{\beta}-\beta)\right]'R'\left(R\,M_n^{-1}\,R'\right)^{-1}R\left[n^{3/4}(\hat{\beta}-\beta)\right],\end{aligned}$$

similarly as (10). Once again, it follows directly from the first part and the continuous mapping theorem that

$$Q_n\to_d Z'S^{1/2}M^{-1}R'\left(RM^{-1}R'\right)RM^{-1}S^{1/2}Z.$$

Moreover, we have

$$\begin{aligned}P_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t^2+O_p(n^{-1})\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma^2(z_{t-1})+O_p(n^{-1/2})\to_d S.\end{aligned}$$

Note that $\sigma^2(s)$ is integrable, and therefore, we have from Park and Phillips (2001, Theorem 3.2) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma^2(z_{t-1})\to_d L(1,0)\int_{-\infty}^{\infty}\sigma^2(s)\,ds$$

and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma^2(z_{t-1})(\varepsilon_t^2-1)=O_p(n^{-1/4}).$$

The second part now follows immediately.

To prove the third part, note that we have under H_1 that

$$\begin{aligned} Q_n &= n^{3/4}\left(R(\hat\beta-\beta)+(R\beta-r)\right)'\left(RM_n^{-1}R'\right)^{-1}n^{3/4}\left(R(\hat\beta-\beta)+(R\beta-r)\right)\\ &= \left[n^{3/4}(\hat\beta-\beta)\right]'R'\left(RM_n^{-1}R'\right)^{-1}R\left[n^{3/4}(\hat\beta-\beta)\right]\\ &\quad + 2\,n^{3/4}(R\beta-r)'\left(RM_n^{-1}R'\right)^{-1}R\left[n^{3/4}(\hat\beta-\beta)\right]\\ &\quad + n^{3/2}(R\beta-r)'\left(RM_n^{-1}R'\right)^{-1}(R\beta-r). \end{aligned}$$

The second and third terms are of orders $O_p(n^{3/4})$ and $O_p(n^{3/2})$, respectively, and diverge as $n\to\infty$. Therefore, $F(\hat\beta)\to_p\infty$ as was to be shown.

Proof of Theorem 3.3 We write

$$\frac{n}{\nu(\sqrt{n})}\left(\hat\beta-\beta\right)=W_n^{-1}\left(\frac{1}{n\,\nu(\sqrt{n})}\sum_{t=1}^n\sigma(\alpha'x_t)x_t\varepsilon_t\right).$$

It is well known that

$$W_n=\frac{1}{n^2}\sum_{t=1}^n x_tx_t'\to_d\int_0^1 V(r)V(r)'dr.$$

Moreover, since $\sigma(\alpha's)s$ is asymptotically homogeneous with asymptotic order $\nu(\lambda)\lambda$ and limit homogeneous function $\tau(\alpha's)s$, we may readily deduce that

$$\frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^n\sigma(\alpha'x_t)x_t\varepsilon_t\to_d\int_0^1\tau(\alpha'V(r))V(r)\,dU(r)$$

by applying Park and Phillips (2001, Theorem 3.3) and the Cramer-Wald device. The stated result in the first part may now be easily obtained.

We now prove the second part. Under H_0, we write $F(\hat\beta)$ as in (10), where

$$\begin{aligned} P_n &= \frac{1}{n\,\nu^2(\sqrt{n})}\sum_{t=1}^n u_t^2-\frac{1}{n}\left[\frac{n}{\nu(\sqrt{n})}(\hat\beta-\beta)\right]'W_n\left[\frac{n}{\nu(\sqrt{n})}(\hat\beta-\beta)\right],\\ Q_n &= \left[\frac{n}{\nu(\sqrt{n})}(\hat\beta-\beta)\right]'R'\left(R\,W_n^{-1}\,R'\right)^{-1}R\left[\frac{n}{\nu(\sqrt{n})}(\hat\beta-\beta)\right]. \end{aligned}$$

We may easily derive that

$$Q_n\to_d Z'W^{-1}R'\left(RW^{-1}R'\right)^{-1}RW^{-1}Z$$

using the continuous mapping theorem. Moreover, we have

$$P_n=\frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^n\sigma^2(z_{t-1})+O_p(n^{-1/2})\to_d\int_0^1\tau^2(\alpha'V(r))\,dr,$$

exactly as in the proof of Theorem 3.1. The stated result in the second part now follows immediately.

