Regression Modelling under General Heterogeneity

Liudas Giraitis, George Kapetaniosi, Yufei Li

Working Paper No. 983 August 2024 ISSN 1473-0278

School of Economics and Finance

Queen Mary University of London

Regression Modelling under General Heterogeneity

Liudas Giraitis[1], George Kapetanios[2], Yufei Li[2]

[1]Queen Mary University of London

[2]King's College London

August 22, 2024

Abstract

This paper introduces and analyses a setting with general heterogeneity in regression modelling. It shows that regression models with fixed or time-varying parameters can be estimated by OLS or time-varying OLS methods, respectively, for a very wide class of regressors and noises, not covered by existing modelling theory. The new setting allows the development of asymptotic theory and the estimation of standard errors. The proposed robust confidence interval estimators permit a high degree of heterogeneity in regressors and noise. The estimates of robust standard errors coincide with the well-known estimator of heteroskedasticity-consistent standard errors by White (1980), but are applicable to more general circumstances than just the presence of heteroscedastic noise. They are easy to compute and perform well in Monte Carlo simulations. Their robustness, generality and ease of use make them ideal for applied work. The paper includes a brief empirical illustration.

Keywords: robust estimation, structural change, time-varying parameters, non-parametric estimation

JEL Classification: C12, C51

1 Introduction

Regression analysis is the cornerstone of statistical theory and practice. Ordinary least squares (OLS) has been applied, within various regression contexts, to build an extensive toolkit, for the exploration of economic and financial datasets. The basic theory underlying OLS estimation and inference in regression models has been mostly settled for the best part of a century (see e.g. Lai and Wei (1982)). Relatedly, the problem of robust estimation has been a focus of empirical work in economics starting with the seminal work by White (1980) and its importance is well-understood in applied econometrics: Angrist and Pischke (2010) indicated that "Leamer (1983) diagnosed his contemporaries' empirical work as suffering from a distressing lack of robustness to changes in key assumptions" and Leamer (2010) reflected that "sooner or later, someone articulates the concerns that gnaw away in each of us and asks if the Assumptions are valid."

Most recent theoretical developments have been focused on allowing more general and robust regression settings. An important elaboration has been the consideration of settings where the regression coefficient changes across different observations within the available dataset. Relatedly, allowing more general structures for regressors and regression errors has also been a major focus.

Our paper aims to provide important extensions to a regression setting and allow a very general environment for regressors and error terms. One practical implication of our heterogeneity setting is that standard errors can be easily computed. Likewise, the investigation of structural change in the parameters of statistical and econometric models has received increasing attention in the literature over the past couple of decades. This development is not surprising. Assuming, wrongly, that the model structure remains fixed over time has clear adverse implications, such as inconsistency of parameter estimators and associated test-statistics, and major forecast failures.

The modelling of deterministic smooth evolution of parameters has a long pedigree in statistics, such as linear processes with time-varying spectral densities introduced by Priestley (1965). The context of such modelling is nonparameteric and has been followed up by Robinson (1989), Robinson (1991), Dahlhaus (1997), some of whom refer to such processes as locally stationary. This approach, however, has not been prominent in other applied areas, such as economics, where, random coefficient models dominate.

Various methods have been proposed to identify and handle structural change. In early contributions, changes were supposed to be deterministic, to occur rarely, and to be abrupt. Testing for the presence of parameter breaks leads back to the ground-breaking work by Chow (1960), see e.g. also Brown et al. (1975), Ploberger and Krämer (1992). More recent standard approaches allow for random evolution of parameters, where changes can be either discrete, as in Markov Switching models Hamilton (1989) or threshold models Tong (1990), continuous as in smooth transition models Terasvirta (1998), or driven by unobservable shocks, as in random coefficient models Nyblom (1989). Cogley and Sargent (2005) use random coefficient models for stochastic volatility, and Primiceri (2005) studies whether changes in parameters or in the variance of shocks - policy or otherwise - gave rise to the period of macroeconomic calmness, referred to as the "Great Moderation", after 1985. In these models, parameters typically evolve as random walks or autoregressive processes.

Building on this work, Giraitis et al. (2014), Giraitis et al. (2018), Dendramis et al. (2021) and others have developed a theoretical framework for random coefficient models and their estimation using kernel methods which performs well in finite samples. They have trivial computational cost and are easy to use in applied work, for example, Chronopoulos et al. (2022) showed the empirical prevalence of persistent volatility, and, that GARCH type volatility structures might be less common, than previously thought. However, a full treatment of estimation and inference within a general regression model, has not, surprisingly, been provided.

This paper combines and extends these two related but distinct work strands in a very general regression context. Firstly, we start by considering OLS estimation and inference, in the presence of heterogeneity or scale factors in both regressors and regression error terms which allow to capture heterogeneity in data. Very little is assumed about these scale factors, i.e. sample moments of regressors and regression error may not possess well defined limits. This necessitates novel theoretical analysis of OLS estimation and its associated standard errors. We clarify the relevance of our theory by showing how standard OLS inference fails even in very common regression settings, while robust estimation trivially solves the inference problem. While the form of our proposed inference coincides with that of White (1980), it is applicable, and indeed necessary, in a much wider context than just heteroscedasticity. For example, standard OLS inference does not apply to stationary autoregressive models with martingale difference innovations, see section 5. Our inferential methods can accommodate many other commonly adopted models, which have no heteroscedastic features, as defined in White (1980), or more generally, regression with missing data. Further, we consider robust OLS estimation of time-varying regression parameters, in our general regression setting which permits a very wide class of regressors and regression noises.

The overall outcome is a general theory for regression models with fixed and time-varying parameters under minimal assumptions on regressors and regression errors. Robust standard errors are also derived and are easy to compute. Extensive simulations illustrate that robust regression estimation procedures are well behaved, unlike commonly used alternatives. This setting is well suited to modelling economic and financial data.

The remainder of this paper is structured as follows. Section 2 presents the regression setting that allows for heterogeneity and dependence and outlines the theoretical results for the fixed parameter case. Section 3 extends the analysis to the time-varying regression parameter case. In Section 4 we show that our approach allows regression for a wide range of missing data patterns. Section 5 discusses the robust estimation of an AR(p) model.

Sections 6 presents Monte Carlo simulation results. In Section 7 we provide an empirical example of the application of the robust regression analysis to the modelling of asset returns, and in Section 8 we conclude. The proofs of our results are given in the Supplemental Material.

2 OLS estimation under general heterogeneity

In this section we focus on the ordinary least squares (OLS) estimation in environment permitting general heterogeneity in regression modelling. We analyze the model

$$y_t = \beta' z_t + u_t, \quad u_t = h_t \varepsilon_t, \tag{1}$$

where β is a p-dimensional vector, $z_t = (z_{1t}, ..., z_{pt})'$ is a stochastic regressor and u_t is an uncorrelated noise. To allow for an intercept, we can set the initial component $z_{1t} = 1$ to be equal to 1.

