Journal of Economic Theory and Econometrics, Vol. 24, No. 3, Sep. 2013, 256–305

Efficient Estimation of Regressions with Nonstationary Heteroskedasticity*

Heetaik Chung† Chang Sik Kim‡

Abstract In this paper, we develop an efficient estimation method and an asymptotic chi-square testing procedure in regression models with errors having conditional heteroskedasticity generated by an integrated covariate in both stationary regression and cointegrating regression. In the presence of nonstationary volatility in the regression errors, it is known that the least squares estimator suffers from the second order biases generated by heteroskedasticity and endogeneity, and the standard chi-square test becomes invalid. It is shown that the efficient estimator proposed in the paper is asymptotically unbiased and follows a mixed normal distribution, and the test based on the estimator has chi-square limit distribution. Finite sample performances also confirm our theoretical findings.

Keywords Nonstationary Volatility, Nonstationary Nonlinear Heteroskedasticity, Heteroskedasticity Generating Function, Feasible Estimation

JEL Classification C22

---

*This research was supported by Handong Global University Research Grants 2010, and also supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012-S1A5A2A01020498).

†School of Management and Economics, Handong University, E-mail: htchung@handong.edu

‡Corresponding author, Department of Economics, Sungkyunkwan University, E-mail: skimcs@skku.edu.

Received April 9, 2013, Revised August 14, 2013, Accepted August 29, 2013

1. INTRODUCTION

Recently, much attention has been paid to nonstationary heteroskedasticity models since Hansen (1995) specifies the error variance as a continuous function of a nearly nonstationary AR process. In particular, a significant theoretical progress has been made to the unit root models with various types of time-varying error variances and heteroskedasticity, for example, Cavaliere (2004), Boswijk (2005), Cavaliere and Taylor (2007), Cavaliere, Rahbek and Taylor (2010), Kim and Park (2010), and Cavaliere, Phillips, Smeekes and Taylor (2012) among many others. They considered unit root models with various types of time varying volatilities. With a similar modeling of time varying heterogeneity, Phillips and Xu (2006) investigate the stationary autoregressive model.

In this paper, we consider a different type of heteroskedasticity in a regression model in the sense that the heteroskedasticity is generated by nonlinear transformation of nonstationary variables possibly including its own regressors in the model. This type of heteroskedasticity was originally introduced by Park (2002), and Chung and Park (2007) recently developed the statistical theories for the general time series regression models with nonstationary nonlinear heteroskedasticity (NNH), i.e., the models with errors having conditional heteroskedasticity generated by a nonlinear function of an integrated covariate. As they demonstrate in the paper, such regressions seem to be abundant in practical time series analysis. For instance, we may naturally expect that the time series regression of consumption function has errors having the conditional heteroskedasticity given as a function of income level. Moreover, for the finance models such as the capital asset pricing model (CAPM) and the forward rate unbiasedness may well be thought of having conditionally heteroskedastic errors that evolve with the market levels and the spot exchange rates, respectively. These examples are well illustrated in Chung and Park (2007).

The regressions with NNH in the errors have statistical features that are very distinct from the usual stationary and cointegrating regressions. Most notably, the presence of NNH in the errors may induce spuriousness in the sense of Granger and Newbold (1974) and Phillips (1986). As shown in Chung and Park (2007), if the NNH is given by an explosive function and an excessive volatility is introduced to the errors, the OLS estimator indeed becomes inconsistent and nonsensical. This type of spuriousness can also be found in nonparametric cointegrating regressions as suggested in Kim and Kim (2010). This, of course, is a rather extreme case, which we do not expect to encounter frequently in the usual practical applications. In many empirically relevant cases, we observe relatively mild NNH and may still estimate the underlying regressions consistently using OLS. However, even in this case, the OLS estimator is in general biased and inefficient asymptotically. To be more specific, OLS estimates with NNH errors are asymptotically second order biased in the sense that their limit distributions are not centered around the true parameter or shifted away from the true parameter just like the linear cointegrating regression models in the literature, for example, Phillips and Hansen (1990) and Park (1992). The second order bias in the presence of long run endogeneity still remains in the limit distribution of OLS estimator, and this second order biases are not necessarily eliminated even when the sample size grows large. Furthermore, the standard Wald statistic does not have chi-square limit null distribution, and the usual chi-square test based on the Wald statistic becomes invalid. Even when there is no long run endogeneity problem, the Wald test remains invalid in the regression with NNH because of the heteroskedasticity as shown in Chung and Park (2002).

In this paper, we develop an efficient method of estimation and a chi-square test applicable for the regressions with mild NNH in the errors. They are defined through some simple modifications of the OLS estimator and the standard Wald test, using a method similar to the canonical cointegrating regression methodology by Park (1992) and the fully-modified OLS approach by Phillips and Hansen (1990). They are based on the preliminary fit for the underlying model to estimate the nuisance parameters, which are then used to construct the required modifications. The newly proposed OLS estimator, called the efficient least squares (ELS) estimator, has a reduced asymptotic variance and yields a mixed normal limit distribution. Moreover, the modified Wald test based on the ELS estimator, which we refer to as the heteroskedasticity and endogeneity-corrected (HEC) Wald test, is asymptotically free of nuisance parameters and has chi-square limit distribution.

Through an extensive set of simulations, we investigate the finite sample performances of the proposed procedures and compare their performances with those of the OLS estimator and the standard Wald test. The new methodology works very well in finite samples, even when the sample size is relatively small. The ELS estimator effectively removes both the finite sample and asymptotic biases, and significantly reduces the finite sample variability of the OLS estimator. Likewise, the HEC Wald test performs extremely well, yielding the actual rejection probabilities under the null hypothesis that are very close to their nominal values. This is true for both the stationary and cointegrating regressions and with various specifications of the HGF's in the errors. It appears that our new methodology provides quite reliable tools to make inference in a wide range of the regressions with mild NNH in the errors.

The rest of the paper is organized as follows. The model and assumptions are presented in Section 2. More specifically, the stationary and cointegrating regressions with NNH in the errors are introduced with a set of regularity conditions, and then we provide some results for the regressions with NNH errors established in Chung and Park (2007). The discussion of their findings is included here to make the paper self-expository and to provide the motivations for the estimators and test statistics developed in the paper. In particular, 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 3, we introduce a new methodology to resolve these problems, and develop an effective method of inference in our regression models with NNH in the errors. Here we assume that the form of NNH is known. Section 4 subsequently extends the methodology in Section 3 to the practically more relevant case, where the form of NNH is unknown and has to be estimated. Section 5 presents the simulation results. 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, and (u_t) is the regression error that is further modeled as

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

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

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

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

Under such specifications, we have

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

and our model (1) becomes the regression model with errors having NNH, introduced by Park (2002) and explored later in Chung and Park (2007). The specification of the given model is the same as theirs, and the regressors (x_t) are either stationary or nonstationary integrated. The regression (1) becomes the stationary regression with NNH in the errors if the regressors are stationary, and 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

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

i.e., the conditional orthogonality holds between the regressor and the regression error as in Chung and Park (2007). This assumption is fairly general since it requires only $\mathbf{E}(x_t \varepsilon_t | \mathscr{F}_{t-1}) = 0$ in the stationary regression. In the cointegration model, we follow the assumption in Park and Phillips (2001) that (x_t) is predetermined with respect to the filtration $(\mathscr{F}_t)$, i.e., (x_t) is $(\mathscr{F}_{t-1})$-measurable, which is stronger than the stationary case. It, however, is commonly used in the analysis of nonstationary nonlinear model in the literature. We allow the regressors to be correlated with the z_{t-1} in both stationary and cointegration cases. 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. In this paper we will consider the class of *asymptotically homogeneous functions* which were first introduced by Park and Phillips (1999) in the asymptotic analysis of nonlinear nonstationary time series. Following their notation, it will be denoted by $\mathbb{H}$.

Definition 2.1. *We let $\sigma \in \mathbb{H}$ 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 class $\mathbb{H}$ include a wide class of transformations defined on $\mathbb{R}$. In particular, it is relevant in many econometric analyses in the sense that the variance of regression errors in time series regression are likely to be correlated with the absolute levels of other covariates. We refer readers to Park and Phillips (1999, 2001) and Park (2002) for the detailed discussion of function classifications and its empirical relevance. Now denote $w_t = z_t$ in the stationary regression and $w_t = x_t$ in the cointegrating regression. We then assume the linear process with proper summability conditions in the innovation of the w_t as

$$\triangle w_t = C(L)\eta_t = \sum_{i=0}^{\infty} C_i \eta_{t-i}, \tag{4}$$

where (η_t) are a sequence of independent and identically distributed (iid) random variables (vectors) with regularity conditions in the definition 2.2 in Chung and Park (2007).

We now present two sets of assumptions, depending upon whether the regressor (x_t) in regression (1) is stationary or nonstationary. They are the same set of assumptions in Chung and Park (2007) to ensure the asymptotics for nonlinear nonstationary transformations developed in Park and Phillips (1999, 2001) and nonlinear nonstationary heteroskedasticity in Park (2002). Throughout the paper, we use as usual the notation $\|\cdot\|$ to denote the Euclidean norm.

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), *that is,* $\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} x_t\varepsilon_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \triangle z_t\right) \to_d (U,V)$ *where* $[nr]$ *is the integer part of* nr *for* $r \in [0,1]$,

(c) $(x_t\varepsilon_t, \mathscr{F}_t)$ *is a martingale difference sequence such that* $(1/n)\sum_{t=1}^{n} \mathbf{E}(\varepsilon_t^2 x_t x_t' | \mathscr{F}_{t-1}) \to_p \Sigma$ *and* $\sup_{t\geq 1}\mathbf{E}(\|x_t\varepsilon_t\|^4 | \mathscr{F}_{t-1}) < \infty$ *a.s., and*

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

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, that is,* $\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \varepsilon_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \triangle x_t\right) \to_d (U,V)$ *where* $[nr]$ *is the integer part of* nr *for* $r \in [0,1]$,

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

(d) (x_t) *is adapted to* $(\mathscr{F}_{t-1})$.

Condition (a) is one possible formulation for the variable that generates the heteroskedasticity in our model, not essential in our theoretical development in this paper, yet it is very plausible specification in many economics applications as mentioned before. The martingale difference requirements in above two assumptions and a strong assumption (d) in the cointegrating regression model are imposed to develop the asymptotics for our models that include nonstationary nonlinearity. A detailed discussion can be found in Chung and Park (2007).

For the development of the efficient method of inference proposed in the paper, we also need additional assumptions on the innovation process of an integrated variable z_t that generates the heteroskedasticity.

Assumption 2.3. *Let* $(\triangle w_t)$ *be a linear process defined in (4) and assume that*
(a) $C(z)$ *is bounded and bounded away from zero for* $|z| \leq 1$*, and*
(b) if we write $C(z)^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i z^i$*, then* $k^p \sum_{i=k+1}^{\infty} \pi_i^2 < \infty$ *for some* $p \geq 9$*.*

To implement the efficient estimator in practice, we will have to use the residual from the preliminary regression. Assumption 2.3 above is necessary to establish the asymptotics of the efficient method using those residuals.[1]

To explain the intuition of the paper more clearly, we need to discuss about the main result of Chung and Park(2007). They established the asymptotic theory for the model given by (1) – (3) and investigated 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. It is shown that the existence of NNH changes both the convergence rate and the limit distribution of OLS estimator, whereas the heterogeneity given as a function of stationary variates does not change the convergence rate.

In the stationary regression case, the convergence rate of the OLS estimator is $\sqrt{n}/\nu(\sqrt{n})$, which means that the OLS estimator is consistent only when the asymptotic order $\nu(\lambda)$ of HGF satisfies the condition $\nu(\lambda)/\lambda \to 0$ as $\lambda \to \infty$. If $\nu(\lambda) = \lambda^k$, the OLS estimator is consistent when $k < 1$. If not, the OLS estimator is not consistent and the regression is meaningless. This result may be called as a spuriousness regression in the sense of Granger and Newbold (1974) and Phillips (1986) but in more general definition. When the heterogeneity is strong such as $\nu(\lambda)/\lambda \to \infty$ as $\lambda \to \infty$, the Wald test statistic also becomes inconsistent.

Even when the OLS estimator is consistent, the Wald test becomes invalid in the classical sense because the limit distribution is nonstandard and depends upon various nuisance parameters. However, there is a special case where the limit distribution of the Wald test statistic becomes chi-square as shown in Chung and Park (2007). We present those two conditions below for the ease exposition of subsequent analysis. The first condition is that

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

where U and V are the limit Brownian motions appearing in assumption 2.1 (b) above. As is well known, they become independent when the long run correlation of $(x_t \varepsilon_t)$ and $(\triangle z_t)$ vanishes. The second condition is given by

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

[1]In fact, the assumption was used in Chang et al. (2001) for the theoretical development of efficient estimation in the cointegrated models with deterministic trend.

where M and Σ are introduced respectively in (a) and (c) above in assumption 2.1. It holds if (x_t) is predetermined with respect to the filtration $(\mathscr{F}_t)$, as we have

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

When $(x_t x_t')$ and (ε_t^2) are conditionally uncorrelated, the condition is likely to be met.

In the cointegrating regression model with NNH errors, the convergence rate of the OLS estimator is $n/\nu(\sqrt{n})$, so if $\nu(\lambda) = \lambda^\kappa$ with some $\kappa \geq 2$, the OLS estimator becomes inconsistent. The excessive and strong volatility would make the cointegration regression spurious, precisely as in the case of the stationary regression. 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 has a faster convergence rate in the cointegrating regression. The OLS estimator is consistent when there is mild heterogeneity, and hence the Wald test becomes consistent. It, however, has nonstandard limiting distribution, and nuisance parameters are present in a very complicated manner just like stationary regression case.