The proof of the third part is essentially identical to that of Theorem 3.1, and therefore, it is omitted.

Proof of Theorem 3.4 To prove the first part, we first rotate (x_t) using the orthogonal matrix H and define

$$H'x_t = \begin{pmatrix} h_1'x_t \\ H_2'x_t \end{pmatrix} = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}. \tag{13}$$

Then we have

$$D_n H' W_n(\hat{\beta} - \beta) = Z_n,$$

where

$$Z_n = \left(n^{-1/4} \sum_{t=1}^n x_{1t}\sigma(x_{1t})\varepsilon_t, n^{-3/4} \sum_{t=1}^n x_{2t}'\sigma(x_{1t})\varepsilon_t \right)',$$

and it follows from Lemma 1 in Chang and Park (2003) that $Z_n \to_d Z$ as was to be shown. Moreover, we have $W_n \to_d W$ as noted earlier in the proof of Theorem 3.3, and the joint convergence of W_n and Z_n is obvious from the proof of Lemma 1 in Chang and Park (2003).

We first consider the case $m = 1$. In this case, we write $F(\hat{\beta})$ as in (10) with

$$\begin{aligned} P_n &= n \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t^2 \right] - \frac{1}{n} \left[n^{7/4} W_n(\hat{\beta} - \beta) \right] W_n^{-1} \left[n^{7/4} W_n(\hat{\beta} - \beta) \right], \\ Q_n &= \left[n^{7/4} W_n(\hat{\beta} - \beta) \right] W_n^{-1} R \left(R W_n^{-1} R \right)^{-1} R W_n^{-1} \left[n^{7/4} W_n(\hat{\beta} - \beta) \right], \end{aligned}$$

under H_0. We may now easily deduce that $Q_n = O_p(1)$ and $P_n \to_p \infty$. Consequently, it follows that $F(\hat{\beta}) \to_p 0$ under H_0, as was to be shown. The proof for the asymptotics of $F(\hat{\beta})$ under H_1 is essentially identical to that of Theorem 3.2, and therefore, it is omitted.

For the more general case $m \geq 2$, we write $F(\hat{\beta})$ as in (10) with

$$\begin{aligned} P_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t^2 - \frac{1}{n} \left[n^{5/4} W_n(\hat{\beta} - \beta) \right]' W_n^{-1} \left[n^{5/4} W_n(\hat{\beta} - \beta) \right], \\ Q_n &= \left[n^{5/4} H D_n^{-1} Z_n \right]' W_n^{-1} R' \left(R W_n^{-1} R' \right)^{-1} R W_n^{-1} \left[n^{5/4} H D_n^{-1} Z_n \right], \end{aligned}$$

under H_0. It follows that

$$P_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t^2 + O_p(n^{-1}) \to_d L_1(1,0) \int_{-\infty}^{\infty} \sigma^2(s)\, ds \tag{14}$$

as shown in the proof of Theorem 3.2. Moreover, if we partition Z_n as $Z_n = (Z_{1n}, Z_{2n}')'$, then we have under H_0

$$\begin{aligned} Q_n &= Z_{2n}' H_2' W_n^{-1} R' \left(R W_n^{-1} R' \right)^{-1} R W_n^{-1} H_2 Z_{2n} \\ &\quad + 2n^{-1/2} \left[Z_{2n}' H_2' W_n^{-1} R' \left(R W_n^{-1} R' \right)^{-1} R W_n^{-1} h_1 Z_{1n} \right] \\ &\quad + n^{-1} \left[Z_{1n} h_1' W_n^{-1} R' \left(R W_n^{-1} R' \right)^{-1} R W_n^{-1} h_1 Z_{1n} \right] \\ &= Z_{2n}' H_2' W_n^{-1} R' \left(R W_n^{-1} R' \right)^{-1} R W_n^{-1} H_2 Z_{2n} + O_p(n^{-1/2}) \\ &\to_d Z_2' H_2' W^{-1} R' \left(R W^{-1} R' \right)^{-1} R W^{-1} H_2 Z_2. \end{aligned} \tag{15}$$

The asymptotics for $F(\hat{\beta})$ under H_0 is now immediate from (14) and (15).