Our primary interest is to expand the OLS estimation procedure to a broad setting outlined by structural assumptions aligned with empirical research. They cover variety of types of potential regression variables that may appear in applied work, and match the generality of the setting used in Giraitis et al. (2024) in testing for absence of correlation and cross-correlation under general heterogeneity.

We start with specification of an uncorrelated noise u_t. We suppose that

$$u_t = h_t \varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a zero mean stationary uncorrelated martingale difference noise and $\{h_t\}$ is a deterministic or stochastic scale factor which is independent of $\{\varepsilon_t\}$. More specifically, we impose

Assumption 2.1. *$\{\varepsilon_t\}$ is a stationary martingale difference (m.d.) sequence with respect to some σ-field filtration $\mathcal{F}_t$:*

$$\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}\varepsilon_t^4 < \infty, \quad \mathbb{E}\varepsilon_t^2 = 1.$$

$\{\varepsilon_t\}$ is independent of $\{h_t\}$. Moreover, variable ε_1 has probability distribution density $f(x)$ and $f(x) \le c < \infty$ when $|x| \le x_0$ for some $x_0 > 0$.

The information set $\mathcal{F}_t$ will be generated by the past history $\mathcal{F}_t = \sigma(\varepsilon_s, \, s \le t)$ and possibly other variables.

We suppose that the regressors $z_t = (z_{1t}, ..., z_{pt})'$ have the following structure which is the key feature of our regression setting. For $k = 1, ..., p$,

$$z_{kt} = \mu_{kt} + g_{kt} \eta_{kt}, \tag{3}$$

where $\eta_t = (\eta_{1t}, ..., \eta_{pt})'$ is a stationary sequence, $g_t = (g_{1t}, ..., g_{pt})'$ are deterministic or stochastic scale factors, and $\mu_t = (\mu_{1t}, ..., \mu_{pt})'$ is a vector of deterministic or stochastic means. We assume that $\{\mu_t, g_t, h_t\}$ are independent of $\{\varepsilon_t, \eta_t\}$.

To account for the intercept in regression model (1), we can set

$$z_{1t} \equiv 1 = \mu_{1t} + g_{1t}\eta_{1t}, \quad \mu_{1t} = 0, \quad g_{1t} = \eta_{1t} = 1. \tag{4}$$

We suppose that in (3) $E\eta_{kt} = 0$ except for the intercept (4) where $\eta_{1t} = 1$.

Regression setting (3) permits regressors $z_t = (z_{1t}, ..., z_{pt})'$ with time-varying conditional mean, $\mu_{kt} = E[z_{kt}|\mathcal{F}_n^*]$, and conditional variance, $g_{kt}^2 = \mathrm{var}(z_{kt}|\mathcal{F}_n^*)$, with respect to the information set $\mathcal{F}_n^* = \sigma\big(\mu_t, g_t, h_t, t = 1, ..., n\big)$ generated by scales and means.

The underlying stationary sequence $\{\eta_t\}$ plays an important role in our environment, although estimation of regression parameter β requires relaxed conditions on $\{\eta_t\}$.

Definition 2.1. *We say that a (univariate) covariance stationary sequence $\{\xi_t\}$ has short memory (SM) if $\sum_{h=-\infty}^{\infty} |\mathrm{cov}(\xi_h, \xi_0)| < \infty$.*

Assumption 2.2. *$\eta_t = (\eta_{1t}, ..., \eta_{pt})'$ is $\mathcal{F}_{t-1}$ measurable sequence, $E[\eta_{kt}^2] = 1$ and $E[\eta_{kt}^4] < \infty$.*

(i) For $k, j = 1, ..., p$, $\{\eta_{kt}\}$ and $\{\eta_{jt}\eta_{kt}\}$ are covariance stationary SM sequences.

(ii) $E[\eta_1\eta_1']$ is a positive definite matrix.

The novelty of this regression setting is that the structural postulation (3) of regressors z_t allows for a very wide class of scale factors and means $\{h_t, g_t, \mu_t\}$ which can be both deterministic and stochastic and brings OLS estimation closer to practical work. Estimation procedure permits triangular arrays of the means and scale factors $\big(\mu_t, g_t, h_t,\ t = 1, ..., n\big) = \big(\mu_{nt}, g_{nt}, h_{nt},\ t = 1, ..., n\big)$ - they may vary with n. We skip the subindex n for the brevity of notation.

Denote for $k = 1, ..., p$,

$$v_k^2 = \sum_{t=1}^n g_{kt}^2 h_t^2, \quad v_{gk}^2 = \sum_{t=1}^n g_{kt}^2. \tag{5}$$

We will write $a_n \asymp b_n$ if $a_n = O_p(b_n)$ and $b_n = O_p(a_n)$.

Assumption 2.3. *The scale factors $h_t \geq 0$ and $g_t \geq 0$ are deterministic or stochastic non-negative variables such that for $k = 1, ..., p$,*

$$\frac{\max_{1\leq t\leq n} g_{kt}^2}{v_{gk}^2} = o_p(1), \quad \frac{\max_{1\leq t\leq n} \mu_{kt}^2}{v_{gk}^2} = o_p(1), \tag{6}$$

$$\frac{\sum_{t=1}^n \mu_{kt}^2}{v_{gk}^2} = O_p(1), \quad \frac{\sum_{t=1}^n \mu_{kt}^2 h_t^2}{v_k^2} = O_p(1), \quad v_k^2 \asymp_p v_{gk}^2. \tag{7}$$

To estimate the parameter $\beta = (\beta_1, ..., \beta_p)'$, we use the standard OLS estimator

$$\widehat{\beta} = \big(\sum_{j=1}^{n} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{n} z_j y_j\big) \tag{8}$$

based on a sample y_j, z_j, $j = 1, ..., n$.

Consistency. First we establish consistency of the OLS estimator $\widehat{\beta}$. Set $D = \mathrm{diag}(v_1, ..., v_p)$.

Theorem 2.1. *Suppose that $(y_1, ..., y_n)$ is a sample of dependent variable (1) and Assumptions 2.1, 2.2 and 2.3 are satisfied. Then, the OLS estimator $\widehat{\beta}$ is consistent:*

$$D(\widehat{\beta} - \beta) = \big(v_1(\widehat{\beta}_1 - \beta_1), ..., v_p(\widehat{\beta}_p - \beta_p)\big)' = O_p(1). \tag{9}$$

Theorem 2.1 implies that the k-th component $\widehat{\beta}_k$ of the OLS estimator $\widehat{\beta}$ is v_k-consistent: $\widehat{\beta}_k - \beta_k = O_p(v_k^{-1})$. It is worth noting that the rate of convergence of $\widehat{\beta}_k$ may differ across k and depart from the parametric rate $\sqrt{n}$. Observe that $v_k, v_{gk} \geq c\sqrt{n}$ if $g_{kt}, h_t \geq c > 0$ for all t, n.

Assumption 2.3 is satisfied by regressors z_t and noises u_t which have bounded $4 + \delta$ moment.