Even though Chung and Park (2007) has made a significant theoretical development for the OLS asymptotics and the Wald test statistics in NNH model as discussed above, the results are somewhat of limited in a practical point of view as they explicitly pointed out. Usual chi-square asymptotics for the Wald type test can only be valid in very limited cases when two conditions (5) and (6) are met in case of stationary regression or when the condition (5) are satisfied in case of the cointegrating regression. However, the condition (5) are very unlikely met in many empirical applications and time series econometric models in the sense that the regressors and the regression error variance are often very closely related even in the limit. Given that asymptotic homogeneous HGF functions are more relevant in many practical works as mentioned before, it is desirable to develop an estimation method and asymptotic chi-square test statistics that can be used in more empirically relevant time series models.

3. EFFICIENT LEAST SQUARES AND TEST STATISTICS

As was shown in Chung and Park (2007), the OLS estimator and the Wald statistic have nonstandard limit theories. In general, the OLS estimator is nonnormal, and it is asymptotically biased and indeed inefficient. 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, and its limiting distribution generally depends on various nuisance parameters. The nonstandard limit theories and the nuisance parameter problems for the OLS estimator and the Wald test in the regressions with NNH are mostly due to the presence of heterogeneity and endogeneity. In this section, we develop the methods to correct for heterogeneity and endogeneity by modifying the OLS estimator and the Wald statistic, and obtain the efficient estimators and valid chi-square tests.

3.1. EFFICIENT LEAST SQUARES

Now we propose a new methodology to do more efficient and robust inference in regressions with NNH in the errors. The estimators developed here are generally more efficient than the OLS estimators for all models considered in the paper, i.e., both the stationary and cointegrating regressions with the NNH generated by asymptotically homogeneous HGF. Moreover, the test statistics developed here have limiting chi-square distributions. Our procedure is closely related to the CCR methodology by Park (1992) and the fully-modified OLS approach by Phillips and Hansen (1990), which were later extended by Chang et al. (2002) and Chang and Park (2003) to obtain the efficient estimator and robust test in the nonlinear regression models with integrated time series.

In this section, we assume that the HGF σ is known. It is only to develop the asymptotic theory in a simpler setup. When σ is unknown, σ will be replaced by its consistent estimate $\hat{\sigma}$ as will be shown in the next section. Certainly, we would need a more restrictive set of technical conditions. The case with unknown HGF will be discussed further in the following section.

To estimate our models more efficiently and construct the chi-square tests, we first define

$$\hat{\varepsilon}_t = \frac{\hat{u}_t}{\sigma(z_{t-1})} \tag{7}$$

where $(\hat{u}_t)$ are the fitted OLS residuals. As denoted in the previous section, $\triangle w_t = \triangle z_t$ in the stationary regression or $\triangle w_t = \triangle x_t$ in the cointegrating regression. We then run the regression

$$\triangle w_t = \hat{\pi}_1 \triangle w_{t-1} + \cdots + \hat{\pi}_p \triangle w_{t-p} + \hat{\eta}_{p,t} \tag{8}$$

to obtain the fitted innovations $(\hat{\eta}_{p,t})$ of $(\triangle w_t)$, whose data generating mechanism is given by (4). In regression (8), we are required to increase p as $n \to \infty$ at an appropriate rate. The reader is referred to Chang and Park (2003, p84) for the detailed discussions on this scheme and for the explicit expansion rate for p.

For the stationary regression with NNH, we define

$$\tilde{\beta}_S = \hat{\beta} - \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2} \sum_{t=1}^{n} \sigma(z_{t-1}) \hat{\eta}_{p,t} \tag{9}$$

where

$$\hat{\sigma}_{x\varepsilon,\eta} = \frac{1}{n} \sum_{t=1}^{n} x_t \hat{\varepsilon}_t \hat{\eta}_{p,t} \quad \text{and} \quad \hat{\sigma}_\eta^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_{p,t}^2$$

and, for the cointegrating regression with NNH,

$$\tilde{\beta}_N = \hat{\beta} - \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \sum_{t=1}^{n} x_t \sigma(z_{t-1}) \hat{\sigma}_{\varepsilon,\eta} \widehat{\Sigma}_\eta^{-1} \hat{\eta}_{p,t} \tag{10}$$

where

$$\hat{\sigma}_{\varepsilon,\eta} = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t \hat{\eta}_{p,t}' \quad \text{and} \quad \widehat{\Sigma}_\eta = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_{p,t} \hat{\eta}_{p,t}'$$

using $(\hat{\varepsilon}_t)$ and $(\hat{\eta}_{p,t})$ defined in (7) and (8), respectively. The newly introduced estimators $\tilde{\beta}_S$ and $\tilde{\beta}_N$ are efficient, as we will show below, and for this reason they will be referred to as the efficient least squares estimator (ELS) in the paper. We assume that (z_t) are observable, which in turn implies that α is known in our model for the cointegrating regression with NNH. The assumption, however, is not crucial. If α is unknown, we may replace it by any of its consistent estimators. Our subsequent theory would not be affected by the replacement of α by its consistent estimator.[2]

The following theorem derives the limiting distributions of the ELS estimator for each of the stationary and cointegrating regressions with NNH driven by asymptotically homogeneous HGF. In what follows, we use the notation

$$\Sigma_* = \Sigma - \frac{\sigma_{x\varepsilon,\eta} \sigma_{x\varepsilon,\eta}'}{\hat{\sigma}_\eta^2} \quad \text{and} \quad \sigma_*^2 = \sigma_\varepsilon^2 - \hat{\sigma}_{\varepsilon,\eta} \widehat{\Sigma}_\eta^{-1} \hat{\sigma}_{\varepsilon,\eta}'$$

where $\sigma_{x\varepsilon,\eta} = \mathbf{E}(x_t \varepsilon_t \eta_t)$ and $\sigma_{\varepsilon,\eta} = \mathbf{E}(\varepsilon_t \eta_t')$. Recall that we set $\sigma_\varepsilon^2 = 1$ for the identification of HGF. We first look at the case of stationary regressions.

[2]For instance, α can be consistently estimated by the nonlinear regression of the squared fitted residuals on (x_t) with a properly specified σ. We may also treat σ nonparametrically, and use a method to nonparametrically estimate the index model.

Theorem 3.1. *Let $\sigma \in \mathbb{H}$, and let Assumption 2.1 (a)–(c) and Assumption 2.3 hold. We have*

$$\sqrt{n}\nu(\sqrt{n})^{-1}(\tilde{\beta}_S - \beta) \to_d M^{-1}N_*$$

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

$$\Sigma_* \int_0^1 \tau^2(V(r))\,dr.$$

As for the cointegrating regressions, we have the following results.

Theorem 3.2. *Let $\sigma \in \mathbb{H}$, and let Assumption 2.2 and Assumption 2.3 hold. We have*

$$n\nu(\sqrt{n})^{-1}(\tilde{\beta}_N - \beta) \to_d W^{-1}Z_*$$

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

$$\sigma_*^2 \int_0^1 \tau^2(\alpha' V(r))V(r)V(r)'\,dr.$$

In all cases, the ELS estimator has asymptotically central normal mixture distribution and therefore it is asymptotically unbiased. More importantly, it has smaller asymptotic variance, compared to the OLS estimator even when the latter also has asymptotic central normal mixture distribution. Note that Σ_* is the conditional variance of $(x_t\varepsilon_t)$ given (η_t), and likewise, σ_*^2 is the conditional variance of (ε_t) conditional on (η_t).

3.2. HETEROSKEDASTICITY AND ENDOGENEITY CORRECTED TEST STATISTICS

In the previous section, we developed an efficient least squares under NNH errors, and it is efficient in the sense that it has neither nonstandard nonnormal nor non-centered limit distribution. The efficient least squares estimator, however, only correct the endogeneity problem that causes the asymptotic distribution of OLS to be nonnormal and asymptotically biased. [3] Since NNH is still present in the regression errors even with the efficient estimator, it is natural

[3] Both endogeneity and heteroskedasticy corrected efficient estimator can be developed, but theoretically it becomes too complicated, and skipped in the present paper.

to suggest the White type heteroskedasticity-corrected Wald test. The motivation for the White correction in Wald test statistic should be the same as in the classical regression model with heteroskedastic errors. Chung and Park (2007) suggested the White corrected Wald test, but it is only working with independence between V and U, i.e., when either the long run correlation of $(x_t\varepsilon_t)$ and (Δz_t) goes to zero in the stationary regression or the long run correlation of (ε_t) and (Δx_t) in the cointegrating regression goes to zero. It can be, however, very limited because the regression error is very likely correlated with the variables that generate the heteroskedasticity. Especially if we consider the $z_t = \alpha' x_t$ in the cointegration regression, the asymptotic long run uncorrelatedness between (ε_t) and (Δx_t) cannot be held. Therefore, the following heteroskedasticity and endogeneity corrected White type Wald test can be very general and useful in many regressions with NNH.

We now define the heteroskedasticity and endogeneity-corrected (HEC) Wald statistic respectively for the stationary regression and cointegrating regression. In what follows, we let

$$\tilde{u}_t = y_t - x_t'\tilde{\beta},$$

where $\tilde{\beta} = \tilde{\beta}_S$ or $\tilde{\beta}_N$, depending upon whether the underlying regression is stationary or nonstationary. [4] For the usual stationary regression, we define

$$G(\tilde{\beta}_S) = \left(R\tilde{\beta}_S - r\right)' \left(R\left(\sum_{t=1}^n x_t x_t'\right)^{-1}\left(\sum_{t=1}^n (x_t u_t)^* (x_t u_t)^{*\prime}\right)\left(\sum_{t=1}^n x_t x_t'\right)^{-1} R'\right)^{-1} \left(R\tilde{\beta}_S - r\right),$$

where

$$(x_t u_t)^* = x_t \tilde{u}_t - \sigma(z_{t-1}) \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2} \hat{\eta}_{p,t}.$$

On the other hand, for the nonstationary cointegrating regression, we define

$$G(\tilde{\beta}_N) = \left(R\tilde{\beta}_N - r\right)' \left(R\left(\sum_{t=1}^n x_t x_t'\right)^{-1}\left(\sum_{t=1}^n u_t^{*2} x_t x_t'\right)\left(\sum_{t=1}^n x_t x_t'\right)^{-1} R'\right)^{-1} \left(R\tilde{\beta}_N - r\right),$$

where

$$u_t^* = \tilde{u}_t - \sigma(z_{t-1}) \hat{\sigma}_{\varepsilon,\eta} \widehat{\Sigma}_\eta^{-1} \hat{\eta}_{p,t}.$$

The following theorem shows that our HEC Wald statistic generally yields chi-square limiting distributions.

---

[4]We may use the fitted OLS residuals $(\hat{u}_t)$, in place of $(\tilde{u}_t)$ here. This will not change our subsequent results. Obviously, however, it would be better to use the fitted residuals $(\tilde{u}_t)$, since they use the more efficient estimator.

Theorem 3.3. *For the stationary regression, we have*

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

if the conditions in Theorem 3.1 and Assumption 2.1 (d) hold. And for the cointegrating regression, we have

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

if the conditions in Theorem 3.2 hold.

The HEC Wald statistic thus has chi-square distribution for all models considered in the paper, including the stationary and cointegrating regressions with NNH in the errors driven by asymptotically homogeneous HGF. In particular, the HEC Wald test is valid for the models with NNH generated by the asymptotically homogeneous HGF's without conditions in (5) and (6). This is because the HEC Wald statistic corrects for endogeneity, as well as for heterogeneity as mentioned above.

4. FEASIBLE ESTIMATION WITH AN ESTIMATED HGF

When HGF is unknown, the HGF needs to be estimated for the actual implementation. In this section we develop a method to estimate HGF consistently and efficiently, and establish the limit distribution of the estimated HGF. We then study the asymptotics of the ELS estimator and HEC Wald statistic constructed using estimated HGF. Clearly, our feasible estimation procedure works only when the mild heteroskedasticity with NNH exists. For example, if the HGF function is given by a simple asymptotically homogeneous function like $\sigma(z) = |z|^\gamma$, our feasible estimator becomes inconsistent when $\gamma \geq 1$ for the stationary regressions or $\gamma \geq 2$ for the nonstationary cointegrating regressions just like OLS estimators in both regressions.

Here we only consider the parametric estimation of HGF function, i.e., the functional form of HGF is known, and therefore we need to estimate parameters in the given HGF function to make the efficient estimator feasible. However, as mentioned before, we may also treat σ nonparametrically, and use a method to nonparametrically estimate the index model, which is left to the subsequent research.

4.1. ESTIMATION OF HGF

Let the square of the HGF $\sigma(\cdot)$ in (2) be specified in a parametric form

$$\sigma^2(z_{t-1}) = g(z_{t-1}, \theta_0) \tag{11}$$

with some known function g. Since the HGF $\sigma(\cdot)$ is an asymptotically homogeneous functions, we have

$$g(\lambda s) = \kappa(\lambda)h(s) + o(\kappa(\lambda))$$

for large λ uniformly in s over any compact interval, where h is locally Riemann integrable. Clearly, the asymptotic order κ should be ν^2, but we denote κ and h as the asymptotic order and limit homogeneous function of g to avoid the notational complications. We may then write

$$u_t^2 = g(z_{t-1}, \theta_0) + \zeta_t, \tag{12}$$

where

$$\zeta_t = \sigma^2(z_{t-1})(\varepsilon_t^2 - 1). \tag{13}$$

The nonlinear regression with integrated regressor with conditionally heteroskedastic errors such as the one in (12) and (13) has been studied by Park (2002). He suggested Weighted Nonlinear Least Squares estimator θ_n^* of the model (12) defined as

$$\theta_n^* = \operatorname*{argmin}_{\theta\in\Theta} Q_n^*(\theta). \tag{14}$$

where

$$Q_n^*(\theta) = \sum_{t=1}^{n} \frac{\left(u_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \theta_0)^2}, \tag{15}$$

Our approach in this paper is to make the nonlinear least square estimator in (15) feasible. That is, least squares residual $(\hat{u}_t)$ is used instead of (u_t). For the actual implementation we replace θ_0 with a consistent estimator for θ_0 such as the ordinary NLS estimator $\hat{\theta}_n$ of (12) where $\hat{u}_t^2$ is the regressand.[5]

If we let

$$\tilde{Q}_n(\theta) = \sum_{t=1}^{n} \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \hat{\theta}_n)^2}, \tag{16}$$

then the NLS estimator $\tilde{\theta}_n$ is defined as the minimizer of $\tilde{Q}_n(\theta)$ over $\theta \in \Theta$, i.e.,

$$\tilde{\theta}_n = \operatorname*{argmin}_{\theta\in\Theta} \tilde{Q}_n(\theta). \tag{17}$$

The asymptotic distribution of the NLS estimator $\tilde{\theta}_n$ is derived as in the standard nonlinear regression theory. We require conditions on the regression function that ensure it is sufficiently smooth as a function of the unknown parameter θ.