Under H_1, on the other hand, we have

$$\begin{aligned} Q_n &= n^{5/2}\left[(R\beta - r)'\left(RW_n^{-1}R'\right)^{-1}(R\beta - r)\right] \\ &\quad + 2n^{5/4}\left[(R\beta - r)'\left(R\,W_n^{-1}R'\right)^{-1}RW_n^{-1}H_2Z_{2n}\right] \\ &\quad + 2n^{3/4}\left[(R\beta - r)'\left(RW_n^{-1}R'\right)^{-1}RW_n^{-1}h_1Z_{1n}\right] \\ &\quad + Z_{2n}'H_2'W_n^{-1}R'\left(RW_n^{-1}R'\right)RW_n^{-1}H_2Z_{2n} \\ &\quad + 2n^{-1/2}\left[Z_{2n}'H_2'W_n^{-1}R'\left(RW_n^{-1}R'\right)^{-1}RW_n^{-1}h_1Z_{1n}\right] \\ &\quad + n^{-1}\left[Z_{1n}h_1'W_n^{-1}R'\left(RW_n^{-1}R'\right)^{-1}RW_n^{-1}h_1Z_{1n}\right] \\ &\to_p \infty \end{aligned} \tag{16}$$

and the stated result follows readily from (14) and (16).

Proof of Theorem 4.1(i) Under H_0, we may write

$$G(\hat{\beta}) = \left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta} - \beta)\right]' R'\left(RM_n^{-1}\hat{Q}_nM_n^{-1}R'\right)^{-1}R\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta} - \beta)\right]$$

with

$$\hat{Q}_n = \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}u_t^2x_tx_t' - \left(R_{1n} + R_{1n}'\right) + R_{2n}, \tag{17}$$

where

$$\begin{aligned} R_{1n} &= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tu_t(\hat{\beta} - \beta)'x_tx_t', \\ R_{2n} &= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tx_t'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'x_tx_t'. \end{aligned}$$

We have

$$R_{1n} = \frac{1}{n}\left[\frac{1}{\sqrt{n}\,\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\varepsilon_tx_t\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta} - \beta)'x_tx_t'\right] = O_p(n^{-1}), \tag{18}$$

which follows immediately upon noticing that $(\varepsilon_tx_t \otimes x_tx_t')$ is a martingale difference sequence, due to Theorem 3.3 in Park and Phillips (2001). Moreover,

$$\|R_{2n}\| \leq \frac{1}{n}\left\|\frac{\sqrt{n}}{\nu(\sqrt{n})}(\hat{\beta} - \beta)\right\|^2\left(\frac{1}{n}\sum_{t=1}^{n}\|x_t\|^4\right) = O_p(n^{-1}). \tag{19}$$

Now we have from (17)–(19)

$$\begin{aligned}\hat{Q}_n &= \frac{1}{n\,\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})(x_t\varepsilon_t)(x_t\varepsilon_t)' + O_p(n^{-1}) \\ &= \frac{1}{n\,\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\Sigma + O_p(n^{-1/2}) \\ &\to_d \Sigma S,\end{aligned}$$

similarly as in the proof of Theorem 3.1. To deduce the stated result, use the continuous mapping theorem, and note that N is distributed as central normal mixture with mixing variate given by ΣS if U and V are independent.

Proof of Theorem 4.1(ii) Under H_0, we may write

$$G(\hat{\beta}) = \left[n^{3/4}(\hat{\beta}-\beta)\right]' R' \left(R\,M_n^{-1}\hat{Q}_n M_n^{-1}R'\right)^{-1} R\left[n^{3/4}(\hat{\beta}-\beta)\right],$$

where we have

$$\begin{aligned}\hat{Q}_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t^2x_tx_t' - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_tu_t(\hat{\beta}-\beta)'x_tx_t' \\ &\quad - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_tx_t'(\hat{\beta}-\beta)u_tx_t' + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_tx_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_tx_t' \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma^2(z_{t-1})(x_t\varepsilon_t)(x_t\varepsilon_t)' + O_p(n^{-1}) \\ &\to_d \Sigma S\end{aligned}$$

similarly as in the proof of Theorem 4.1(i). The stated result now follows immediately.