Lemma 2.1. *Suppose that z_t, u_t are such that $v_k^2 \asymp_p n$, $v_{gk}^2 \asymp_p n$ for $k = 1, ..., p$ and*

$$E z_{kt}^4 \leq c, \quad E|u_t|^{4+\delta} \leq c \;\; \textit{for some } \delta > 0, \tag{10}$$

where $c < \infty$ does not depend on t, n. Then Assumptions 2.3 and 2.4(ii) hold.

It follows directly from the proof of the lemma, that (10) implies $v_k^2 = O_p(n)$, $v_{gk}^2 = O_p(n)$.

Asymptotic normality. Asymptotic normality property for an element $\widehat{\beta}_k$ of the OLS estimator and computation of standard errors requires additional assumptions on scale factors and stationary processes $\{\eta_t, \varepsilon_t\}$.

Assumption 2.4. *(i) For $k, j = 1, ..., p$, $\{\varepsilon_t^2\}$, $\{\eta_{jt}\eta_{kt}\varepsilon_t^2\}$ and $\{\eta_{jt}\varepsilon_t^2\}$ are covariance stationary SM sequences. (ii) For $k = 1, ..., p$,*

$$\frac{\max_{1\leq t\leq n} g_{kt}^2 h_t^2}{v_k^2} = o_p(1), \quad \frac{\max_{1\leq t\leq n} \mu_{kt}^2 h_t^2}{v_k^2} = o_p(1). \tag{11}$$

We will describe standard errors using the following notation.

$$\begin{aligned} &S_{zz} = \textstyle\sum_{t=1}^n z_t z_t', \quad S_{zzuu} = \textstyle\sum_{t=1}^n z_t z_t' u_t^2, \\ &\Omega_n = (E[S_{zz}\,|\mathcal{F}_n^*])^{-1} E[S_{zzuu}\,|\mathcal{F}_n^*](E[S_{zz}\,|\mathcal{F}_n^*])^{-1} = (\omega_{jk}). \end{aligned} \tag{12}$$

Generality of our regression setting limits asymptotic theory for $\widehat{\beta}_t$ we can establish. While multivariate asymptotic theory for $\widehat{\beta}_t$ is not available, we can derive the asymptotic normality property for linear combinations $a'\widehat{\beta}$ of elements of $\widehat{\beta}$ and we can build feasible asymptotic theory for components β_k of β.

Theorem 2.2. *Suppose that assumptions of Theorem 2.1 are satisfied and Assumption 2.4 holds. Then, for any $a = (a_1, ..., a_p)' \neq 0$, the OLS estimator $\widehat{\beta}$ has property:*

$$\frac{a'D(\widehat{\beta} - \beta)}{\sqrt{a'D\Omega_n Da}} \to_d \mathcal{N}(0, 1). \tag{13}$$

The t-statistic for the parameter β_k, $k = 1, ..., p$ satisfies

$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\omega_{kk}}} \to_d \mathcal{N}(0, 1). \tag{14}$$

Property (13) does not appear to be workable since it requires estimation of D, Ω_n. One practical application of (13) is that it implies (14), where standard error $\sqrt{\omega_{kk}}$ can be consistently estimated by computing

$$\widehat{\Omega}_n = S_{zz}^{-1} S_{zz\widehat{u}\widehat{u}} S_{zz}^{-1} = (\widehat{\omega}_{jk}), \quad \widehat{u}_t = y_t - \widehat{\beta}' z_t. \tag{15}$$

Corollary 2.1. *Under the assumption of Theorem 2.2, for $k = 1, ..., p$, as $n \to \infty$,*

$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} \to_d \mathcal{N}(0, 1), \quad \frac{\widehat{\omega}_{kk}}{\omega_{kk}} = 1 + o_p(1), \quad \sqrt{\omega_{kk}} \asymp_p v_k^{-1}. \tag{16}$$

This result is the main contribution of Section 2. It enables an easy computation of standard errors and building confidence intervals for the regression parameters β_k contained in $\beta = (\beta_1, ..., \beta_p)'$. The novelty of this finding is that the order of the standard error $\sqrt{\omega_{kk}}$ may differ from the standard one $n^{-1/2}$, they may be random, and they do not require to exhibit asymptotic behaviour, i.e. to be proportional to $n^{-1/2}$.

Notable, the estimator $\widehat{\Omega}_n$ of robust standard errors in (12) coincides with the well-known estimator of heteroskedasticity-consistent standard errors by White (1980). The regular estimator of standard errors in OLS regression estimation is:

$$\widehat{\Omega}_n^{(st)} = S_{zz}^{-1} \widehat{\sigma}_u^2, \quad \widehat{\sigma}_u^2 = n^{-1} \sum_{j=1}^{n} \widehat{u}_j^2. \tag{17}$$

Differently from the robust standard errors $\sqrt{\widehat{\omega}_{kk}}$, they produce size distortions in estimation of β_k in regression model (1), see Section 6. This emphasizes the flexibility of the robust standard errors and good performance of the normal approximation (16) even when we divert to very complex regression models with arbitrary scale factors g_t, h_t and means μ_t.

3 Time-varying OLS estimation under general heterogeneity

In this section we extend the regression setting (1) to allow for time-varying parameter β_t. We consider the model

$$y_j = \beta_j' z_j + u_j, \quad j = 1, ..., n, \tag{18}$$

where regressors z_j and regression noise u_j are as in (3) and (2) and permit the same degree of heterogeneity as the model (1).

Of primary concern is to develop a point-wise estimation procedure for the path $\beta_1, ..., \beta_n$ of the time-varying parameter β_t in model (18). While $\{\eta_j, \varepsilon_j\}$ and $\{\mu_j, g_j, h_j\}$ will remain the same as in Section 2, the estimator and assumptions require some amendments.

The idea of estimation of time-varying parameter β_t is rather straightforward and will be reduced to large extent to estimation theory for regression model with a fixed parameter. Assume that there are kernel weights $b_{n,tj} \geq 0$ such that $b_{n,tj} \to 0$ as $|t-j| \to \infty$. We multiply both sides of (18) by $b_{n,tj}^{1/2}$, to obtain a regression model:

$$\widetilde{y}_j \quad = \quad \beta_j' \widetilde{z}_j + \widetilde{u}_j, \quad j = 1, ..., n,$$

where $\widetilde{y}_j = b_{n,tj}^{1/2} y_j$, $\widetilde{z}_j = b_{n,tj}^{1/2} z_j$ and $\widetilde{u}_j = b_{n,tj}^{1/2} u_j$. With this, we do not change the structure of regressors $\widetilde{z}_t = (\widetilde{z}_{1t}, ..., \widetilde{z}_{pt})'$ and $\widetilde{u}_t$, and only multiply the means μ_t and scales g_t, h_t by $b_{n,tj}^{1/2}$ in settings (3) and (2).