[5] Park and Phillips(2001) showed the asymptotics of the NLS estimator when the errors are homoskedastic.

Define

$$\dot{g} = \left(\frac{\partial g}{\partial \theta_i}\right),\ \ddot{g} = \left(\frac{\partial^2 g}{\partial \theta_i \partial \theta_j}\right), \dddot{g} = \left(\frac{\partial^3 g}{\partial \theta_i \partial \theta_j \partial \theta_k}\right),$$

to be all vectors, arranged by the lexicographic ordering of their indices. $\ddot{G}$ is the second derivatives of g in matrix form and $\ddot{g}$ is obtained from $\ddot{G}$ by stacking its rows into a column vector. Also $(\dot{\kappa}, \ddot{\kappa}, \dddot{\kappa})$ and $(\dot{h}, \ddot{h}, \dddot{h})$ are the asymptotic orders and the limit homogeneous functions of $(\dot{g}, \ddot{g}, \dddot{g})$ respectively.

Let $\dot{\tilde{Q}}_n$ and $\ddot{\tilde{Q}}_n$ be the first and second derivatives of $\tilde{Q}_n$ with respect to θ defined in the usual way i.e., $\dot{\tilde{Q}}_n = \partial \tilde{Q}_n / \partial \theta$ and$\ddot{\tilde{Q}}_n = \partial^2 \tilde{Q} / \partial \theta \partial \theta'$. We have

$$\dot{\tilde{Q}}_n(\theta) = -\sum_{t=1}^{n} \frac{\dot{g}(z_{t-1}, \theta)(\hat{u}_t^2 - g(z_{t-1}, \theta))}{g(z_{t-1}, \hat{\theta}_n)^2},$$

$$\ddot{\tilde{Q}}_n(\theta) = \sum_{t=1}^{n} \frac{\dot{g}(z_{t-1}, \theta)\dot{g}(z_{t-1}, \theta)'}{g(z_{t-1}, \hat{\theta}_n)^2} - \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1}, \theta)}{g(z_{t-1}, \hat{\theta}_n)^2}(\hat{u}_t^2 - g(z_{t-1}, \theta))$$

ignoring a constant, which is unimportant. The asymptotic distribution of $\tilde{\theta}_n$ can be obtained from the first order Taylor expansion of $\dot{\tilde{Q}}_n$, which is written as

$$\dot{\tilde{Q}}_n(\tilde{\theta}_n) = \dot{\tilde{Q}}_n(\theta_0) + \ddot{\tilde{Q}}_n(\theta_n)(\tilde{\theta}_n - \theta_0), \tag{18}$$

where θ_n lies in the line segment connecting $\tilde{\theta}_n$ and θ_0. We have $\dot{\tilde{Q}}_n(\tilde{\theta}_n) = 0$ if $\tilde{\theta}_n$ is an interior solution to the minimization problem.

Let $\dot{g}$ be asymptotically homogeneous. For an appropriately chosen normalizing sequence υ_n, it follows immediately from the sample covariance asymptotics that $\upsilon_n^{-1}\dot{\tilde{Q}}_n(\theta_0) \to_d \dot{Q}(\theta_0)$ for some random vector $\dot{Q}(\theta_0)$. Also, if we let

$$\ddot{Q}_n^0(\theta_0) = \sum_{t=1}^{n} \frac{\dot{g}(z_{t-1}, \theta_0)\dot{g}(z_{t-1}\theta_0)'}{g(z_{t-1}, \theta_0)^2},$$

then $\upsilon_n^{-1}\ddot{Q}_n^0(\theta_0)\upsilon_n^{-1\prime} \to_p \ddot{Q}(\theta_0)$ for some random matrix $\ddot{Q}(\theta_0)$, due to Lemma A6 and sample mean asymptotics in Theorem 3.3 in Park and Phillips(2001).

Therefore, under the following suitable conditions that ensure $\upsilon_n^{-1}\ddot{\tilde{Q}}_n(\theta_n)\upsilon_n^{-1\prime} =$

$\upsilon_n^{-1}\ddot{Q}_n^0(\theta_0)\upsilon_n^{-1\prime} + o_p(1)$ and $\ddot{Q}(\theta_0) > 0$ a.s., we may expect from (18) that

$$\begin{aligned}\upsilon_n'(\tilde{\theta}_n - \theta_0) &= -\left(\upsilon_n^{-1}\ddot{\tilde{Q}}_n(\theta_n)\upsilon_n^{-1\prime}\right)^{-1}\upsilon_n^{-1\prime}\dot{\tilde{Q}}_n(\theta_0) \\ &= -\left(\upsilon_n^{-1}\ddot{Q}_n^0(\theta_n)\upsilon_n^{-1\prime}\right)^{-1}\upsilon_n^{-1\prime}\dot{\tilde{Q}}_n(\theta_0) + o_p(1) \\ &\to_d -\ddot{Q}(\theta_0)^{-1}\dot{Q}(\theta_0) \qquad (19)\end{aligned}$$

as $n \to \infty$.

Park and Phillips (2001) specified the set of the following sufficient conditions that lead to (19), and hence we need to check those conditions to have asymptotic distributions of the weighted nonlinear least squares estimators in (17).

AD1: $\upsilon_n^{-1}\dot{\tilde{Q}}_n(\theta_0) \to_d \dot{Q}(\theta_0)$ as $n \to \infty$.

AD2: $\upsilon_n^{-1}\ddot{\tilde{Q}}_n(\theta_0)\upsilon_n^{-1\prime} = \upsilon_n^{-1}\ddot{Q}_n^0(\theta_0)\upsilon_n^{-1\prime} + o_p(1)$ for large n.

AD3: $\upsilon_n^{-1}\ddot{\tilde{Q}}_n(\theta_0)\upsilon_n^{-1\prime} \to_p \ddot{Q}(\theta_0)$ as $n \to \infty$.

AD4: $\ddot{Q}(\theta_0) > 0$ a.s.

AD5: There is a sequence μ_n such that $\mu_n \upsilon_n^{-1} \to_{a.s.} 0$ such that

$$\sup_{\theta \in N_n} \|\mu_n^{-1}(\ddot{\tilde{Q}}_n(\theta) - \ddot{\tilde{Q}}_n(\theta_0))\mu_n^{-1\prime}\| \to_p 0$$

where $N_n = \{\theta : \|\mu_n'(\theta - \theta_0)\| \leq 1\}$.

AD1-AD5 are the conditions for the case when the asymptotic order is dependent upon θ.[6] This type of asymptotically homogeneous function is more general such that it covers the type of function with a constant asymptotic order, and it is more relevant to have the function with the convergent rate depending upon θ since it covers much wider class of functions. Moreover, we do not need to presume consistency to derive the asymptotic distributions under these conditions.

We now present the asymptotic distribution of $\tilde{\theta}_n$. For notational brevity, we write $\dot{\kappa}_0(\cdot) = \dot{\kappa}(\cdot, \theta_0)$. To properly formulate a sufficient set of conditions for AD5, define a neighborhood of θ_0 by

$$N(\varepsilon, \lambda) = \{\theta : \left\| \frac{\dot{\kappa}_0(\lambda)'}{\kappa_0(\lambda)}(\theta - \theta_0) \right\| \leq \lambda^{-1+\varepsilon}\}$$

[6]The reader is referred to Park and Phillips (2001) for the asymptotics of the nonlinear estimators.

for $\varepsilon > 0$ given.

Theorem 4.1. *Let Assumption 2.1 (a)-(c) or Assumption 2.2 hold, and if*
(a) $\dot{g}/g \in \mathbb{H}$,
(b) for any $\bar{s} > 0$ *given, there exists* $\varepsilon > 0$ *such that as* $\lambda \to \infty$

$$\left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right) (\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \left| \frac{\ddot{g}(\lambda s, \theta_0)}{g^2(\lambda s, \theta_0)} \right| \right) \right\| \to 0, \quad (20)$$

$$\lambda^{-1+\varepsilon} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right) (\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\varepsilon, \lambda)} \left| \frac{\ddot{g}(\lambda s, \theta)}{g^2(\lambda s, \theta_0)} \right| \right) \right\| \to 0, \quad (21)$$

$$\lambda^{-1+\varepsilon} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right) (\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\varepsilon, \lambda)} \left| \frac{\dddot{g}(\lambda s, \theta)}{g^2(\lambda s, \theta_0)} \right| \right) \right\| \to 0; \quad (22)$$

$$\lambda^{-1} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0^2} \right) (\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \left| \frac{\dot{g}(\lambda s, \theta)}{g^2(\lambda s, \theta_0)} \right| \right) \right\| \to 0 \quad (23)$$

(c) $\int_{|s| \le \delta} \frac{\dot{h}(s, \theta_0) \dot{h}(s, \theta_0)'}{h^2(s, \theta_0)} ds > 0$ *for all* $\delta > 0$,

(d) $\frac{\sqrt{n}\, \dot{\kappa}_0(\sqrt{n})}{\kappa_0(\sqrt{n})} \to \infty$ *as* $n \to \infty$,
(e) $\frac{\sqrt{n}}{\kappa_0(\sqrt{n})} \to \infty$ *as* $n \to \infty$ *for the stationary regression or* $\frac{n}{\kappa_0(\sqrt{n})} \to \infty$ *as* $n \to \infty$ *for the cointegrating regression,*

Then

$$\frac{\sqrt{n}\, \dot{\kappa}_0(\sqrt{n})'}{\kappa_0(\sqrt{n})} (\tilde{\theta}_n - \theta_0) \to_d \left(\int_0^1 \frac{\dot{h}(V, \theta_0) \dot{h}(V, \theta_0)'}{h(V, \theta_0)^2} \right)^{-1} \int_0^1 \frac{\dot{h}(V, \theta_0)}{h(V, \theta_0)} dW$$

as $n \to \infty$.

Since OLS residual $\hat{u}$ is used for the consistent nonlinear estimation of the parameter θ, we impose Condition (e) that guarantees the consistency of OLS estimators in our model. Compared to the results in Park and Phillips (2001) and Park (2002), we need to impose an additional condition (23) to derive the asymptotics of the weighted nonlinear least squares. It is naturally expected so since we have weight function in the denominator to take the heteroskedasticity into account in the estimation. To control the behavior of this weight function when it is getting near zero, we have to limit the class of homogeneous functions in our consideration to get the asymptotic results. Hence the classes of functions we are considering are by no means exhaustive. However most of functions considered in the previous literature satisfy the condition including power functions with nonnegative powers, logistic functions, and all the other distribution function-like functions.

Park (2002) suggested the weighted least squares estimator in (14) and established the asymptotic results. However, the results of Theorem 4.1 are different from the previously developed asymptotics in two ways. First, our asymptotic results is clearly based on the residual $\hat{u}_t$ instead of unobservable u_t, and secondly we use $g(z_{t-1},\hat{\theta}_n)$ instead of $g(z_{t-1},\theta_0)$.

4.2. FEASIBLE ESTIMATION

We have defined the efficient estimator of our NNH regression models in section 3, the estimators in (9) and (10) are not feasible in that the true value of θ_0 and hence the NNH function $\sigma(z_{t-1})$ are assumed to be given, which is not the case in most of econometric models we are considering. It is, therefore, essential to define the feasible efficient estimator and develop the asymptotic theory for a host of empirical applications.

Once the consistent estimation of weight nonlinear least squares estimator $\tilde{\theta}_n$ is obtained, then we now can define the feasible efficient estimator of β just in the same way as in (9) and (10) using $\tilde{\theta}_n$. That is, we let

$$\hat{\varepsilon}_t^{\dagger} = \frac{\hat{u}_t}{\hat{\sigma}(z_{t-1})},$$

using estimated HGF where $\hat{\sigma}(z_{t-1}) = \sigma(z_{t-1},\tilde{\theta}_n)$, and then the feasible efficient estimator for the stationary regression with NNH can be defined as

$$\tilde{\beta}_S^{\dagger} = \hat{\beta} - \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_{\eta}^{2}} \sum_{t=1}^{n} \hat{\sigma}(z_{t-1})\hat{\eta}_{p,t}$$

where

$$\hat{\sigma}_{x\varepsilon,\eta}^{\dagger} = \frac{1}{n}\sum_{t=1}^{n} x_t \hat{\varepsilon}_t^{\dagger}\hat{\eta}_{p,t},$$

and, for the cointegrating regression with NNH,

$$\tilde{\beta}_N^{\dagger} = \hat{\beta} - \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t \hat{\sigma}(z_{t-1})\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\hat{\Sigma}_{\eta}^{-1}\hat{\eta}_{p,t}$$

where

$$\hat{\sigma}^{\dagger}_{\varepsilon,\eta} = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}^{\dagger}_t \hat{\eta}'_{p,t}.$$

The efficient feasible estimators are expected to have the same limiting distribution as given in Theorem 3.1 and 3.2 under suitable conditions. As for the limit distributions of the estimators $\beta^{\dagger}_S$ and $\beta^{\dagger}_N$, we have the following theorem.