Proof of Theorem 4.1(iii) Under H_0, we have

$$G(\hat{\beta}) = \left[\frac{n}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right]' R' \left(RW_n^{-1}\hat{Q}_nW_n^{-1}R'\right)^{-1} R\left[\frac{n}{\nu(\sqrt{n})}(\hat{\beta}-\beta)\right]$$

with

$$\begin{aligned}\hat{Q}_n &= \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}u_t^2x_tx_t' - \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tu_t(\hat{\beta}-\beta)'x_tx_t' \\ &\quad - \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tx_t'(\hat{\beta}-\beta)u_tx_t' + \frac{1}{n^2\,\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tx_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_tx_t'.\end{aligned}$$

The rest of the proof is essentially identical to the proof of Theorem 4.1(i), except that $\sigma^2(\alpha's)ss'$ is asymptotically homogeneous with asymptotic order $\nu^2(\lambda)\lambda^2$ and limit homogeneous function $\tau^2(\alpha's)ss'$. The details are therefore omitted.

Proof of Theorem 4.1(iv) Let $m = 1$. Under H_0, we may write

$$G(\hat{\beta}) = \left[n^{7/4} W_n(\hat{\beta} - \beta)\right] W_n^{-1} R \left(R\, W_n^{-1} \hat{Q}_n\, W_n^{-1} R\right)^{-1} R W_n^{-1} \left[n^{7/4} W_n(\hat{\beta} - \beta)\right],$$

where we have

$$\begin{aligned} \hat{Q}_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t^2 x_t^2 - 2 \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t^3 u_t (\hat{\beta} - \beta) + \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t^4 (\hat{\beta} - \beta)^2 \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t^2 \sigma^2(x_t) + o_p(1) \\ &\to_d L(1,0) \int_{-\infty}^{\infty} s^2 \sigma^2(s)\, ds, \end{aligned}$$

from which the stated result follows easily. The rest of the proof is entirely analogous with the proof of Theorem 4.1(ii), and is therefore omitted.

We now let $m \geq 2$, and define (x_{1t}) and (x_{2t}) to be given as in (13). Under H_0, we may write

$$G(\hat{\beta}) = \left[n^{5/4} H D_n^{-1} Z_n\right]' W_n^{-1} R' \left(R W_n^{-1} H S_n H' W_n^{-1} R'\right)^{-1} R W_n^{-1} \left[n^{5/4} H D_n^{-1} Z_n\right],$$

where $Z_n = D_n H' W_n (\hat{\beta} - \beta)$ and

$$\begin{aligned} S_n &= n^{-3/2} \sum_{t=1}^n \hat{u}_t^2 H' x_t x_t' H \\ &= n^{-3/2} \sum_{t=1}^n u_t^2 H' x_t x_t' H - 2\, n^{-3/2} \sum_{t=1}^n u_t x_t' (\hat{\beta} - \beta) H' x_t x_t' H \\ &\quad + n^{-3/2} \sum_{t=1}^n x_t' (\hat{\beta} - \beta)(\hat{\beta} - \beta)' x_t H' x_t x_t' H. \end{aligned} \tag{20}$$

We may deduce that

$$\begin{aligned} n^{-3/2} \sum_{t=1}^n u_t^2 H' x_t x_t' H &= \begin{pmatrix} \frac{1}{n} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{1t}^2\right] & \frac{1}{\sqrt{n}} \left[\frac{1}{n} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{1t} x_{2t}'\right] \\ \frac{1}{\sqrt{n}} \left[\frac{1}{n} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{2t} x_{1t}\right] & \frac{1}{n^{3/2}} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{2t} x_{2t}' \end{pmatrix} \\ &= \begin{pmatrix} O_p(n^{-1}) & O_p(n^{-1/2}) \\ O_p(n^{-1/2}) & \frac{1}{n^{3/2}} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{2t} x_{2t}' \end{pmatrix} \end{aligned} \tag{21}$$

and

$$\frac{1}{n^{3/2}} \sum_{t=1}^n \sigma^2(x_{1t}) \varepsilon_t^2 x_{2t} x_{2t}' \to_d \int_{-\infty}^{\infty} \int_0^1 \sigma^2(s) V_2(r) V_2(r)' L(dr, 0) ds. \tag{22}$$