This leads to a regression model with a fixed parameter β_t:

$$\widetilde{y}_j = \beta_t' \widetilde{z}_j + \widetilde{u}_j + r_j, \quad j = 1, ..., n \tag{19}$$

which includes an additional error term $r_j = (\beta_j' - \beta_t')\tilde{z}_j$. It is rather easy to show that the term r_j is negligible because of "smoothness" of parameter β_j when j is close to t, and because of the down-weighting of data when j is distant from t. It follows directly from (19) that β_t can be estimated by the OLS estimator

$$\widehat{\beta}_t = \Big(\sum_{j=1}^{n} \tilde{z}_j \tilde{z}_j'\Big)^{-1} \Big(\sum_{j=1}^{n} \tilde{z}_j \tilde{y}_j\Big).$$

Asymptotic theory for $\widehat{\beta}_t$ will succeed from the estimation theory for fixed parameter in Section 2 by setting $r_j = 0$ in (19).

Returning to the notation of the original regression (18), we obtain the following time-varying OLS estimator of the regression parameter β_t:

$$\widehat{\beta}_t \quad = \quad \Big(\sum_{j=1}^{n} b_{n,tj} z_j z_j'\Big)^{-1} \Big(\sum_{j=1}^{n} b_{n,tj} z_j y_j\Big). \tag{20}$$

The weights $b_{n,tj}$ are generated like this. Set

$$b_{n,tj} = K(\frac{|t-j|}{H}), \ t, j = 1, ..., n, \tag{21}$$

where $H = H_n$ is a bandwidth parameter such that $H \to \infty$ and $H = o(n)$. The kernel function K is bounded and there exist a_0, $\delta > 0$ and $\theta > 3$ such that

$$\begin{aligned} K(x) &\geq a_0 > 0, \ 0 \leq x \leq \delta, \\ K(x) &\leq Cx^{-\theta}, \ x > \delta. \end{aligned} \tag{22}$$

For example, (22) is satisfied by functions $K(x) = I(x \in [0,1])$ and $K(x) = p(x)$ where $p(x)$ is probability density of the standard normal distribution.

We need a smoothness assumption on the time-varying parameter β_j which can be deterministic or stochastic.

Assumption 3.1. *For some $\gamma \in (0,1]$, for $t, j = 1, ..., n$ it holds*

$$E||\beta_t - \beta_j||^2 \leq c(\frac{|t-j|}{n})^{2\gamma}, \tag{23}$$

where $c < \infty$ does not depend on t, j, n.

Regressors z_t and standard errors have the same structure as in Section 2. While assumptions on stationary process $\{\eta_t\}$ and m.d. noise $\{\varepsilon_t\}$ remain the same as in Section 2, for simplicity of presentation, we replace complex assumptions on the scale factors g_t, h_t and the means μ_t by simple sufficient assumptions similar to those used in Lemma 2.1. As before, scale factors $\{h_t, g_t, \mu_t\}$ can be deterministic or stochastic, they may vary with n and they are independent of $\{\eta_t, \varepsilon_t\}$.

Denote

$$v_{kt}^2 = \textstyle\sum_{j=1}^n b_{n,tj}^2 g_{kj}^2 h_j^2, \ \ v_{gk,t}^2 = \sum_{j=1}^n b_{n,tj}^2 g_{kj}^2, \quad k = 1, ..., p.$$

Assumption 3.2. *For $k = 1, ..., p$, $v_{kt}^2 \asymp_p H$, $v_{gk,t}^2 \asymp_p H$ and*

$$Ez_{kt}^4 \leq c, \quad E|u_t|^{4+\delta} \leq c \textit{ for some } \delta > 0, \tag{24}$$

where $c < \infty$ does not depend on t, n.

The proof of Lemma 2.1 shows that (24) implies $v_{kt}^2 = O_p(H)$, $v_{gk,t}^2 = O_p(H)$.

To describe standard errors, we will use the following notation.

$$S_{zz,t} = \textstyle\sum_{j=1}^n b_{n,tj} z_j z_j', \quad S_{zzuu,t} = \sum_{j=1}^n b_{n,tj}^2 z_j z_j' u_j^2,$$

$$\Omega_{nt} = E[S_{zz,t}]^{-1}E[S_{zzuu,t}]E[S_{zz,t}]^{-1} = (\omega_{jk,t}).$$

The next theorem establishes consistency rate and the asymptotic normality property for the components of the time-varying OLS estimator $\widehat{\beta}_t = (\widehat{\beta}_{1t}, ..., \widehat{\beta}_{pt})'$.

Theorem 3.1. *Suppose that $(y_1, ..., y_n)$ is a sample from a regression model (18). Assume that Assumptions 2.1, 2.2, 3.1 and 3.2 are satisfied. Then, for $1 \le t = t_n \le n$ and $k = 1, ..., p$, the following holds:*

$$\widehat{\beta}_{kt} - \beta_{kt} = O_p\big(H^{-1/2} + (H/n)^{\gamma}\big), \tag{25}$$

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} \to_d \mathcal{N}(0,1) \quad \text{if } H = o(n^{2\gamma/(2\gamma+1)}), \tag{26}$$

and $\sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}$.

The unknown standard errors $\sqrt{\omega_{kk,t}}$ can be estimated using the estimator

$$\widehat{\Omega}_{nt} = S_{zz,t}^{-1}S_{zz\widehat{u}\widehat{u},t}S_{zz,t}^{-1} = (\widehat{\omega}_{jk,t}), \quad \widehat{u}_j = y_j - \widehat{\beta}_j' z_j. \tag{27}$$

Corollary 3.1. *Under assumption of Theorem 3.1, for $k = 1, ..., p$, assuming that $H = o(n^{2\gamma/(2\gamma+1)})$ it holds*

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} \to_d \mathcal{N}(0,1), \quad \frac{\widehat{\omega}_{kk,t}}{\omega_{kk,t}} = 1 + o_p(1). \tag{28}$$

Computation of standard errors $\sqrt{\widehat{\omega}_{kk,t}}$ is straightforward. It is worth noting that under general heterogeneity the scale factors h_t, g_t, μ_t in model (18) are unknown, standard errors might be random and the limit of $H^{1/2}\sqrt{\omega_{kk,t}}$ may not exist. The univariate asymptotic normality for a component $\widehat{\beta}_{kt}$ of $\widehat{\beta}_t$ still can be shown, although such general environment does not allow establishing multivariate asymptotic theory for $\widehat{\beta}_t$.

The estimator $\widehat{\Omega}_{nt}$ of robust standard errors in (27) is a time-varying version of heteroskedasticity-consistent estimator of standard errors by White (1980). Simulations confirm that it does not produce coverage distortions in estimation of β_t under settings considered in this section.

Example 3.1. *A typical example of a deterministic time-varying parameter β_t which satisfies Assumption 3.1, is $\beta_t = \beta_{t,n} = g(t/n)$, $t = 1, ..., n$, where $g(\cdot)$ is a deterministic smooth function that has property $|g(x) - g(y)| \le C|x - y|$. Such β_t satisfies (23) with $\gamma = 1$.*

A standard example of a stochastic smooth parameter β_t is a re-scaled random walk $\beta_t = \beta_{t,n} = n^{-1/2}\sum_{j=1}^{t} e_j$, $t = 1, ..., n$, where $\{e_j\}$ is an i.i.d. sequence with $E[e_t] = 0$ and $E[e_j^2] < \infty$. It satisfies (23) with $\gamma = 1/2$, that is for $t > s$,

$$E(\beta_t - \beta_s)^2 \;=\; n^{-1}E(\textstyle\sum_{j=s+1}^{t} e_j)^2 \le C(t-s)/n.$$

4 Regression with missing data

Given the interest in empirical regression analysis, when some observations y_t or regressors z_t might be missing, see, e.g., Enders (2022), we present below somewhat unexpected results on regression estimation showing that the results of Section 2 and 3 allow to cover various settings of missing data patterns.