Theorem 4.2. *Let Assumption 2.3 hold. If (i) conditions in Theorem 3.1 and Theorem 4.1 hold, and there exists $\varepsilon > 0$ such that as $\lambda \to \infty$ for any $\bar{s} > 0$ given,*

$$\lambda^{-1+\varepsilon} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0}\right)(\lambda)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N(\varepsilon,\lambda)} \left| \frac{\dot{g}(\lambda s,\theta)}{g(\lambda s,\theta_0)} \right| \right) \right\| \to 0, \tag{24}$$

then

$$\sqrt{n}\nu(\sqrt{n})^{-1}(\tilde{\beta}^{\dagger}_S - \beta) \to_d M^{-1}N_*$$

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

$$\Sigma_* \int_0^1 \tau^2(V(r))\,dr.$$

and if (ii) conditions in Theorem 3.2 and Theorem 4.1 and the additional condition in (24) hold, then

$$n\nu(\sqrt{n})^{-1}(\tilde{\beta}^{\dagger}_N - \beta) \to_d W^{-1}Z_*$$

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

$$\sigma^2_* \int_0^1 \tau^2(\alpha' V(r))V(r)V(r)'\,dr.$$

Compared to the results in Theorem 3.1-2, the conditions on the NNH functions are quite restrictive. This is because we are considering feasible estimation with estimated HGF rather than the efficient estimation with known true parameter values in HGF. As mentioned before, the commonly used homogeneous functions in nonlinear nonstationarity in the literature including power function with nonnegative powers and logistic functions are all included in the class of functions that satisfies the conditions imposed in Theorem 4.1 and 4.2.

The asymptotic results we have in Theorem 4.2 are the same as those of infeasible efficient estimator in the previous section, and we have central mixed normal distributions in both stationary and cointegration regressions. The estimators are asymptotically unbiased and more efficient as mentioned before, and the following simulation results support our theoretical findings.

We now define the feasible the heteroskedasticity and endogeneity-corrected (HEC) Wald statistic using estimated HGF that was introduced in the previous section. We let

$$\tilde{u}_t^\dagger = y_t - x_t' \tilde{\beta}^\dagger$$

where $\tilde{\beta}^\dagger = \tilde{\beta}_S^\dagger$ or $\tilde{\beta}_N^\dagger$, depending upon whether the underlying regression is stationary or nonstationary. Then we define $G(\tilde{\beta}_S^\dagger)$ and $G(\tilde{\beta}_N^\dagger)$ exactly the same way as in the previous section but using $\tilde{\beta}^\dagger$ and $\tilde{u}_t^\dagger$ instead of $\tilde{\beta}$ and $\tilde{u}_t$ such that

$$G(\tilde{\beta}_S^\dagger) = \left(R\tilde{\beta}_S^\dagger - r\right)' \left(R\,(\textstyle\sum_{t=1}^n x_t x_t')^{-1} (\sum_{t=1}^n (x_t u_t)^{**} (x_t u_t)^{**\prime})\,(\sum_{t=1}^n x_t x_t')^{-1} R'\right)^{-1} \left(R\tilde{\beta}_S^\dagger - r\right),$$

where

$$(x_t u_t)^{**} = x_t \tilde{u}_t^\dagger - \hat{\sigma}(z_{t-1}) \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2} \hat{\eta}_{p,t}$$

and

$$G(\tilde{\beta}_N^\dagger) = \left(R\tilde{\beta}_N^\dagger - r\right)' \left(R\,(\textstyle\sum_{t=1}^n x_t x_t')^{-1} \left(\sum_{t=1}^n u_t^{**2} x_t x_t'\right)(\sum_{t=1}^n x_t x_t')^{-1} R'\right)^{-1} \left(R\tilde{\beta}_N^\dagger - r\right)$$

where

$$u_t^{**} = \tilde{u}_t^\dagger - \hat{\sigma}(z_{t-1}) \hat{\sigma}_{\varepsilon,\eta}^\dagger \widehat{\Sigma}_\eta^{-1} \hat{\eta}_{p,t}$$

The following theorem shows that the HEC Wald statistic $G(\tilde{\beta}^\dagger)$ generally yields chi-square limiting distribution.

Theorem 4.3. *We have*

$$G(\tilde{\beta}^\dagger) \to_d \chi_q^2,$$

if one of the conditions (i) or (ii) in Theorem 4.2 holds.

The White type heteroskedasticity corrected test statistic based on the estimation of β and θ has very nice chi-square asymptotics as expected. In fact, the OLS based Wald test statistic is not valid in the classical sense since it has nonstandard limit distribution and does involve nuisance parameters in a very complicated way as mentioned before unless the specific extra conditions explained in (5) and (6) are met. With proper conditions, we can notice that the White correction indeed works well not only for the classical heteroskedasticity, but also for the complicated nonlinear nonstationary heteroskedasticity, and it also works for NNH generated by integrable functions as shown in Chung and Park (2007).

5. SIMULATION RESULTS

This section investigates the finite sample performances of the various estimators and test statistics considered in the paper. We examine both the stationary and cointegrating regressions with NNH. We choose the following asymptotically homogenous functions for the HGF.

$$(a)\ \ \sigma(z) = |z|^{\gamma},$$
$$(b)\ \ \sigma(z) = \frac{e^{\gamma z}}{1+e^{\gamma z}},$$

where $\gamma > 0$. As for the power function (a), we let the values of the parameter γ as $1/4, 1/2$ and 1 for the stationary regression, and $1/2, 1$ and 2 for the cointegrating regression. The function is homogeneous (and hence, asymptotically homogeneous) with asymptotic order $\nu(\lambda) = \lambda^{\gamma}$. Therefore, both of the OLS and ELS estimators are consistent if $\nu(\lambda)/\lambda \to 0$, that is, if $\gamma < 1$. The consistency condition for the cointegrating regression is $\nu(\lambda)/\lambda^2 \to 0$ and estimators are consistent as long as $\gamma < 2$. As for the logistic function (b) γ is given as 1 both for stationary and the cointegrating regression. The asymptotic order for the logistic function (b) is $\nu(\lambda) = 1$, and the limit homogeneous function is $\tau(s) = 1\{s \geq 0\}$. Since the asymptotic order is constant, the consistency condition is satisfied for the case of the logistic HGF.

As for the stationary regression, we generate $(\varepsilon_t), (x_t)$ and (z_t) 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. We deliberately set these variables to satisfy all the conditions in Assumption 2.1. In order to demonstrate the difference in the performance of between OLS and ELS and between the Wald and the HEC Wald statistic, we let neither (5) nor (6) is satisfied. The limit Brownian motions U and V in (b) have variances 4 and 2, respectively and the covariance 2. We also $M = \mathbf{E}x_t^2 = 2$ in (a), $\Sigma = 4$ in (c). In this setup, both heterogeneity and endogeneity are present for the stationary regression. As was shown by Chung and Park (2007), the OLS estimator becomes asymptotically inefficient and biased, and the standard Wald test is invalid.

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

$$\varepsilon_t = e_{1t} + e_{2t}$$
$$\triangle x_t = e_{1,t-1}$$
$$\triangle z_t = e_{1t}.$$

As in the stationary regression (e_{1t}) and (e_{2t}) are independent iid standard normals, and $(\varepsilon_t), (x_t)$ and (z_t) satisfy all the conditions in Assumption 2.2. We let $z_{t-1} = x_t$ in (a) and hence the limit Brownian motions U and V in (b) have variances 2 and 1, respectively with covariance 1. Therefore, both endogeneity and heterogeneity are also present in the cointegrating regression.

In the simulations, we look at the samples of sizes 100 and 500 in order to consider both the samples of moderate and relatively large sizes. The performances of the estimators and the tests are evaluated based on 100,000 iterations. To implement our methodology, we estimate HGF using the weighted NLS estimation method explained in the previous section and obtain the fitted innovation using the first order VAR on (Δz_t) for the stationary regression or on (Δx_t) for the cointegrating regression. The simulation results are summarized in Figures 1–2 and Tables 1–2. The estimated densities of the various estimators are presented in the figures, and the actual rejection probabilities for the tests are reported in the tables.

Our simulation results, both for the estimations and tests, are largely consistent with our asymptotic theories developed in the paper. In general, the OLS estimator is inefficient and biased. The ELS estimator proposed in the paper, on the other hand, has much smaller variances, and virtually no biases, compared to the OLS estimator. Our correction procedure seems to work very well, i.e., it significantly reduces the variability of the estimator and effectively corrects for the biases. Even for the samples of relatively small size, the ELS estimator performs rather well. Likewise, the HEC Wald tests are much more reliable than the standard Wald test which yields the rejection probabilities quite different from their nominal values. The HEC Wald test generally yields the rejection probabilities that are fairly close to the nominal sizes.

As mentioned before, the OLS estimator has nonzero asymptotic bias if the HGF is asymptotically homogeneous and if there is endogeneity. Our simulations clearly show some significant biases for both the stationary and cointegrating regression with logistic HGF. Moreover, those biases do not vanish as the sample size gets large. With power function, the bias of OLS estimator is clearly observed for the cointegrating regression, but not so obvious for the stationary regression. However, the ELS estimator has no noticeable bias, and is much more concentrated around the true parameter value both for the logistic and power function type HGF. When $\gamma = 1$ for the stationary regression, or $\gamma = 2$ for the cointegrating regression, the OLS estimators are inconsistent. In this case we cannot estimate the HGF consistently using $(\hat{u}_t)$. Therefore, direct comparison the performances between the ELS and the OLS does not seem to be meaningful. However, even in this case, the performance of ELS estimator dominates that of OLS. Especially, HEC Wald tests seem reliable. This is probably because the second order bias correction for endogeneity and heteroskedasticity improves the performance of estimator even when the HGF function cannot be consistently estimated.[7]

Since our simulation models have both heteroskedasticity and endogeneity, the standard Wald test has nonstandard limit distribution. Therefore, the test based on the standard Wald statistic is expected to yield the rejection probabilities that are different from the nominal sizes. However, the HEC Wald test is expected to correct for both heteroskedasticity and endogeneity and have no significant size distortion.

In the cointegrating regression, the standard Wald test over-reject the null hypothesis both for the logistic and for the power function type HGF, and the actual rejection rates are getting larger as the power of the HGF increases. This is quite expected because the distortion becomes more serious as the degree of the heterogeneity is getting stronger. The actual rejection probabilities of the HEC Wald test are very close to the nominal sizes compared to the performances of the standard Wald test. They are becoming closer to the exact nominal sizes as the sample size increases.

In the stationary regression, the actual rejection probabilities of the standard Wald test are smaller than the nominal size and getting smaller as the power of the HGF increases. The HEC Wald test, however, shows much less distortion and are getting closer to the nominal sizes as the sample size increases as in the cointegration regression. With logistic HGF, the rejection probabilities of the two Wald tests show significantly large discrepancy in the small sample, but as the sample size increases they are getting closer to the nominal sizes.

6. CONCLUSIONS

Chung and Park (2007) has demonstrated the empirical relevancy of the regression models with NNH in the errors and subsequently established the statis-

[7]This finding can be further developed in the future research. We thank a referee who raised an issue of the better performance of ELS estimator in this particular case.

tical theories for such models. The message they tried to deliver is twofold. First, there seems to be plenty of examples for such regressions, and many of the time series regressions that are commonly run by the researchers and practitioners in the fields of economics and finance have at least some NNH features in them. Second, we need new statistical procedures to effectively analyze such regressions. If relied on the standard OLS methodology, we get asymptotically biased and inefficient estimators and invalid tests.

In the paper, we propose an efficient method of estimation and a heteroskedasticity corrected asymptotic chi-square testing procedure, which we may use for the inference in the regressions with NNH in the errors. Our methodology is widely applicable for general time series regressions that are stationary or cointegrated. We also develop feasible methods to implement the proposed procedures and develop their asymptotic theories. The implementation of the proposed procedure simply requires that the OLS estimator and Wald test be modified using the nuisance parameter estimates that are obtained by the preliminary fit of the underlying model. We show that those feasible methods correct the asymptotic bias, improve the efficiency of the estimator, and make Wald test to have asymptotically chi-square.

The performance of this method in the finite sample was investigated using simulation. Simulation results are consistent with the asymptotic theories developed in the paper. The ELS estimator significantly reduces the variability and effectively corrects for the biases, and the HEC Wald tests yields rejection probabilities that are fairly close to chi-square distribution.

There are some issues need to be resolved. One is how to find the integrated covariate that generates NNH. The variable is likely to be correlated with regressors in the regression model, but it can be other variable outside of the regression model. Therefore, it may not be easy to find it in practice. More proper but computationally more demanding method is letting the covariate as a latent random walk process and extracting it from nonlinear Kalman filter. The research on this issue is left to the subsequent research of 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 stationary regressor (x_t) in our model.