Moreover, if we let

$$T_n = (T'_{1n}\,,\,T'_{2n})' = n^{5/4}H'(\hat{\beta}-\beta) = W_n^{-1}\left(\frac{1}{\sqrt{n}}h_1 Z_{1n} + H_2 Z_{2n}\right) = O_p(1),$$

then we may easily derive that

$$\begin{aligned} & n^{-3/2}\sum_{t=1}^{n} u_t x'_t(\hat{\beta}-\beta)H'x_t x'_t H \\ &= n^{-11/4}\sum_{t=1}^{n}\sigma(x_{1t})\varepsilon_t\Big(x_{1t}T_{1n} + x'_{2t}T_{2n}\Big)\begin{pmatrix} x_{1t}^2 & x_{1t}x'_{2t} \\ x_{2t}x_{1t} & x_{2t}x'_{2t}\end{pmatrix} \\ &= \begin{pmatrix} O_p(n^{-2}) & O_p(n^{-3/2}) \\ O_p(n^{-3/2}) & O_p(n^{-1})\end{pmatrix}, \end{aligned} \tag{23}$$

and that

$$\begin{aligned} & n^{-3/2}\sum_{t=1}^{n} x'_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t H'x_t x'_t H \\ &= n^{-4}\sum_{t=1}^{n}(x_{1t}\ x'_{2t})T_n T'_n\begin{pmatrix} x_{1t} \\ x_{2t}\end{pmatrix}\begin{pmatrix} x_{1t}^2 & x_{1t}x'_{2t} \\ x_{2t}x_{1t} & x_{2t}x'_{2t}\end{pmatrix} \\ &= \begin{pmatrix} O_p(n^{-1}) & O_p(n^{-1}) \\ O_p(n^{-1}) & O_p(n^{-1})\end{pmatrix}. \end{aligned} \tag{24}$$

Now it follows from (20)–(24) that

$$HS_nH' \to_d H_2\left(\int_{-\infty}^{\infty}\int_0^1 \sigma^2(s)V_2(r)V_2(r)'L(dr,0)\,ds\right)H'_2.$$

To complete the proof, note that we have due to Theorem 3.4

$$n^{5/4}HD_n^{-1}Z_n \to_d \begin{pmatrix} 0 \\ H_2 Z_2 \end{pmatrix},$$

where $H_2 Z_2$ is distributed as central normal mixture with mixing variate

$$H_2\left(\int_{-\infty}^{\infty}\int_0^1 \sigma^2(s)V_2(r)V_2(r)'L(dr,0)\,ds\right)H'_2$$

and is independent of L_1.

References

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307-327.

Chang, Y., Park, J.Y., 2003. Index models with integrated time series,Journal of Econometrics 114, 73-106.

de Jong, R.M., 2002. Appendum to Asymptotics for nonlinear transformations of integrated time series,Mimeograph, Department of Economics, Michigan State University.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation,Econometrica 50, 987-1008.

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

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. (Academic Press, New York).

Hansen, B.E., 1995. Regression with nonstationary volatility, Econometrica 63, 1113-1132.

Miller, J.I., Park, J.Y., 2005. Nonlinearity, nonstationarity and thick tails: How they interact to generate persistency in memory, Mimeograph, Department of Economics, Rice University.

Park, J.Y., 1992. Canonical cointegrating regressions, Econometrica 60, 119-143.

Park, J.Y., 2002. Nonlinear nonstationary heteroskedasticity, Journal of Econometrics 110, 383-415.

Park, J.Y., 2003. Strong approximations for nonlinear transformations of integrated time series, Mimeograph, Department of Economics, Rice University.

Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series,Econometric Theory 15, 269-298.

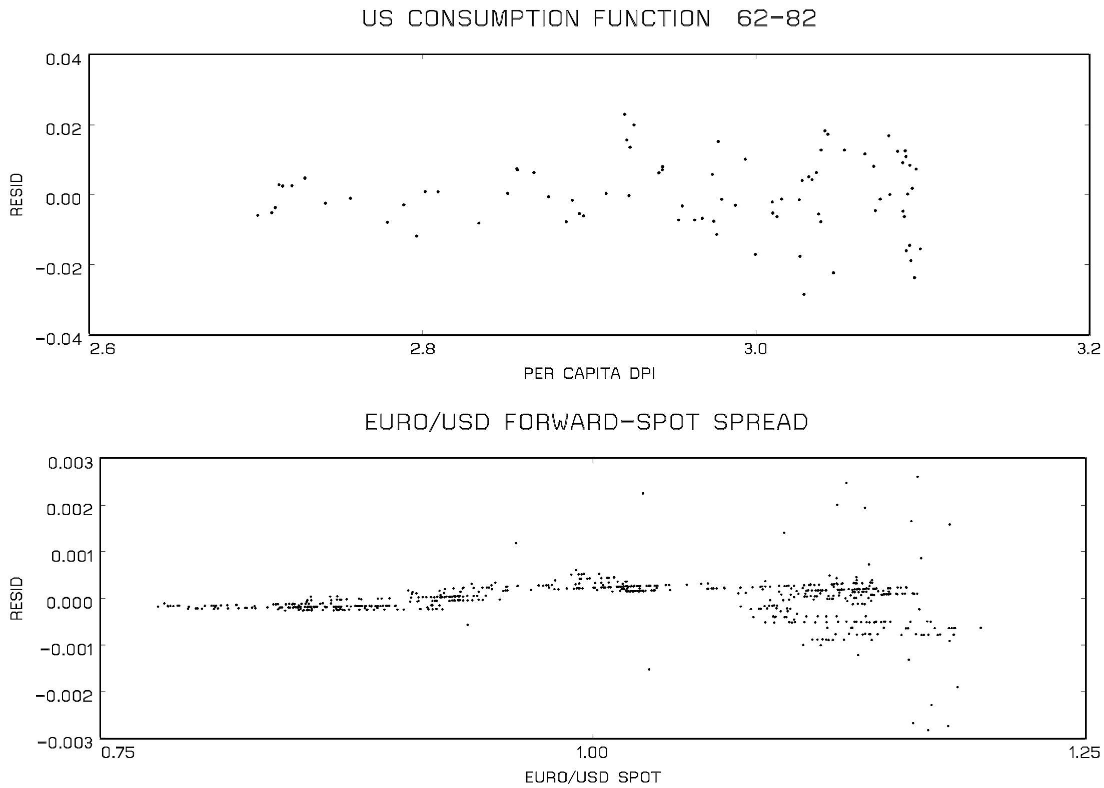
Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series, Econometrica 69, 117-161.

Phillips, P.C.B., 1986. Understanding spurious regression in econometrics, Journal of Econometrics 33, 311-340.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regressions with I(1) processes, Review of Economic Studies 57, 99-125.