Estimation of fixed parameter. Suppose that y_t follows regression model (1) with fixed parameter β, regressors z_t as in (3) and regression noise u_t as in (2) of Section 2. Our primary interest is to estimate parameter β using subsample of y_t, z_t's:

$$(y_{j_1}, z_{j_1}), ..., (y_{j_N}, z_{j_N}), \quad N \le n.$$

To that purpose, we introduce missing data indicator τ_t, $t = 1, ..., n$: we set $\tau_t = 1$ if both y_t and z_t are observed, otherwise $\tau_t = 0$. Overall, missing data indicator τ_t is a sequence of random or deterministic variables, i.e. the indicator of regularly missing, block-wise missing or randomly missing data. Then, setting $(y_t, z_t) = (0, 0)$ for any time period t, where either y_t or z_t is missing, we arrive at a sample $\widetilde{y}_1, ..., \widetilde{y}_n$, $\widetilde{y}_t = \tau_t y_t$, from regression model

$$\widetilde{y}_t = \beta' \widetilde{z}_t + \widetilde{u}_t, \quad t = 1, ..., n, \tag{29}$$

where regressors $\widetilde{z}_t$ and the noise $\widetilde{u}_t$ are obtained from z_t and u_t by multiplying the means and scale factors by τ_t:

$$\begin{aligned} \widetilde{z}_{kt} &= \widetilde{\mu}_{kt} + \widetilde{g}_{kt}\eta_{kt}, \quad \widetilde{\mu}_{kt} = \tau_t \mu_{kt}, \quad \widetilde{g}_{kt} = \tau_t g_{kt}, \\ \widetilde{u}_t &= \widetilde{h}_t \varepsilon_t, \quad \widetilde{h}_t = \tau_t h_t. \end{aligned} \tag{30}$$

We will impose the following assumptions on the missing data indicator, z_t, u_t and scale factors g_t, h_t which imply that the model (29) is covered by the setting (1) of Section 2.

Assumption 4.1. *Missing data indicator $\{\tau_t\}$ is independent of $\{\varepsilon_t, \eta_t\}$ in (2) and (3).*

Assumption 4.2. (i) *Regressors z_t and the noise u_t satisfy moment assumption (10), and there exists $c > 0$ such that scale factors $g_{kt} \ge c > 0$ and $h_t \ge c > 0$ are bounded away from zero, where c does not depend on k, t, n.*

(ii) *ε_t, η_t satisfy Assumptions 2.1, 2.2, and 2.4(i).*

In such a case we are able to estimate regression model with missing data.

Proposition 4.1. *Suppose that $\widetilde{y}_1, ..., \widetilde{y}_n$ is a sample of regression model (29) and Assumptions 4.1 and 4.2 are satisfied. Assume that $N = \sum_{t=1}^n \tau_t \asymp_p n$. Then, the OLS estimator $\widehat{\beta}$, based on $\widetilde{y}_1, ..., \widetilde{y}_n$ and $\widetilde{z}_1, ..., \widetilde{z}_n$ fulfills properties (16) of Corollary 2.1 and $\sqrt{\omega_{kk}} \asymp_p n^{-1/2}$.*

It is worth noting that missing data does not have direct impact on the estimator $\widehat{\beta}$ nor the estimates of the standard errors $\sqrt{\widehat{\omega}_{kk}}$. Proposition 4.1 shows that ignoring missing data does not impact estimation. That is, the researcher can estimate parameter β and standard errors $\sqrt{\omega_{kk}}$ using estimators based on a subsample y_{j_t}, z_{j_t}, $t = 1, ..., N$,

$$\begin{aligned} \widetilde{\beta} &= \Big(\sum_{t=1}^{N} z_{j_t} z_{j_t}'\Big)^{-1} \Big(\sum_{t=1}^{N} z_{j_t} y_{j_t}\Big), \quad \widetilde{\Omega}_n = \widetilde{S}_{zz}^{-1} \widetilde{S}_{zz\widehat{u}\widehat{u}} \widetilde{S}_{zz}^{-1} = (\widetilde{\omega}_{jk}), \\ &\widetilde{S}_{zz} = \textstyle\sum_{t=1}^{N} z_{j_t} z_{j_t}', \quad \widetilde{S}_{zz\widehat{u}\widehat{u}} = \sum_{t=1}^{N} z_{j_t} z_{j_t}' \widehat{u}_{j_t}^2, \quad \widetilde{u}_{j_t} = y_{j_t} - \widehat{\beta}' z_{j_t}. \end{aligned}$$

These estimates have property, $\widetilde{\omega}_{kk}^{-1/2}(\widetilde{\beta}_k - \beta_k) \to_d \mathcal{N}(0, 1)$.

Estimation of time-varying parameter. Assume that y_t follows regression model (18) with time-varying parameter β_t, where regressors z_t and regression noise u_t are as in (3) and (2). In such a case, estimation of parameter β_t, $t = 1, ..., n$ from subsample $(y_{j_1}, z_{j_1}), ..., (y_{j_N}, z_{j_N})$, $N \le n$ builds on the results of Section 3. Firstly, we set $(y_t, z_t) = (0, 0)$ for any $t = 1, ..., n$, where either y_t or z_t is not observed, to obtain a sample $\widetilde{y}_1, ..., \widetilde{y}_n$, $\widetilde{y}_t = \tau_t y_t$, from regression model

$$\widetilde{y}_j = \beta_j' \widetilde{z}_j + \widetilde{u}_j, \quad j = 1, ..., n, \tag{31}$$

where regressors $\widetilde{z}_t$ and the noise $\widetilde{u}_t$ are defined as in (30). Under Assumptions 4.1 and 4.2, parameter β_t, $t = 1, ..., n$ can be estimated by the estimator $\widehat{\beta}_t$ given in (20) with kernel weights $b_{n,tj}$ as in (21), as long as the missing data pattern has property:

$$N_t = \sum_{j=1}^{n} \tau_j b_{n,tj} \asymp_p H. \tag{32}$$

The latter holds, if e.g. $\tau_j = 1$ for $|j - t| \le \epsilon H$ for some $\epsilon > 0$.