Proof of Theorem 3.1. We let

$$\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S-\beta)=M_n^{-1}\left[\frac{1}{\sqrt{n}\,\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\,\hat{\eta}_{p,t}\right)\right].$$

The first term converges to M^{-1} by the Assumption 2.1(a). For the second term we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\,\hat{\eta}_{p,t}\right)\\
&=\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\,\eta_t\right)+o_p(1)\\
&\to_d \int_0^1\tau(V(r))d\{U(r)-\frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}C^{-1}(1)V(r)\}\\
&=\int_0^1\tau(V(r))dU_*(r),
\end{aligned}$$

using Theorem 10(b) of Chang et al. (2001). The stated result follows directly, upon noticing that V is independent of U_*. □

Proof of Theorem 3.2. We write

$$\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}_N-\beta)=W_n^{-1}\left[\frac{1}{n\,\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(\alpha' x_t)x_t\left(\varepsilon_t-\hat{\sigma}_{\varepsilon,\eta}\widehat{\Sigma}_\eta^{-1}\,\hat{\eta}_{p,t}\right)\right].$$

Note that the asymptotic order and the limit homogeneous function of $\sigma(\alpha' s)s$ is $\nu(\lambda)\lambda$ and $\tau(\alpha' s)s$. Then it can be easily deduced from Theorem 10(b) in Chang et al. (2001) that

$$\begin{aligned}
&\frac{1}{n\,\nu(\sqrt{n})}\sum_{t=1}^{n}x_t\sigma(\alpha' x_t)\left(\varepsilon_t-\hat{\sigma}_{\varepsilon,\eta}\widehat{\Sigma}_\eta^{-1}\,\hat{\eta}_{p,t}\right)\\
&=\frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(\alpha' x_t)x_t\left(\varepsilon_t-\sigma_{\varepsilon,\eta}\Sigma_\eta^{-1}\,\eta_t\right)+o_p(1)\\
&\to_d \int_0^1\tau(\alpha' V(r))V(r)d\{U(r)-\sigma_{\varepsilon,\eta}\Sigma_\eta^{-1}C^{-1}(1)V(r)\}\\
&=\int_0^1\tau(\alpha' V(r))V(r)dU_*(r).
\end{aligned}$$

and, given the independence of U_* and V, the stated result follows immediately. □

Proof of Theorem 3.3(i). Under H_0, we let

$$G(\tilde{\beta}_S) = \left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S - \beta)\right]' R' \left(R M_n^{-1} P_n^* M_n^{-1} R'\right)^{-1} R \left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S - \beta)\right],$$

where we may write

$$\begin{aligned}
P_n^* &= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}(x_t u_t)^*(x_t u_t)^{*\prime} \\
&= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\hat{\eta}_{p,t}\right)\left(x_t\varepsilon_t - \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\hat{\eta}_{p,t}\right)' \\
&\quad - (R_{1n} + R_{1n}') + R_{2n}.
\end{aligned}$$

It can be readily deduced from equations (50) and (66) in Chang et al. (2001) that

$$\begin{aligned}
&\frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\hat{\eta}_{p,t}\right)\left(x_t\varepsilon_t - \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\hat{\eta}_{p,t}\right)' \\
&= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\left(x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\right)\left(x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\right)' + o_p(1) \\
&= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\Sigma_* \\
&+ \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}\,\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\right. \\
&\left.\left(\left(x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\right)\left(x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\right)' - \Sigma_*\right)\right] + o_p(1) \\
&= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\Sigma_* + o_p(1) \\
&\to_d \Sigma_*\int_0^1 \tau^2(V(r))dr.
\end{aligned}$$

We may also easily establish that

$$\begin{aligned}
R_{1n} &= \frac{1}{n}\left[\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}_{x\varepsilon,\eta}}{\hat{\sigma}_\eta^2}\hat{\eta}_{p,t}\right)\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S-\beta)\right)' x_t x_t'\right] \\
&= \frac{1}{n}\left[\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\right)\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S-\beta)\right)' x_t x_t'\right] \\
&\quad + o_p(1) \\
&= o_p(1)
\end{aligned}$$

which follows immediately upon noticing that $\{x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma_\eta^2}\eta_t\} \otimes x_t x_t'$ is a martingale difference sequence, due to Theorem 3.3 in Park and Phillips(2001) and Theorem 3.1.

As for R_{2n} which is

$$R_{2n} = \frac{1}{n}\left[\frac{1}{n}\sum_{t=1}^{n}x_t x_t'\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S-\beta)\right)\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S-\beta)\right)' x_t x_t'\right],$$

we can establish

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

To deduce the stated result, use the continuous mapping theorem and Theorem 3.1 and notice U_* is central normal mixture with mixing variate $\Sigma_* S$ and that U_* and V are independent. □

Proof of Theorem 3.3(ii). Under H_0, we may write

$$G(\tilde{\beta}_N) = \left[\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}_N-\beta)\right]' R'\left(RW_n^{-1}P_n^*W_n^{-1}R'\right)^{-1}R\left[\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}_N-\beta)\right],$$

where

$$\begin{aligned}
P_n^* &= \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}u_t^{*2}x_t x_t' \\
&= \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_t x_t'^2(\alpha' x_t)\left(\varepsilon_t - \hat{\sigma}_{\varepsilon,\eta}\widehat{\Sigma}_\eta^{-1}\hat{\eta}_{p,t}\right)^2 - (R_{1n}+R_{1n}') + R_{2n}.
\end{aligned}$$

Following the similar argument in the proof of Theorem 3.3(i) we can establish

$$
\begin{aligned}
P_n^* &\to_d \sigma_*^2 \int_0^1 \tau^2(\alpha' V(r)) V(r) V(r)' dr \\
R_{1n} &= \frac{1}{n^3 \nu(\sqrt{n})} \sum_{t=1}^n \left(\sigma(z_{t-1}) \left(\varepsilon_t - \hat{\sigma}_{\varepsilon,\eta} \widehat{\Sigma}_\eta^{-1} \hat{\eta}_{p,t} \right) \left(\frac{n}{\nu(\sqrt{n})} (\tilde{\beta}_S - \beta) \right)' x_t \right) x_t x_t' \\
&= o_p(1), \\
R_{2n} &= \frac{1}{n^4} \sum_{t=1}^n \left(x_t' \left(\frac{n}{\nu(\sqrt{n})} (\tilde{\beta}_S - \beta) \right) \left(\frac{n}{\nu(\sqrt{n})} (\tilde{\beta}_S - \beta) \right)' x_t \right) x_t x_t' \\
&= O_p(n^{-1}).
\end{aligned}
$$

The stated result follows immediately from Theorem 3.2. □

Proof of Theorem 4.1. The proof of Theorem 4.1 has the following steps. We first set up the objective functions with OLS residual $\hat{u}_t$, and obtain the asymptotic distribution of the nonlinear least squares estimator of θ_0 in Lemma A.1 as follows.

$$\hat{\theta}_n^* = \underset{\theta \in \Theta}{\operatorname{argmin}}\, Q_n^*(\theta). \tag{25}$$

where

$$Q_n^*(\theta) = \sum_{t=1}^n \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \theta_0)^2}.$$

With the preliminary limiting behaviors based on (25), we then derive the asymptotic results of the weighted nonlinear least squares estimator based on the ordinary least squares estimator $\hat{\theta}_n$ based on the That is, the weighted NLS estimator $\tilde{\theta}_n$ is defined as

$$\tilde{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \tilde{Q}_n(\theta). \tag{26}$$

where

$$\tilde{Q}_n(\theta) = \sum_{t=1}^n \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \hat{\theta}_n)^2}.$$

The asymptotic distribution of $\tilde{\theta}_n$ in (26) both for the stationary and cointegration regressions will be explored.

The following lemma shows the asymptotics of nonlinear least squares estimator under asymptotically homogeneous HGF functions. The proof is based on Park and Phillips (2001), Park (2002) and Chang et al. (2001). However the proof is more complicated in the sense that we explicitly use the residual from the preliminary regression and heteroskedasticity correction in the objective function, and hence we need some additional restrictions on the nonlinear functions compared to the previous results. For the consistent estimation of OLS in the stationary and cointegration regressions, we need

(i) $\dfrac{\sqrt{n}}{\kappa_0^{1/2}(\sqrt{n})} \to \infty$ as $n \to \infty$ for the stationary regression,

and

(ii) $\dfrac{n}{\kappa_0^{1/2}(\sqrt{n})} \to \infty$ as $n \to \infty$ for the cointegrating regression,

respectively. We will first start with stationary regression, and the extension to the cointegrating regression will be almost the same and straightforward.

Lemma A.1. *Under the same assumptions (a)-(d) in the Theorem 5, we have*

$$\frac{\sqrt{n}\,\dot{\kappa}_0(\sqrt{n})'}{\kappa_0(\sqrt{n})}(\hat{\theta}_n^* - \theta_0) \to_d \left(\int_0^1 \frac{\dot{h}(V,\theta_0)\dot{h}(V,\theta_0)'}{h(V,\theta_0)^2}\right)^{-1} \int_0^1 \frac{\dot{h}(V,\theta_0)}{h(V,\theta_0)} dU$$

as $n \to \infty$.

Proof of Lemma A.1. **(i) Stationary Regression** : It suffices to show that AD1-AD5 are satisfied for the stated asymptotic distribution of nonlinear least squares estimator $\hat{\theta}_n^*$ defined in (25).

AD1: Note that

$$\hat{u}_t^2 - g(z_{t-1},\theta_0) = g(z_{t-1},\theta_0)\left(\varepsilon_t^2 - 1\right) - 2u_t x_t'(\hat{\beta}-\beta) + x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t$$

since

$$\hat{u}_t^2 = u_t^2 - 2u_t x_t'(\hat{\beta}-\beta) + (\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta),$$

then we have

$$\begin{aligned}
\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \dot{Q}_n^*(\theta_0) &= -\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^n \frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}(\hat{u}_t^2 - g(z_{t-1},\theta_0)) \\
&= -\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^n \frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)}(\varepsilon_t^2 - 1) \\
&\quad + \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^n \frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} \times \\
&\quad \left(2u_t x_t'(\hat{\beta}-\beta) - x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)'x_t\right).
\end{aligned} \tag{27}$$

By condition (a) and Theorem 3.3 in Park and Phillips (2001),

$$\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\sum_{t=1}^{n}\frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)}(\varepsilon_t^2-1)\to_p \int_0^1 \frac{\dot{h}(V,\theta_0)}{h(V,\theta_0)}dU.$$

Now, we need to show

$$\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\sum_{t=1}^{n}\frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}\left(2u_t x_t'(\hat{\beta}-\beta)-x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)'x_t\right)=o_p(1).$$

Ignoring the constant, we have

$$\begin{aligned}
&\left\|\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\sum_{t=1}^{n}\frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}u_t x_t'(\hat{\beta}-\beta)\right\|\\
&\le \sup_{|s|\le\bar{s}}\left\|\frac{\dot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)}\right\|\left\|\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\sum_{t=1}^{n}u_t x_t'(\hat{\beta}-\beta)\right\|\\
&= \sup_{|s|\le\bar{s}}\left\|\frac{\dot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)}\right\|\left\|(\hat{\beta}-\beta)\right\|\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\left\|\sum_{t=1}^{n}\sigma(z_{t-1})x_t\varepsilon_t\right\|\\
&\to 0,
\end{aligned}$$

since

$$\frac{\sqrt{n}}{\kappa_0^{1/2}}(\hat{\beta}-\beta)=O_p(1),\ \ \frac{1}{\sqrt{n}\kappa_0^{1/2}}\sum_{t=1}^{n}\sigma(z_{t-1})x_t\varepsilon_t=O_p(1)$$

from the Theorem 3.1 and the Theorem 3.3 of Park and Phillips (2001), and

$$\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0^2}\right)^{-1}\sup_{|s|\le\bar{s}}\left\|\frac{\dot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)}\right\|\to 0$$

from the assumption (23) as $n\to\infty$. Similarly, we have

$$\begin{aligned}
&\left\|\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1}\sum_{t=1}^{n}\frac{\dot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)'x_t\right\|\\
&= \sup_{|s|\le\bar{s}}\left\|\frac{\dot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)}\right\|\left\|\frac{\sqrt{n}}{\kappa_0^{1/2}}(\tilde{\beta}_S-\beta)\right\|^2\left(\frac{1}{n}\sum_{t=1}^{n}\|x_t\|^2\right).\\
&\to 0,
\end{aligned}$$

and therefore the condition AD1 is satisfied.

AD2: Since

$$\ddot{Q}_n^*(\theta_0) = \sum_{t=1}^{n} \frac{\dot{g}(z_{t-1},\theta_0)\dot{g}(z_{t-1},\theta_0)'}{g(z_{t-1},\theta_0)^2} - \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}(\hat{u}_t^2 - g(z_{t-1},\theta_0)),$$

we need to show

$$\begin{aligned}
&\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}(\hat{u}_t^2 - g(z_{t-1},\theta_0)) \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1'} \\
&= \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \left[\sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)}(\varepsilon_t^2 - 1)\right] \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1'} \\
&- 2\left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \left[\sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} u_t x_t'(\hat{\beta} - \beta)\right] \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1'} \\
&+ \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \left[\sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} x_t'(\hat{\beta} - \beta)'(\hat{\beta} - \beta)' x_t\right] \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1'} \\
&= o_p(1).
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
&\left\| \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0} \otimes \frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^{n} \frac{\ddot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2}(\varepsilon_t^2 - 1) \right\| \\
&\leq \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0}\right)^{-1} \left(\sup_{|s|\leq \bar{s}} \left| \frac{\ddot{g}(\sqrt{n}s,\theta_0)}{g(\sqrt{n}s,\theta_0)^2} \right|\right) \right\| \frac{1}{n} \sum_{t=1}^{n} \left|(\varepsilon_t^2 - 1)\right| \\
&\to_p 0.
\end{aligned}$$

As for the second term, we have

$$\begin{aligned}
&\left\| \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0} \otimes \frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0}\right)^{-1} \sum_{t=1}^{n} \frac{\ddot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} u_t x_t'(\hat{\beta} - \beta) \right\| \\
&\leq \frac{1}{n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0}\right)^{-1} \left(\sup_{|s|\leq \bar{s}} \left| \frac{\ddot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)} \right|\right) \right\| \left\|(\hat{\beta} - \beta)\right\| \left\| \sum_{t=1}^{n} \sigma(z_{t-1}) x_t \varepsilon_t \right\| \\
&\to_p 0.
\end{aligned}$$

As for the third term, we have

$$\left\| \left(\frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0} \otimes \frac{\sqrt{n}\,\dot{\kappa}_0}{\kappa_0} \right)^{-1} \sum_{t=1}^{n} \frac{\ddot{g}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} (\hat{\beta}-\beta)' x_t x_t' (\hat{\beta}-\beta) \right\|$$
$$\leq \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\leq \bar{s}} \left| \frac{\ddot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \left\| x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)' x_t \right\|$$
$$\leq \frac{1}{n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\leq \bar{s}} \left| \frac{\ddot{g}(\sqrt{n}s,\theta_0)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \left\| \frac{\sqrt{n}}{\kappa_0^{1/2}} (\hat{\beta}-\beta) \right\|^2 \left(\frac{1}{n} \sum_{t=1}^{n} \|x_t\|^2 \right)$$
$$\to 0.$$

AD3 : AD3 is satisfied by applying Lemma A6(b) and Theorem 3.3 in Park and Phillips (2001). We have $\ddot{Q}(\theta_0)$ given as

$$\ddot{Q}(\theta_0) = \int_0^1 \frac{\dot{h}(V,\theta_0)\dot{h}(V,\theta_0)'}{h(V,\theta_0)^2}.$$