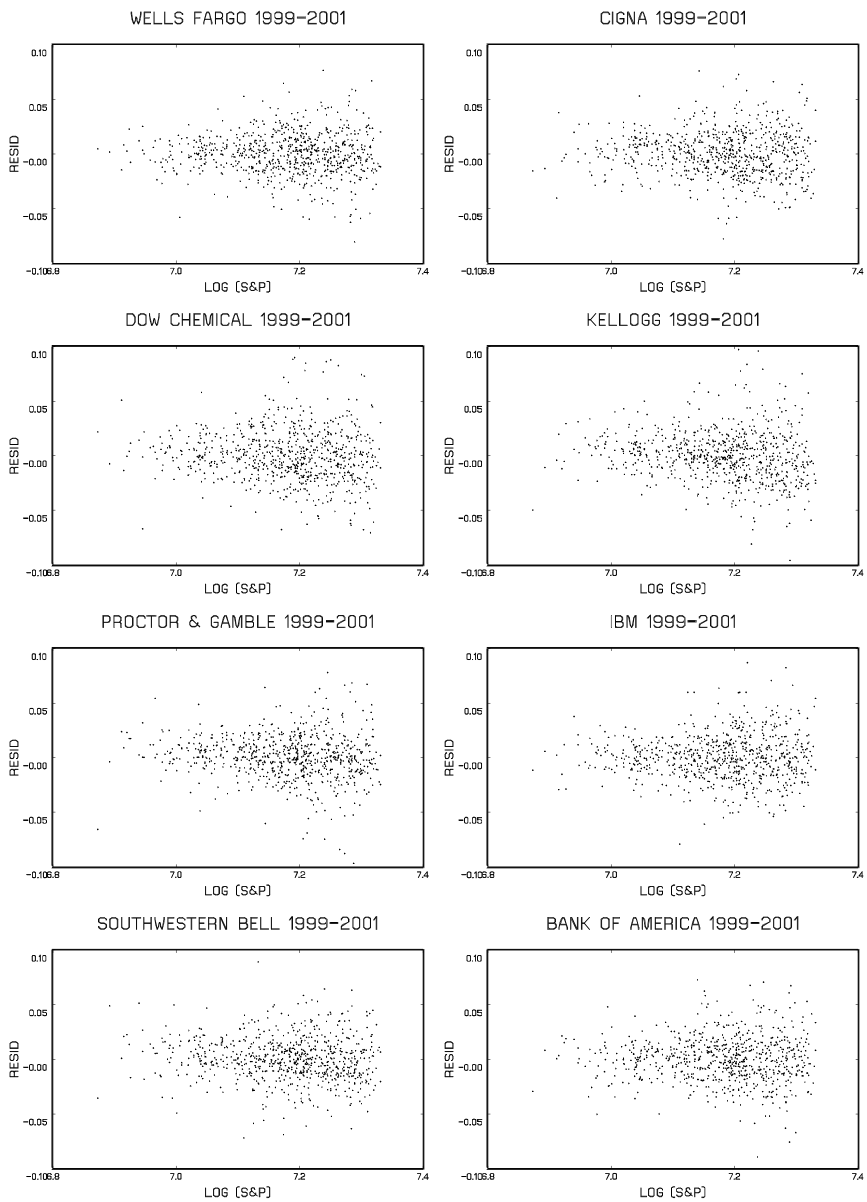
Shephard, N., 2005. Stochastic Volatility: Selected Readings. (Oxford University Press: Oxford).

Figure 1: Residuals from US Consumption Function and EURO/USD Forward-Spot Regressions



For the upper panel, we fit the regression for the US consumption function,$c_t = \alpha + \beta\, \iota_t + u_t$ with (c_t) and (ι_t) denoting respectively the log of per capita real personal consumption expenditure and disposal personal income, and plot the residuals $(\hat{u}_t)$ against (ι_t). Both variables used in the regression are seasonally adjusted, and the data are quarterly and collected for the period of 1962:1 – 1982:4. During this period, the parameter values α and β appear to be reasonably stable. The lower panel shows the EURO/USD forward-spot spreads, plotted against the EURO/USD spot rates. The underlying regression model is specified as $f_t = \alpha + \beta\, s_t + u_t$, where (s_t) and (f_t) are the daily spot and one month forward EURO/USD exchange rates. The data are obtained from DATASTREAM, for the period of 04/20/2001 – 01/19/2004.

Figure 2: Residuals from CAPM Regressions



In this figure, the residuals from the CAPM regression are plotted against the log level of the S&P 500 composite price index. The regression model fitted here is given by $(r_{it} - r_{ft}) = \beta\,(r_{mt} - r_{ft}) + u_t$, where we use the daily return for the individual stocks and the S&P 500 composite price index respectively for (r_{it}) and (r_{mt}), and the Fama-French risk free rate for (r_{ft}). The stock return data are collected from CRSP, for the period of 01/04/1999-12/31/2001. The period is selected arbitrarily for our illustration purpose as the selected period gives us relatively more clustered observations on the market.