Proposition 4.2. *Suppose that $\widetilde{y}_1, ..., \widetilde{y}_n$ is a sample of regression model (31) with missing data and Assumptions 4.1, 3.1 and 4.2 are satisfied. Assume that $1 \le t = t_n \le n$ and (32) holds. Then $\widehat{\beta}_{kt}$, $k = 1, ..., p$ satisfy properties (25) and (26) of Theorem 3.1, $\sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}$, and Corollary 3.1 holds.*

5 Estimation of a stationary AR(p) model with an m.d. noise

The robust OLS estimation theory of Section 2 extends to estimation of a stationary AR(p) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t, \tag{33}$$

where ε_t is a stationary martingale difference noise, and parameters $\phi_0, ..., \phi_p$ are such that (33) has a stationary solution $y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} |a_j| < \infty$ and $\mu = E y_t$. We assume that ε_t satisfies Assumption 2.1. This model can be written as regression model (1), $y_t = \beta' z_t + \varepsilon_t$, where $\beta = (\beta_1, ..., \beta_{p+1})' = (\phi_0, ..., \phi_p)'$ is a fixed parameter and regressors $z_t = (z_{1t}, z_{2t}, ..., z_{p+1,t})' = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})'$ are stationary variables. The theoretical results of Section 2 allow to estimate the parameter β by the robust OLS estimator $\widehat{\beta}$ and permit the evaluation of the robust standard errors.

Theorem 5.1. *Suppose that $(y_1, ..., y_n)$ is a sample from a stationary AR(p) model (33) and $E\varepsilon_t^8 < \infty$. Then, the estimates $\widehat{\beta}_k$, $k = 1, ..., p+1$ satisfy (16) and $\sqrt{\omega_{kk}} \asymp_p n^{-1/2}$.*

Monte Carlo findings in Section 6.4 show that the robust OLS estimation produces correct 95% confidence intervals for β_k while the standard OLS method leads to coverage distortions, when the noise ε_t is not i.i.d. This suggests that robust OLS estimation has a wider range of applications, than just heteroscedasticity, and is applicable for regression settings not covered by the standard OLS estimation and inference theory.

The asymptotic theory for Whittle estimators of parameters of a stationary ARMA model with a stationary m.d. noise is provided in Giraitis et al. (2018). They do not consider estimation of standard errors and constrain the practical implementation of their results to AR(1) and MA(1) models.

6 Monte Carlo Simulations

In this section, we explore the finite sample performance of the robust and standard OLS estimation methods in regression settings, outlined in Sections 2 and 3. We examine the impact of time-varying deterministic and stochastic parameters, means, scale factors and heteroskedasticity of regression noise on estimation. Comparison of simulation results for standard and robust estimation methods shows that, despite the generality of our regression setting, estimation based on the robust standard errors produces well-sized coverage intervals for fixed and time-varying regression parameters β, β_t, while application of the standard confidence intervals leads to severe distortion of coverage rates.

6.1 Estimation of fixed parameters

We generate arrays of samples of regression model with fixed parameter and an intercept:

$$y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + u_t, \quad u_t = h_t \varepsilon_t, \quad \beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.4, 0.3)'. \qquad (34)$$

We set the sample size to $n = 1500$ and conduct 1000 replications and set the nominal coverage probability at 0.95. (Estimation results for $n = 200, 800$ are available upon request).

This model includes three parameters and three regressors. We set $z_{1t} = 1$ and define

$$z_{kt} = \mu_{kt} + g_{kt} \eta_{kt}, \quad k = 2, 3, \qquad (35)$$

Table 1: Robust OLS estimation in Model 6.1.

Parameters	Bias	RMSE	CP	CP_{st}	SD
β_1	-0.00570	0.04579	95.0	79.2	0.04544
β_2	0.00206	0.03407	95.4	72.7	0.03401
β_3	0.00204	0.03495	94.0	72.9	0.03489

$$\mu_{kt} = 0.5\sin(\pi t/n) + 1, \quad \eta_{kt} = 0.5\eta_{k,t-1} + \xi_{kt},$$

where $\xi_{2t} = \varepsilon_{t-1}$ and $\xi_{3t} = \varepsilon_{t-2}$. The stationary martingale difference noise ε_t in u_t is generated by an GARCH$(1,1)$ process

$$\varepsilon_t = \sigma_t e_t, \quad \sigma_t^2 = 1 + 0.7\sigma_{t-1}^2 + 0.2\varepsilon_{t-1}^2, \quad e_t \sim i.i.d.\,\mathcal{N}(0,1). \tag{36}$$

Model 6.1. *y_t follows (34) with deterministic scale factors. We set: $h_t = 0.3(t/n)$ and $g_{2,t} = g_{3,t} = 0.4(t/n)$.*

Model 6.2. *y_t follows (34) with stochastic scale factors. We set*

$$h_t = \Big|\frac{1}{2\sqrt{n}}\sum_{j=1}^{t}\zeta_j\Big| + 0.25, \quad g_{2t} = g_{3t} = \Big|\frac{1}{2\sqrt{n}}\sum_{j=1}^{t}\nu_{kj}\Big| + 0.25.$$

The generating noises $\{\zeta_j, \nu_{2j}, \nu_{3j}\}$ are i.i.d. $\mathcal{N}(0,1)$ and independent of $\{\varepsilon_j\}$.

Models 6.1 and 6.2 are regression models with fixed parameters. To verify the validity of the asymptotic normal approximation of Corollary 2.1 in finite samples, we compute empirical coverage rates (CP) for 95% confidence intervals used in robust OLS estimation, for parameter β. For comparison, we compute the coverage rates CP_{st} for standard confidence intervals based on the standard errors (17) used in the standard OLS estimation. The robust and standard OLS procedures share the same estimator $\widehat{\beta}$, and whence Bias, root mean square error (RMSE) and standard deviation (SD). Their confidence intervals are based on different standard errors, as the variances in their normal approximation are different.

Table 1 reports estimation results for Model 6.1 which contains determinist scale factors. It shows that coverage rate CP for robust confidence intervals is close to the nominal 95%, while coverage rate CP_{st} of the standard confidence intervals drops below 80%. The Bias, RMSE, and SD are small.

Table 2 shows estimation results for Model 6.2 which includes stochastic scale factors. It shows that coverage rate CP for robust confidence intervals is close to the nominal 95%, while standard estimation method produces coverage distortions for parameters β_2 and β_3.

Table 2: Robust OLS estimation in Model 6.2.

Parameters	Bias	RMSE	CP	CP_{st}	SD
β_1	-0.00420	0.05117	94.6	92.2	0.05100
β_2	0.00208	0.03205	94.6	87.4	0.03199
β_3	0.00071	0.01542	94.8	85.3	0.01541

6.2 Estimation of time-varying parameters

In this section we examine the validity of the normal approximation for the estimator $\widehat{\beta}_t$, (20), of time-varying parameter β_t established in Corollary 3.1 of Section 3. Subsequently, we replace the fixed regression parameter β in the regression model (34) by a time-varying parameter $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$:

$$y_t = \beta_{1t} + \beta_{2t} z_{2t} + \beta_{3t} z_{3t} + u_t, \quad u_t = h_t \varepsilon_t, \tag{37}$$

where $z_{1t} = 1$ and z_{2t}, z_{3t} are defined with μ_{2t}, μ_{3t} and η_{2t}, η_{3t} as in (35).

We consider two models. In Model 6.3, parameters and scale factors are deterministic. Model 6.4 combines deterministic and stochastic parameters and scale factors.