The terms related to R_1 *and* R_2 *disappear as was shown in the Proof of AD3 and with the* condition (d).
AD4 : AD4 is already assumed in condition (c) of the Theorem 4.1.
AD5 : This step is very similar to the proof of Theorem 5.3 in Park and Phillips (2001). To show AD5, fix δ such that $0 < \delta < \varepsilon/3$, and define $\mu_n = n^{1/2-\delta}\,\dot{\kappa}_n$ and $\nu_n = n^{1/2}\,\dot{\kappa}_n$ so that $\mu_n \nu_n^{-1} \to 0$ as required. We first write

$$\ddot{Q}_n^*(\theta) - \ddot{Q}_n^*(\theta_0)$$
$$= (\ddot{D}_{1n}(\theta) + \ddot{D}_{1n}(\theta)') + \ddot{D}_{2n}(\theta) + \ddot{D}_{3n}(\theta) - \ddot{D}_{4n}(\theta) + 2\ddot{D}_{5n}(\theta) - \ddot{D}_{6n}(\theta),$$

where

$$\ddot{D}_{1n}(\theta) = \sum_{t=1}^{n} \frac{\dot{g}(z_{t-1},\theta_0)(\dot{g}(z_{t-1},\theta) - \dot{g}(z_{t-1},\theta_0))'}{g(z_{t-1},\theta_0)^2}$$
$$\ddot{D}_{2n}(\theta) = \sum_{t=1}^{n} \frac{(\dot{g}(z_{t-1},\theta) - \dot{g}(z_{t-1},\theta_0))(\dot{g}(z_{t-1},\theta) - \dot{g}(z_{t-1},\theta_0))'}{g(z_{t-1},\theta_0)^2}$$
$$\ddot{D}_{3n}(\theta) = \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta)(g(z_{t-1},\theta) - g(z_{t-1},\theta_0))}{g(z_{t-1},\theta_0)^2}$$
$$\ddot{D}_{4n}(\theta) = \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta) - \ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} (u_t^2 - g(z_{t-1},\theta_0))$$
$$\ddot{D}_{5n}(\theta) = \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta) - \ddot{G}(z_{t-1},\theta_0)}{g(z_{t-1},\theta_0)^2} u_t x_t'(\hat{\beta}-\beta)$$

$$\ddot{D}_{6n}(\theta) = \sum_{t=1}^{n} \frac{(\ddot{G}(z_{t-1},\theta) - \ddot{G}(z_{t-1},\theta_0))}{g(z_{t-1},\theta_0)^2} x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)'x_t$$

and define

$$\overline{\omega}_{in}^2(\theta) = ||\mu_n^{-1}\ddot{D}_{in}(\theta)\mu_n^{-1'}||,$$

for i= 1,...,6. For all $\theta \in N_n$ where $N_n = \{\theta : ||\mu_n'(\theta-\theta_0)|| \leq 1\}$, we have

$$\overline{\omega}_{1n}^2(\theta) \leq \sum_{t=1}^{n} \left\|\mu_n^{-1}\dot{g}(z_{t-1},\theta_0)\right\| \left\|(\mu_n \otimes \mu_n)^{-1} \frac{\ddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\|, \tag{28}$$

$$\overline{\omega}_{2n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n)^{-1} \frac{\ddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\|^2, \tag{29}$$

$$\begin{aligned}\overline{\omega}_{3n}^2(\theta) \leq & \sum_{t=1}^{n} \left\|\mu_n^{-1}\dot{g}(z_{t-1},\theta_0)\right\| \left\|(\mu_n \otimes \mu_n)^{-1} \frac{\ddot{g}(z_{t-1},\theta)}{g^2(z_{t-1},\theta_0)}\right\| \\ & + \frac{1}{2}\sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n)^{-1} \frac{\ddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\| \left\|(\mu_n \otimes \mu_n)^{-1} \frac{\ddot{g}(z_{t-1},\theta)}{g^2(z_{t-1},\theta_0)}\right\| \end{aligned} \tag{30}$$

$$\overline{\omega}_{4n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1} \frac{\dddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\| |(\varepsilon_t^2 - 1)|, \tag{31}$$

$$\overline{\omega}_{5n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1} \frac{\dddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\| \left\|u_t x_t'(\hat{\beta}-\beta)\right\|, \tag{32}$$

$$\overline{\omega}_{6n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1} \frac{\dddot{g}(z_{t-1},\overline{\theta})}{g^2(z_{t-1},\theta_0)}\right\| \left\|x_t'(\hat{\beta}-\beta)'(\hat{\beta}-\beta)'x_t\right\| \tag{33}$$

where $\overline{\theta}$ lies in the line segment connecting θ and θ_0. We now need to evaluate the Brownian motion V at the maximum or the minimum of the sample path of V, which can be defined by

$$s_{\max} = \max_{0\leq r\leq 1} V(r) \text{ and } s_{\min} = \min_{0\leq r\leq 1} V(r),$$

and let $\bar{s} = \max(s_{\max}, -s_{\min}) + 1$, then we have for large n,

$$\sup_{\theta\in N_n} \left|\frac{\ddot{g}(z_{t-1},\theta)}{g^2(z_{t-1},\theta_0)}\right| \leq \sup_{|s|\leq\bar{s}} \sup_{\theta\in N_n} \left|\frac{\ddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)}\right|,$$

for all $t = 1, ..., n$. It now follows from (28)- (33) that

$$\overline{\omega}_{1n}^2(\theta) \le \frac{n^{3\delta}}{\sqrt{n}} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\ddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \times \frac{1}{n}\sum_{t=1}^{n} \left\| (\dot{\kappa}_0)^{-1} \dot{g}(z_{t-1},\theta_0) \right\|, \tag{34}$$

$$\overline{\omega}_{2n}^2(\theta) \le \frac{n^{4\delta}}{n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\ddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\|^2, \tag{35}$$

$$\overline{\omega}_{3n}^2(\theta) \le \frac{n^{3\delta}}{\sqrt{n}} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\ddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \times \frac{1}{n}\sum_{t=1}^{n} \left\| (\dot{\kappa}_0)^{-1} \dot{g}(z_{t-1},\theta_0) \right\| + \frac{n^{4\delta}}{2n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\ddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\|^2, \tag{36}$$

$$\overline{\omega}_{4n}^2(\theta) \le \frac{n^{3\delta}}{\sqrt{n}} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\dddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \times \frac{1}{n}\sum_{t=1}^{n} \left| (\varepsilon_t^2 - 1) \right|, \tag{37}$$

$$\overline{\omega}_{5n}^2(\theta) \le \frac{n^{3\delta}}{\sqrt{n}} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\dddot{g}(\sqrt{n}s,\theta)}{g^2(\sqrt{n}s,\theta_0)} \right| \right) \right\| \frac{1}{n} \left\| (\hat{\beta} - \beta) \right\| \left\| \sum_{t=1}^{n} \sigma(z_{t-1}) x_t \varepsilon_t \right\|, \tag{38}$$

$$\overline{\omega}_{6n}^2(\theta) \le \frac{n^{3\delta}}{\sqrt{n}} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \otimes \frac{\dot{\kappa}_0}{\kappa_0} \right)^{-1} \left(\sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{\dddot{g}(\sqrt{n}s,\theta)}{g(\sqrt{n}s,\theta_0)} \right| \right) \right\| \frac{1}{n} \left\| \frac{\sqrt{n}}{\kappa_0^{1/2}} (\hat{\beta} - \beta) \right\|^2 \left(\frac{1}{n}\sum_{t=1}^{n} \|x_t\|^2 \right), \tag{39}$$

from which we may easily deduce that $\overline{\omega}_{in}^2(\theta) = o_{a.s.}(1)$, $i = 1, ...6$, uniformly in $\theta \in N_n$ due to (21) and (22). Now AD5 follows immediately from (34)-(39).
(ii) Cointegrating regression : The proof is exactly the same as in the stationary regression except for the asymptotic order, that is,

$$\frac{n}{\kappa_0^{1/2}}(\hat{\beta} - \beta) = O_p(1) \text{ and } \frac{1}{n\kappa_0^{1/2}} \sum_{t=1}^{n} \sigma(z_{t-1}) x_t \varepsilon_t = O_p(1),$$

and hence skipped here to save space. □

Main Proof of Theorem 4.1. **(i) Stationary Regression** : Now, the objective function is different $\tilde{Q}_n(\theta)$ from $Q_n^*(\theta)$ because $g(z_{t-1}, \hat{\theta}_n)^2$ is used instead of $g(z_{t-1}, \theta_0)^2$. However, $\tilde{Q}_n(\theta)$ can be written as a function of $Q_n^*(\theta)$ as follows. Note that

$$\begin{aligned}
\tilde{Q}_n(\theta) &= \sum_{t=1}^{n} \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \theta_0)^2 - \{g(z_{t-1}, \hat{\theta}_n)^2 - g(z_{t-1}, \theta_0)^2\}} \\
&= \sum_{t=1}^{n} \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \theta_0)^2} \left(1 - \frac{g(z_{t-1}, \hat{\theta}_n)^2 - g(z_{t-1}, \theta_0)^2}{g(z_{t-1}, \theta_0)^2}\right)^{-1} \\
&= \sum_{t=1}^{n} \frac{\left(\hat{u}_t^2 - g(z_{t-1}, \theta)\right)^2}{g(z_{t-1}, \theta_0)^2} (1 + G^*(\cdot)) \qquad (40)
\end{aligned}$$

where

$$G^*(\cdot) = \frac{2\dot{g}(z_{t-1}, \bar{\theta})'}{g(z_{t-1}, \theta_0)^2}(\hat{\theta}_n - \theta_0),$$

which can be obtained by Taylor expansion. If we let

$$\ell(\hat{\theta}_n) = \frac{g(z_{t-1}, \hat{\theta}_n)^2 - g(z_{t-1}, \theta_0)^2}{g(z_{t-1}, \theta_0)^2},$$

then

$$\begin{aligned}
\left(1 - \frac{g(z_{t-1}, \hat{\theta}_n)^2 - g(z_{t-1}, \theta_0)^2}{g(z_{t-1}, \theta_0)^2}\right)^{-1} &= \frac{1}{1 - \ell(\theta_0)} + \frac{\dot{\ell}(\bar{\theta})}{\left(1 - \ell(\bar{\theta})\right)^2}(\hat{\theta}_n - \theta_0) \\
&= 1 + \frac{2\dot{g}(z_{t-1}, \bar{\theta})}{g(z_{t-1}, \theta_0)^2}(\hat{\theta}_n - \theta_0)
\end{aligned}$$

where $\overline{\theta}$ lies between $\hat{\theta}_n$ and θ_0. We have

$$G^*(\bar{\theta}) = \frac{2\dot{g}(z_{t-1}, \bar{\theta})}{g(z_{t-1}, \theta_0)^2}(\hat{\theta}_n - \theta_0),$$

and therefore,

$$\tilde{Q}_n(\theta) = Q_n^*(\theta) + Q_n^*(\theta)G^*\left(\bar{\theta}\right)$$

Similarly, we have

$$\begin{aligned}\dot{\tilde{Q}}_n(\theta) &= -\sum_{t=1}^{n} \frac{\dot{g}(z_{t-1},\theta)(\hat{u}_t^2 - g(z_{t-1},\theta))}{g(z_{t-1},\hat{\theta}_n)^2} \\ &= \dot{Q}_n^*(\theta) + \dot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right),\end{aligned}$$

and

$$\begin{aligned}\ddot{\tilde{Q}}_n(\theta) &= \sum_{t=1}^{n} \frac{\dot{g}(z_{t-1},\theta)\dot{g}(z_{t-1},\theta)'}{g(z_{t-1},\hat{\theta}_n)^2} - \sum_{t=1}^{n} \frac{\ddot{G}(z_{t-1},\theta)}{g(z_{t-1},\hat{\theta}_n)^2}(\hat{u}_t^2 - g(z_{t-1},\theta)) \\ &= \ddot{Q}_n^*(\theta) + \ddot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right).\end{aligned}$$

Therefore, it suffices to show the following conditions to have all the sufficient conditions satisfied.

AD1*: $\upsilon_n^{-1}\dot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right) \to_p 0$ as $n \to \infty$.

AD2*: $\upsilon_n^{-1}\ddot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right)\upsilon_n^{-1\prime} = o_p(1)$ for large n.

AD3*: $\upsilon_n^{-1}\ddot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right)\upsilon_n^{-1\prime} \to_p 0$ as $n \to \infty$.

AD4*: $\ddot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right) > 0$ a.s.