Table 1: Rejection Probabilities of Wald Tests in Stationary Regression

	Wald			HC Wald		
			HGF: Integrable			
	1%	5%	10%	1%	5%	10%
$n = 100$	2.16	8.88	16.03	1.31	8.20	16.17
$n = 200$	2.13	8.85	15.86	1.27	8.08	15.82
$n = 500$	1.86	8.26	15.02	1.34	7.63	14.86
			HGF: Logistic			
	1%	5%	10%	1%	5%	10%
$n = 100$	1.29	5.77	11.24	1.59	6.74	12.62
$n = 200$	1.32	6.15	11.81	1.44	6.85	12.82
$n = 500$	1.36	6.67	12.82	1.44	7.11	13.64
			HGF: $\|z\|^{0.25}$			
	1%	5%	10%	1%	5%	10%
$n = 100$	0.95	4.41	8.72	1.28	5.09	9.78
$n = 200$	0.84	4.08	8.48	0.99	4.46	8.94
$n = 500$	0.79	4.15	8.37	0.88	4.25	8.62
			HGF: $\|z\|^{0.5}$			
	1%	5%	10%	1%	5%	10%
$n = 100$	0.78	3.84	7.86	1.04	4.48	8.80
$n = 200$	0.67	3.56	7.52	0.82	3.87	8.01
$n = 500$	0.64	3.52	7.49	0.69	3.71	7.66
			HGF: $\|z\|$			
	1%	5%	10%	1%	5%	10%
$n = 100$	0.62	3.48	7.44	0.75	3.83	8.02
$n = 200$	0.56	3.27	7.14	0.62	3.41	7.40
$n = 500$	0.51	3.15	6.96	0.53	3.24	7.13
			HGF: Exponential			
	1%	5%	10%	1%	5%	10%
$n = 100$	1.01	5.15	10.30	1.14	6.08	11.93
$n = 200$	1.12	5.74	11.34	1.07	6.34	12.61
$n = 500$	1.33	7.15	14.12	1.24	7.63	15.35

Table 2: Rejection Probabilities of Wald Tests in Cointegrating Regression

	Wald			HC Wald		
HGF: Integrable						
	1%	5%	10%	1%	5%	10%
$n = 100$	0.26	1.38	2.88	1.28	6.69	12.82
$n = 200$	0.10	0.60	1.37	1.18	6.21	12.19
$n = 500$	0.01	0.11	0.28	1.12	5.89	11.60
HGF: Logistic						
	1%	5%	10%	1%	5%	10%
$n = 100$	1.50	5.94	10.77	1.64	6.78	12.50
$n = 200$	1.50	5.75	10.54	1.46	6.49	12.36
$n = 500$	1.53	5.73	10.18	1.31	6.32	12.11
HGF: $\|z\|^{0.5}$						
	1%	5%	10%	1%	5%	10%
$n = 100$	6.87	16.20	23.87	1.93	7.44	13.42
$n = 200$	6.62	16.17	23.97	1.64	6.81	12.81
$n = 500$	6.46	15.85	23.58	1.42	6.40	12.15
HGF: $\|z\|$						
	1%	5%	10%	1%	5%	10%
$n = 100$	11.77	22.99	31.24	2.11	8.14	14.56
$n = 200$	11.44	22.86	31.19	1.77	7.35	13.81
$n = 500$	11.26	22.63	30.88	1.52	6.93	13.29
HGF: $\|z\|^2$						
	1%	5%	10%	1%	5%	10%
$n = 100$	18.30	31.26	39.78	2.30	8.99	16.36
$n = 200$	18.09	31.10	39.59	1.91	8.40	15.49
$n = 500$	17.90	30.91	39.46	1.78	8.11	15.26
HGF: Exponential						
	1%	5%	10%	1%	5%	10%
$n = 100$	17.78	30.64	37.78	0.78	6.40	15.85
$n = 200$	21.02	34.27	41.22	0.82	7.43	17.79
$n = 500$	24.57	37.68	44.44	1.05	8.88	20.32

Figure 3: Densities of OLS estimators

Stationary regression | Cointegrating regression