Model 6.3. *y_t follows (37) with ε_t as in (36). The scale factors h_t, g_{2t}, g_{3t} and parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ are deterministic:*

$$\begin{aligned} h_t &= 0.5\sin(2\pi t/n) + 1, \quad g_{2t} = g_{3t} = 0.5\sin(\pi t/n) + 1. \\ \beta_{1t} &= 0.5\sin(0.5\pi t/n) + 1, \quad \beta_{2t} = 0.5\sin(\pi t/n) + 1, \quad \beta_{3t} = 0.5\sin(2\pi t/n) + 1. \end{aligned}$$

Model 6.4. *y_t follows (37) with $\varepsilon_t \sim i.i.d.\,\mathcal{N}(0,1)$ and scale factors:*

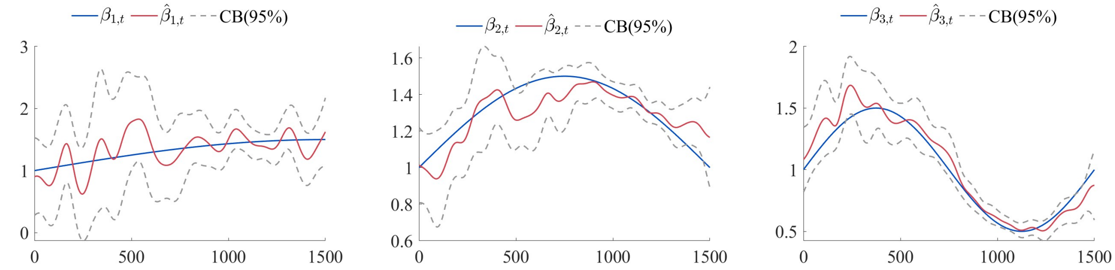
$$h_t = 0.5\sin(2\pi t/n) + 1, \quad g_{2t} = \Big|\frac{1}{2\sqrt{n}}\sum_{j=1}^{t} \zeta_j\Big| + 0.25, \quad g_{3t} = 0.5\sin(\pi t/n) + 1.$$

Parameters β_{1t}, β_{2t} are the same as in Model 6.3, and parameter β_{3t} is stochastic:

$$\beta_{3t} = \Big|\frac{1}{2\sqrt{n}}\sum_{j=1}^{t} \nu_j\Big| + 0.3(t/n), \qquad \zeta_j, \nu_j \sim i.i.d.\,\mathcal{N}(0,1).$$

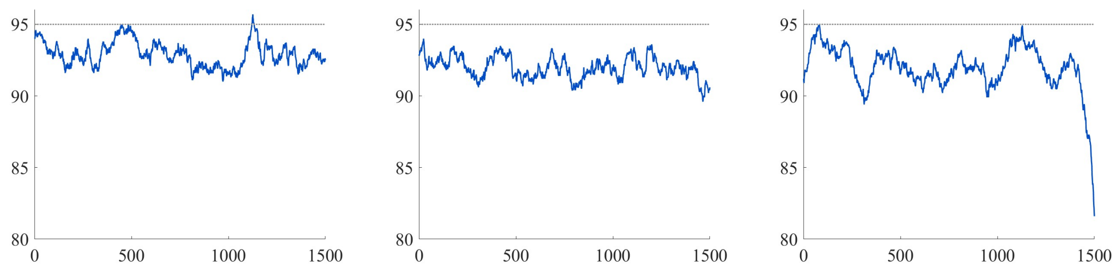
We estimate β_t through the estimator $\widehat{\beta}_t$, (20), where the weights $b_{n,tj} = K(|t-j|/H)$ are computed using the Gaussian kernel function $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ with the bandwidth $H = n^h$, $h = 0.4, 0.5, 0.6, 0.7$.

Figure 1 displays parameter estimation results for a single simulation from Model 6.3. It depicts the estimates $\widehat{\beta}_{k1}, ..., \widehat{\beta}_{kn}$ (red line) of the true parameters β_{kt} (blue line), $k = 1, 2, 3$ obtained with the bandwidth $H = n^{0.5}$, and their point-wise 95% confidence intervals (grey



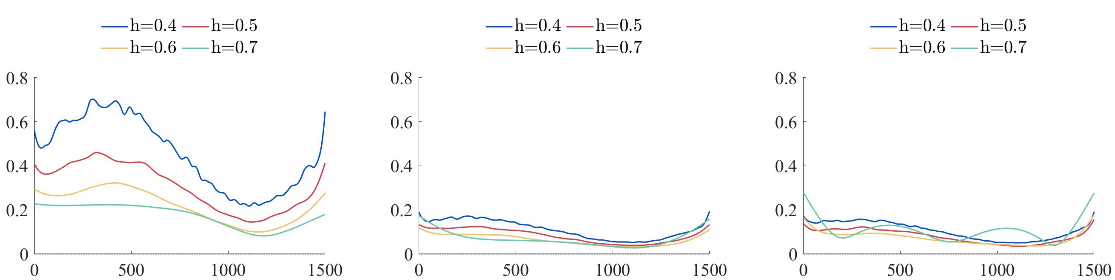
(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 1: Robust 95% confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3: $n = 1500$, bandwidth $H = n^{0.5}$. Single replication.



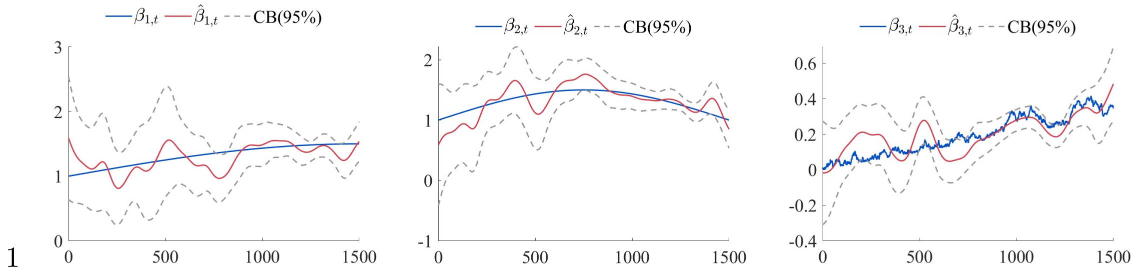
(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 2: Coverage rates (in %) of robust confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3: $n = 1500$, bandwidth $H = n^{0.5}$.



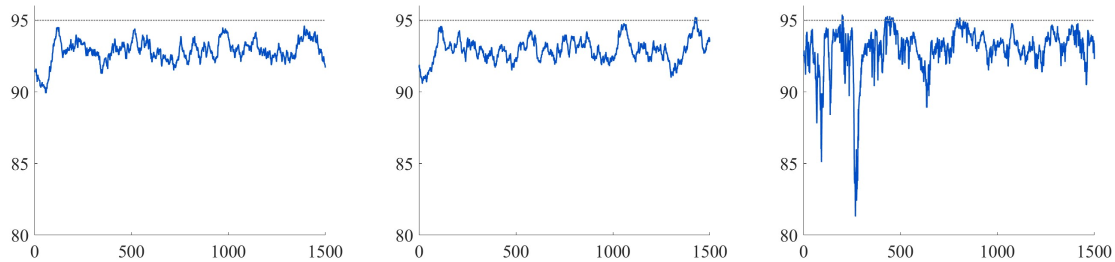
(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 3: RMSE for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3: $n = 1500$, bandwidth $H = n^h$, $h = 0.4, 0.5, 0.6, 0.7$.