AD5*: There is a sequence μ_n such that $\mu_n\upsilon_n^{-1} \to_{a.s.} 0$ such that

$$\sup_{\theta\in N_n}\left\|\mu_n^{-1}\left([\ddot{Q}_n^*(\theta) - \ddot{Q}_n^*(\theta_0)]G^*\left(\bar{\theta}\right)\right)\mu_n^{-1\prime}\right\| \to_p 0$$

All the conditions AD1*–AD5* are readily seen to be satisfied. For example, we have

$$\begin{aligned}\left\|\upsilon_n^{-1}\dot{Q}_n^*(\theta)G^*\left(\bar{\theta}\right)\right\| &\le \left\|\upsilon_n^{-1}\dot{Q}_n^*(\theta)\right\|\left\|G^*\left(\bar{\theta}\right)\right\| \\ &\le \sup_{|s|\le\bar{s}}\ \sup_{\theta\in N(\varepsilon,\lambda)}\left|\frac{\dot{g}(\lambda s,\theta)}{g^2(\lambda s,\theta_0)}\right|\left\|\hat{\theta}_n - \theta_0\right\|\left\|\upsilon_n^{-1}\dot{Q}_n^*(\theta)\right\| \to 0,\end{aligned}$$

since

$$\upsilon_n^{-1}\dot{Q}_n^*(\theta) = O_p(1)$$

and
$$\sup_{|s|\leq\bar{s}}\sup_{\theta\in N(\varepsilon,\lambda)}\left|\frac{\dot{g}(\lambda s,\theta)}{g^2(\lambda s,\theta_0)}\right|\left\|\hat{\theta}_n-\theta_0\right\|\to 0$$

from the asymptotic result of ordinary nonlinear least squares estimator $\hat{\theta}_n$ as
$$\frac{\sqrt{n}\,\dot{\kappa}_0'}{\kappa_0}\left(\tilde{\theta}_n-\theta_0\right)=O_p(1),$$

and the assumption (23)
$$\lambda^{-1+\varepsilon}\left\|\left(\frac{\dot{\kappa}_0}{\kappa_0}\right)(\lambda)^{-1}\left(\sup_{|s|\leq\bar{s}}\sup_{\theta\in N(\varepsilon,\lambda)}\left|\frac{\dot{g}(\lambda s,\theta)}{g^2(\lambda s,\theta_0)}\right|\right)\right\|\to 0$$

Therefore
$$\sup_{|s|\leq\bar{s}}\sup_{\theta\in N(\varepsilon,\lambda)}\left|\frac{\dot{g}(\lambda s,\theta)}{g^2(\lambda s,\theta_0)}\right|\left\|\hat{\theta}_n-\theta_0\right\|=o_{a.s.}(1)$$

uniformly in $\theta\in N(\varepsilon,\lambda)$.Therefore, AD1* is satisfied. The proof for the cointegration regression will be the same and will be skipped to save space. □

Proof of Theorem 4.2(i). When x_t is stationary and $\sigma(\cdot)$ is asymptotically homogenous, we may write
$$\begin{aligned}\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S^{\dagger}-\beta)&=M_n^{-1}\left[\frac{1}{\sqrt{n}\,\nu(\sqrt{n})}\sum_{t=1}^{n}\left(x_t\sigma(z_{t-1})\varepsilon_t-\frac{\hat{\sigma}_{x\varepsilon,\eta}^{\dagger}}{\hat{\sigma}_{\eta}^{2}}\hat{\sigma}(z_{t-1})\hat{\eta}_{p,t}\right)\right]\\&=M_n^{-1}\left[A_n+B_n\right].\end{aligned}$$

where
$$\begin{aligned}A_n&=\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\hat{\sigma}_{x\varepsilon,\eta}^{\dagger}}{\hat{\sigma}_{\eta}^{2}}\hat{\eta}_{p,t}\right)\quad\text{and}\\B_n&=\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\frac{\hat{\sigma}_{x\varepsilon,\eta}^{\dagger}}{\hat{\sigma}_{\eta}^{2}}\left(\sigma(z_{t-1})-\hat{\sigma}(z_{t-1})\right)\hat{\eta}_{p,t}.\end{aligned}$$

As for B_n, using equation (60) of Chang et al. (2001), we have
$$\begin{aligned}&\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\left(\sigma(z_{t-1},\tilde{\theta}_n)-\sigma(z_{t-1},\theta_0)\right)\hat{\eta}_{p,t}\\&=\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\left(\sigma(z_{t-1},\tilde{\theta}_n)-\sigma(z_{t-1},\theta_0)\right)\eta_t+o_p(1).\end{aligned}$$

Note that

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{n}\nu(\sqrt{n})} \sum_{t=1}^{n} \left(\sigma(z_{t-1},\tilde{\theta}_n) - \sigma(z_{t-1},\theta_0)\right) \eta_t \right\| \\
&\leq \sup_{\bar{\theta}\in N_n} \left\| \frac{\sigma(z_{t-1},\tilde{\theta}_n) - \sigma(z_{t-1},\theta_0)}{\sigma(z_{t-1},\theta_0)} \right\| \left\| \frac{1}{\sqrt{n}\nu(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1})\eta_t \right\| \\
&\leq \sup_{|s|\leq\bar{s}} \sup_{\bar{\theta}\in N_n} \left\| \frac{\dot{\sigma}(\sqrt{ns},\bar{\theta})}{\sigma(\sqrt{ns},\theta_0)} \right\| \left\| \tilde{\theta}_n - \theta_0 \right\| \left\| \frac{1}{\sqrt{n}\nu(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1})\eta_t \right\|.
\end{aligned}$$

where $\overline{\theta}$ lies between $\tilde{\theta}_n$ and θ_0. Since η_t is a martingale difference sequence, we have

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

Note that

$$\frac{\dot{\sigma}(\sqrt{ns},\bar{\theta})}{\sigma(\sqrt{ns},\theta_0)} = \frac{\dot{g}(\lambda s,\theta)}{g(\lambda s,\theta_0)},$$

and hence we have

$$\frac{\sqrt{n}\,\dot{\kappa}_0'}{\kappa_0}\left(\tilde{\theta}_n - \theta_0\right) = O_p(1),$$

from the Theorem 4.1 and

$$\sqrt{n}^{-1+\varepsilon} \sup_{|s|\leq\bar{s}} \sup_{\bar{\theta}\in N_n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0}\right) (\sqrt{n})^{-1} \frac{\dot{\sigma}(\sqrt{ns},\bar{\theta})}{\sigma(\sqrt{ns},\theta_0)} \right\| \to 0,$$

from the assumption (24). Therefore

$$\left\| \frac{1}{\sqrt{n}\nu(\sqrt{n})} \sum_{t=1}^{n} \left(\sigma(z_{t-1},\tilde{\theta}_n) - \sigma(z_{t-1},\theta_0)\right) \eta_t \right\| = o_{a.s.}(1)$$

uniformly in $\theta \in N(\varepsilon,\lambda)$.

We now need to show that

$$\frac{1}{n}\sum_{t=1}^{n} x_t \hat{\varepsilon}_t^{\dagger} \hat{\eta}_{p,t} \to_p \sigma_{x\varepsilon,\eta}.$$

Notice that

$$\begin{aligned}
\hat{\varepsilon}_t^{\dagger} &= \frac{\hat{u}_t}{\sigma(z_{t-1},\tilde{\theta}_n)} = \frac{\hat{u}_t}{\sigma(z_{t-1},\theta_0)+\left(\sigma(z_{t-1},\tilde{\theta}_n)-\sigma(z_{t-1},\theta_0)\right)} \\
&= \frac{\hat{u}_t}{\sigma(z_{t-1},\theta_0)\left[1+\frac{\sigma(z_{t-1},\tilde{\theta}_n)-\sigma(z_{t-1},\theta_0)}{\sigma(z_{t-1},\theta_0)}\right]} \\
&= \frac{\hat{u}_t}{\sigma(z_{t-1},\theta_0)} + \frac{\hat{u}_t}{\sigma(z_{t-1},\theta_0)}\Phi(\cdot) = \hat{\varepsilon}_t + \hat{\varepsilon}_t\,\Phi(\cdot)
\end{aligned}$$

where

$$\Phi\left(\overline{\theta}\right) = -\frac{\dot{\sigma}(z_{t-1},\overline{\theta})}{\sigma(z_{t-1},\,\theta_0)}(\tilde{\theta}_n-\theta_0),$$

and $\overline{\theta}$ lies between $\tilde{\theta}_n$ and θ_0 by the same expansion in (40). Since

$$\sqrt{n}^{-1+\varepsilon}\sup_{|s|\le\bar{s}}\sup_{\bar{\theta}\in N_n}\left\|\left(\frac{\dot{\kappa}_0}{\kappa_0}\right)(\sqrt{n})^{-1}\frac{\dot{\sigma}(\sqrt{ns},\bar{\theta})}{\sigma(\sqrt{ns},\theta_0)}\right\| \to 0,$$

it follows that

$$\left\|\frac{1}{n}\sum_{t=1}^{n}x_t\hat{\varepsilon}_t\Phi\left(\overline{\theta}\right)\hat{\eta}_{p,t}\right\| = o_{a.s.}(1)$$

uniformly in $\theta\in N(\varepsilon,\lambda)$. Therefore, we have

$$\begin{aligned}
\hat{\sigma}^{\dagger}_{x\varepsilon,\eta} &= \frac{1}{n}\sum_{t=1}^{n}x_t\hat{\varepsilon}_t^{\dagger}\hat{\eta}_{p,t} = \frac{1}{n}\sum_{t=1}^{n}x_t\hat{\varepsilon}_t\hat{\eta}_{p,t}+o_p(1) \\
&\to_p \sigma_{x\varepsilon,\eta}.
\end{aligned}$$

By the Theorem 10 in Chang et al. (2001), we have

$$\begin{aligned}
A_n &= \frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\hat{\eta}_{p,t}\right) \\
&= \frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t-\frac{\sigma_{x\varepsilon,\eta}}{\sigma^2_{\eta}}\eta_t\right)+o_p(1) \\
&\to_d \int_0^1\tau(V(r))dU_*(r),
\end{aligned}$$

from which the stated result follows directly upon noticing that V is independent of U_*. □

Proof of Theorem 4.2(ii). For the cointegrating regression where $\sigma(\cdot)$ is asymptotically homogenous, we may write

$$\begin{aligned}\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}_N^{\dagger}-\beta) &= W_n^{-1}\left[\frac{1}{n\,\nu(\sqrt{n})}\sum_{t=1}^{n}x_t\left(\sigma(\alpha'x_t)\varepsilon_t-\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\widehat{\Sigma}_{\eta}^{-1}\hat{\sigma}(\alpha'x_t)\hat{\eta}_{p,t}\right)\right]\\ &= W_n^{-1}\left[A_n+B_n\right].\end{aligned}$$

where

$$\begin{aligned}A_n &= \frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(\alpha'x_t)x_t\left(\varepsilon_t-\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\widehat{\Sigma}_{\eta}^{-1}\,\hat{\eta}_{p,t}\right)\quad\text{and}\\ B_n &= \frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}x_t\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\widehat{\Sigma}_{\eta}^{-1}\left(\sigma(\alpha'x_t)-\hat{\sigma}(\alpha'x_t)\right)\hat{\eta}_{p,t}.\end{aligned}$$

As for A_n, from equation (66) in the proof of Theorem 10 in Chang et al. and $\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\to_p\sigma_{\varepsilon,\eta}$ as is shown in the proof of Theorem 4.2(i), we have

$$\begin{aligned}&\frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(\alpha'x_t)x_t\left(\varepsilon_t-\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\widehat{\Sigma}_{\eta}^{-1}\,\hat{\eta}_{p,t}\right)\\ &=\frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(\alpha'x_t)x_t\left(\varepsilon_t-\sigma_{\varepsilon,\eta}\Sigma_{\eta}^{-1}\,\eta_t\right)+o_p(1)\\ &\to_d\int_0^1\tau(\alpha'V(r))V(r)dU_*(r).\end{aligned}$$

As for B_n, following the same line in the proof of Theorem 4.2(i),

$$\left\|\frac{1}{n\nu(\sqrt{n})}\sum_{t=1}^{n}x_t\hat{\sigma}_{\varepsilon,\eta}^{\dagger}\widehat{\Sigma}_{\eta}^{-1}\left(\hat{\sigma}(\alpha'x_t)-\sigma(\alpha'x_t)\right)\hat{\eta}_{p,t}\right\|=o_{a.s.}(1)$$

uniformly in θ. Now the stated result follows directly upon noticing that V is independent of U_*. □

Proof of Theorem 4.3(i). Under H_0, we let

$$G(\tilde{\beta}_S^{\dagger})=\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S^{\dagger}-\beta)\right]'R'\left(RM_n^{-1}\tilde{P}_n^{*}M_n^{-1}R'\right)^{-1}R\left[\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}_S^{\dagger}-\beta)\right],$$

and we may write

$$\tilde{P}_n^{*}=\frac{1}{n\,\nu^2(\sqrt{n})}\sum_{t=1}^{n}(x_tu_t)^{**}(x_tu_t)^{**\prime}=A_n-(R_{1n}+R_{1n}')+R_{2n}$$

where

$$A_n = \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\left(\sigma(z_{t-1})x_t\varepsilon_t - \hat{\sigma}(z_{t-1})\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\times \left(\sigma(z_{t-1})x_t\varepsilon_t - \hat{\sigma}(z_{t-1})\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)'$$

$$R_{1n} = \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\left(\sigma(z_{t-1})x_t\varepsilon_t - \hat{\sigma}(z_{t-1})\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\left(x_tx_t'(\tilde{\beta}^{\dagger}_S-\beta)\right)'$$

$$R_{2n} = \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}x_tx_t'(\tilde{\beta}^{\dagger}_S-\beta)(\tilde{\beta}^{\dagger}_S-\beta)'x_tx_t'$$

As for A_n, we can rewrite it as

$$\begin{aligned}A_n &= \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)' \\
&- \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right) \qquad (41)\\
&\times(\hat{\sigma}(z_{t-1})-\sigma(z_{t-1}))\left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)' \\
&- \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}(\hat{\sigma}(z_{t-1})-\sigma(z_{t-1})) \\
&\times\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\,\sigma(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)' \qquad (42)\\
&+ \frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}(\hat{\sigma}(z_{t-1})-\sigma(z_{t-1}))^2\left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)'.\end{aligned}$$

It can be readily deduced from equations (50) and (66) in Chang et al. (2001) that

$$\begin{aligned}&\frac{1}{n\nu^2(\sqrt{n})}\sum_{t=1}^{n}\sigma^2(z_{t-1})\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\left(x_t\varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)' \qquad (43)\\
&\to_d \Sigma^*\int_0^1\tau^2(V(r))dr.\end{aligned}$$

The second term in A_n (41) can be written as

$$\left\| \frac{1}{n\,v^2(\sqrt{n})} \sum_{t=1}^{n} \sigma(z_{t-1}) \left(x_t \varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right) (\hat{\sigma}(z_{t-1}) - \sigma(z_{t-1})) \left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right)' \right\|$$

$$\leq \sup_{\bar{\theta} \in N_n} \left\| \frac{\sigma(z_{t-1}, \tilde{\theta}_n) - \sigma(z_{t-1}, \theta_0)}{\sigma(z_{t-1}, \theta_0)} \right\| \times$$

$$\left\| \frac{1}{n\,v^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \left(x_t \varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right) \left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right)' \right\|$$

$$\leq \sup_{|s| \leq \bar{s}} \sup_{\bar{\theta} \in N_n} \left\| \frac{\dot{\sigma}(\sqrt{ns}, \bar{\theta})}{\sigma(\sqrt{ns}, \theta_0)} \right\| \left\| \tilde{\theta}_n - \theta_0 \right\| \times$$

$$\left\| \frac{1}{n\,v^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \left(x_t \varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right) \left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right)' \right\|.$$

Since $x_t \varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma^2_{\eta}} \eta_t$ is a martingale difference sequence, and we have

$$\frac{1}{n\,v^2(\sqrt{n})} \sum_{t=1}^{n} \sigma^2(z_{t-1}) \left(x_t \varepsilon_t - \frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right) \left(\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}} \hat{\eta}_{p,t} \right)' = O_p(1)$$

following the same argument in (43). Note also that

$$\tilde{\theta}_n - \theta_0 = O_p(\sqrt{n}\,\dot{\kappa}_0(\sqrt{n})^{-1} \kappa_0(\sqrt{n})),$$

and

$$\sqrt{n}^{-1+\varepsilon} \sup_{|s| \leq \bar{s}} \sup_{\bar{\theta} \in N_n} \left\| \left(\frac{\dot{\kappa}_0}{\kappa_0} \right) (\sqrt{n})^{-1} \frac{\dot{\sigma}(\sqrt{ns}, \bar{\theta})}{\sigma(\sqrt{ns}, \theta_0)} \right\| \to 0,$$

which was shown in the proof of the Theorem 4.2(i). Therefore, the second term in A_n (41) is $o_{a.s.}(1)$ uniformly in $\theta \in N(\varepsilon, \lambda)$. Since the fourth term in A_n is also $o_{a.s.}(1)$ uniformly in $\theta \in N(\varepsilon, \lambda)$ in the same way, we have

$$A_n \to_d \Sigma^* \int_0^1 \tau^2(V(r))dr.$$

Moreover, we may also easily establish that

$$
\begin{aligned}
R_{1n} &= \frac{1}{n}\left[\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\left(\sigma(z_{t-1})x_t\varepsilon_t - \hat{\sigma}(z_{t-1})\frac{\hat{\sigma}^{\dagger}_{x\varepsilon,\eta}}{\hat{\sigma}^2_{\eta}}\,\hat{\eta}_{p,t}\right)\times\right.\\
&\left.\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_S-\beta)\right)' x_t x_t'\right]\\
&= \frac{1}{n}\left[\frac{1}{\sqrt{n}\nu(\sqrt{n})}\sum_{t=1}^{n}\sigma(z_{t-1})\{x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma^2_{\eta}}\,\eta_t\}\times\right.\\
&\left.\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_S-\beta)\right)' x_t x_t'\right] + o_p(1)\\
&= o_p(1)
\end{aligned}
$$

which follows immediately upon noticing that $\{x_t\varepsilon_t - \frac{\sigma_{x\varepsilon,\eta}}{\sigma^2_{\eta}}\,\eta_t\}\otimes x_t x_t'$ is a martingale difference sequence, due to Theorem 3.3 in Park and Phillips (2001) and Theorem 4.1. As for R_{2n} which is

$$
R_{2n} = \frac{1}{n}\left[\frac{1}{n}\sum_{t=1}^{n}x_t x_t'\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_S-\beta)\right)\left(\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_S-\beta)\right)' x_t x_t'\right],
$$

we can establish

$$
\|R_{2n}\| \le \frac{1}{n}\left\|\frac{\sqrt{n}}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_S-\beta)\right\|^2\left(\frac{1}{n}\sum_{t=1}^{n}\|x_t\|^4\right) = O_p(n^{-1}).
$$

Since U_* is central normal mixture with mixing variate $\Sigma_* S$ and that U_* and V are independent, we deduce the stated results in the theorem 4.3 in the stationary case. □

Proof of Theorem 4.3(ii). Under H_0, we let

$$
G(\tilde{\beta}_N) = \left[\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_N-\beta)\right]' R'\left(RW_n^{-1}\tilde{P}^*_n W_n^{-1}R'\right)^{-1}R\left[\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_N-\beta)\right],
$$

and we may write

$$
\tilde{P}^*_n = \frac{1}{n^2\nu^2(\sqrt{n})}\sum_{t=1}^{n}u_t^{**2}x_t x_t' = A_n - (R_{1n}+R_{1n}') + R_{2n}.
$$

where

$$A_n = \frac{1}{n^2 \nu^2(\sqrt{n})} \sum_{t=1}^{n} x_t \left(\sigma(z_{t-1})\varepsilon_t - \hat{\sigma}(z_{t-1})\hat{\sigma}^{\dagger}_{\varepsilon,\eta}\widehat{\Sigma}^{-1}_{\eta}\,\hat{\eta}_{p,t}\right) \times \left(\sigma(z_{t-1})\varepsilon_t - \hat{\sigma}(z_{t-1})\hat{\sigma}^{\dagger}_{\varepsilon,\eta}\widehat{\Sigma}^{-1}_{\eta}\,\hat{\eta}_{p,t}\right)' x_t',$$

$$R_{1n} = \frac{1}{n^2 \nu^2(\sqrt{n})} \sum_{t=1}^{n} x_t \left(\sigma(z_{t-1})\varepsilon_t - \hat{\sigma}(z_{t-1})\hat{\sigma}^{\dagger}_{\varepsilon,\eta}\widehat{\Sigma}^{-1}_{\eta}\,\hat{\eta}_{p,t}\right) \left(x_t x_t'(\tilde{\beta}^{\dagger}_N - \beta)\right)',$$

$$R_{2n} = \frac{1}{n^2 \nu^2(\sqrt{n})} \sum_{t=1}^{n} x_t x_t'(\tilde{\beta}^{\dagger}_N - \beta)(\tilde{\beta}^{\dagger}_N - \beta)' x_t x_t'.$$

As for A_n, the lines of proof will be the same as the previous proof, and therefore we have

$$A_n \to_d \sigma^2_* \int_0^1 \tau^2(\alpha' V(r)) V(r) V(r)' dr$$

We may also easily establish that

$$R_{1n} = \frac{1}{n}\left[\frac{1}{n^2 \nu(\sqrt{n})} \sum_{t=1}^{n} x_t \left(\sigma(z_{t-1})\varepsilon_t - \hat{\sigma}(z_{t-1})\hat{\sigma}^{\dagger}_{\varepsilon,\eta}\widehat{\Sigma}^{-1}_{\eta}\,\hat{\eta}_{p,t}\right) \times \left(\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_N - \beta)\right)' x_t x_t'\right]$$
$$= O_p(n^{-1}),$$

$$R_{2n} = \frac{1}{n}\left[\frac{1}{n^3} \sum_{t=1}^{n} \left(x_t'\left(\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_N - \beta)\right)\left(\frac{n}{\nu(\sqrt{n})}(\tilde{\beta}^{\dagger}_N - \beta)\right)' x_t\right)(x_t x_t')\right]$$
$$= O_p(n^{-1}),$$

The rest of the proof is almost identical to the proof of Theorem 3.3 and is omitted. □

REFERENCES

Boswijk, H. P. (2005). Adaptive testing for a unit root with nonstationary volatility. Working paper, Amsterdam School of Economics

Cavaliere, G. (2004). Unit root tests under time-varying variance shifts. Econometric Reviews 23, 259–92.

Cavaliere, G., and A. M. R. Taylor (2007). Testing for unit roots in time series models with nonstationary volatility. Journal of Econometrics 140, 919–47.

Cavaliere, G., A. Rahbek., and A. M. R. Taylor (2010).Testing for co-integration in vector autoregressions with non-stationary volatility. Journal of Econometrics 158, 7–24.

Cavaliere, G., P. C. B. Phillips, S. Smeekes, and A. M. R. Taylor (2012). Lag length selection for unit root tests in the presence of nonstationary volatility, Cowles Foundation Discussion Paper No. 1844.

Chang, Y., and J. Y. Park (2003). Index models with integrated time series. Journal of Econometrics 114, 73–106.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. Econometrics Journal 4, 1–36.

Chung, H., and J. Y. Park (2007). Nonstationary nonlinear heteroskedasticity in regression. Journal of Econometrics 137, 230–259.

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

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

Hansen, B. E. (1995). Regression with nonstationary volatility. Econometrica 63, 1113–1132.

Kim, C. S., and J. Y. Park (2010). Cointegrating Regressions with Time Heterogeneity. Econometric Reviews 29, 397–438.

Kim, J., and C. S. Kim (2010). Local Linear Estimation of Nonparametric Cointegrating Regression. Journal of Economic Theory and Econometrics 21, 23–42.

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., and P. C .B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series, Econometric Theory 15, 269–298.

Park, J. Y., and P. C. B. Phillips (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. (1991). Optimal Inference in Cointegrating System. Econometrica 59, 283–306.

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

Phillips, P. C. B., and Xu, K.-L., (2006). Inference in autoregression under heteroskedasticity, Journal of Time Series Analysis 27, 289–308.

Table 1: Rejection Probabilities of Wald Tests in Stationary Regression

	$n = 100$			$n = 500$		
	HGF: $e^z/(1+e^z)$					
	1%	5%	10%	1%	5%	10%
Wald	1.26	5.67	10.84	1.32	6.18	12.00
HEC Wald	1.81	6.65	11.97	1.40	5.74	10.82
	HGF: $\|z\|^{0.25}$					
	1%	5%	10%	1%	5%	10%
Wald	0.92	4.51	8.85	0.76	4.14	8.34
HEC Wald	1.87	6.87	12.33	1.16	5.30	10.32
	HGF: $\|z\|^{0.5}$					
	1%	5%	10%	1%	5%	10%
Wald	0.72	3.81	7.91	0.52	3.07	6.57
HEC Wald	1.85	6.94	12.46	1.22	5.37	10.36
	HGF: $\|z\|^1$					
	1%	5%	10%	1%	5%	10%
Wald	0.37	2.37	5.47	0.20	1.65	4.23
HEC Wald	1.47	6.09	11.39	0.95	4.54	9.34

Table 2: Rejection Probabilities of Wald Tests in Cointegrating Regression

	$n=100$			$n=500$		
	HGF: $e^z/(1+e^z)$					
	1%	5%	10%	1%	5%	10%
Wald	1.27	5.59	10.86	1.18	5.03	9.49
HEC Wald	1.62	6.49	11.94	1.08	5.20	10.20
	HGF: $\|z\|^{0.5}$					
	1%	5%	10%	1%	5%	10%
Wald	7.04	16.32	23.96	6.41	15.93	23.78
HEC Wald	1.90	6.95	12.43	1.20	5.44	10.33
	HGF: $\|z\|$					
	1%	5%	10%	1%	5%	10%
Wald	11.99	23.32	31.55	8.44	18.80	26.94
HEC Wald	2.11	7.52	13.34	1.22	5.46	10.60
	HGF: $\|z\|^2$					
	1%	5%	10%	1%	5%	10%
Wald	13.58	25.80	34.50	11.03	22.56	31.04
HEC Wald	2.13	7.98	14.33	1.22	5.73	11.25

Figure 1: Densities of OLS and ELS estimators in Stationary Regressions

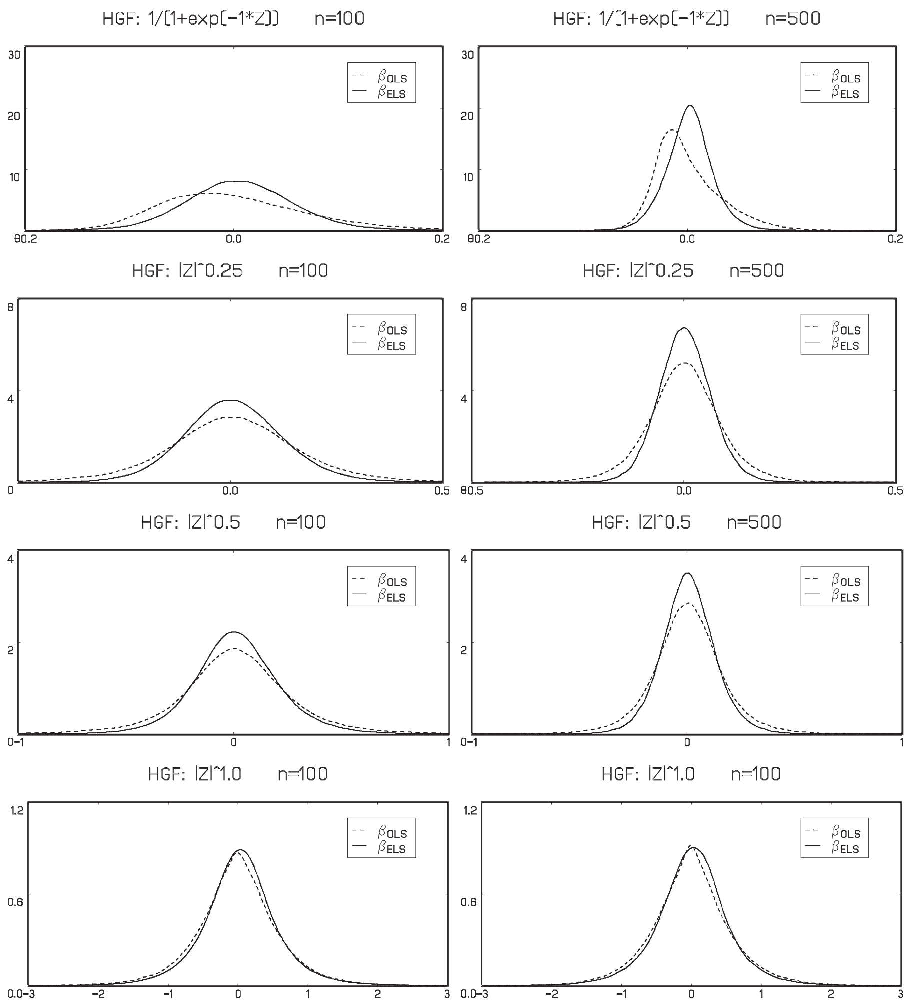

Figure 2: Densities of OLS and ELS estimators in Cointegrating Regressions