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 4: Robust 95% confidence bands for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.4: $n = 1500$, bandwidth $H = n^{0.5}$. Single replication.



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 5: Coverage rates (in %) of robust confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.4: $n = 1500$, bandwidth $H = n^{0.5}$.

dashed lines), computed using robust standard errors. The robust time-varying confidence intervals cover the true parameters β_{kt}, $t = 1, ..., n$ for most of the times.

Figure 2 reports the point-wise empirical coverage rates (blue line) in time-varying robust estimation of parameters $\beta_{kt}, k = 1, 2, 3$ which are close to the nominal 95% for most of the times. Figure 3 shows the RMSE's for different choices of the bandwidth $H = n^h$, $h = 0.4, 0.5, 0.6, 0.7$. As expected, the RMSE depends on the smoothness of the parameter β_{kt} and often is minimized by moderately large values of H, e.g. $H = n^{0.6}$.

Figure 4 reports estimation results for a single simulation from Model 6.4, and Figure 5 displays point-wise empirical coverage rates for robust 95% confidence intervals. For deterministic parameters β_{1t} and β_{2t}, estimation quality is good and results are similar to those obtained for Model 6.3. For the stochastic parameter β_{3t}, the robust point-wise confidence intervals cover the path of stochastic parameter β_{3t} for most of the times, see Figure 4(c). Figure 5(c) shows that coverage rates of robust time-varying confidence intervals for β_{3t} might be slightly affected by stochastic changes in parameter and scale factors. Nevertheless, they are still satisfactory and not far from the nominal 95% coverage.

Table 3: Robust OLS estimation in Model 6.1 with block missing data (Type 1).

Parameters	Bias	RMSE	CP	CP_{st}	SD
β_1	-0.00818	0.04983	94.60	74.60	0.04915
β_2	0.00356	0.03875	94.00	67.90	0.03859
β_3	0.00246	0.03840	93.80	70.00	0.03832

Table 4: Robust OLS estimation in Model 6.1 with randomly missing data (Type 2).

Parameters	Bias	RMSE	CP	CP_{st}	SD
β_1	-0.00567	0.05732	94.30	66.60	0.05704
β_2	0.00144	0.04251	95.20	63.50	0.04249
β_3	0.00289	0.04128	94.80	64.70	0.04118

6.3 Estimation of regression parameter with missing data

To examine the impact of missing data on the robust and standard OLS estimation based on partially observed data $(y_{j_1}, z_{j_1}), (y_{j_2}, z_{j_2}), ..., (y_{j_N}, z_{j_N})$, we use two types of missing data patterns over the time period $1, ..., 1500$.

Type 1. The block of data $j \in [650, 850]$ is missing.

Type 2. 500 single observations are missing at randomly selected times.

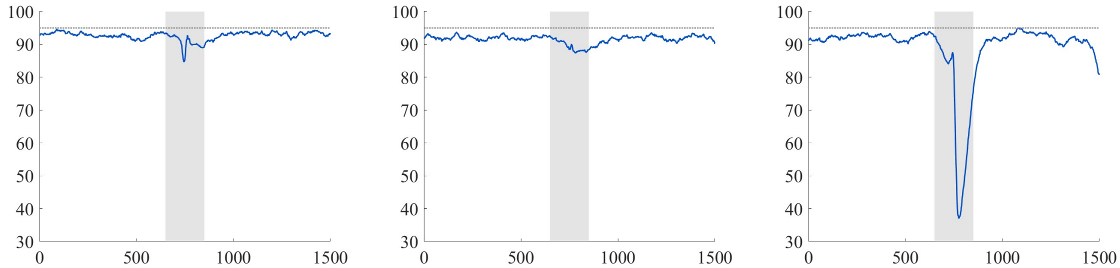
Tables 3 and 4 report robust and standard estimation results for Model 6.1 with fixed parameter. Table 3 shows that block missing data (Type 1) do not lead to visible changes of Bias, RMSE and SD, and coverage rate for robust confidence intervals is still around 95%. At the same time, coverage rate CP_{st} of the standard confidence intervals is significantly distorted.

Table 4 shows that randomly missing data do not affect the coverage rate of robust confidence intervals which is close to the nominal 95%, while the coverage rate of the standard confidence intervals hovers around 65%. This emphasises the flexibility of the robust OLS estimation of the fixed parameter in presence of block missing data

Figure 6 – 8 report estimation results for Model 6.3 with time-varying parameter β_t.

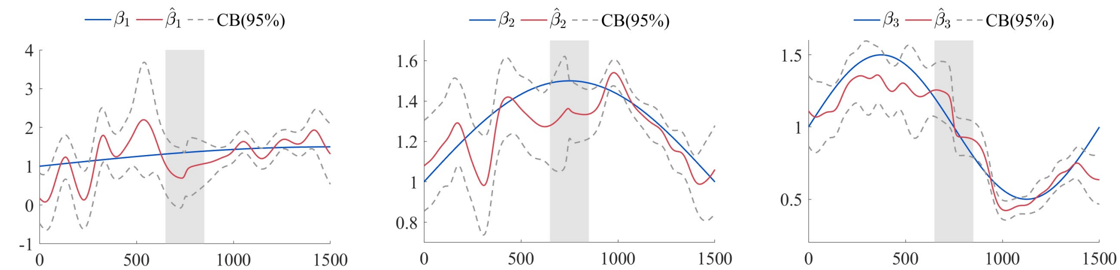
Figure 6 shows the coverage rates in time-varying robust estimation with block missing data (Type 1, shaded region) for $t = 1, ..., 1500$. The coverage is close to the nominal 95%, with some distortion for parameters $\beta_{1,t}$ and $\beta_{2,t}$ and larger distortion for parameter $\beta_{3,t}$ in the shaded region. The distortion peaks at the centre of the block, as expected. Although the width of missing data block, 200, exceeds the bandwidth $H = n^{0.5} = 39$ used in the estimation of β_t, the coverage distortion seems to be offset by the smooth down-weighting of data and the performance of the robust time-varying OLS estimation exceeds our expectations.

Figure 7 reports the path of the estimator $\widehat{\beta}_{kt}$ and the point-wise robust confidence



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 6: Coverage rates (in %) of robust confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3 with block missing data (Type 1), $n = 1500$, bandwidth $H = n^{0.5}$.



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 7: Robust 95% confidence bands for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3 with block missing data (Type 1), $n = 1500$, bandwidth $H = n^{0.5}$. Single replication.

intervals, for a single simulation. The robust confidence intervals become wider in the shaded region, which may explain the satisfactory coverage performance in that time period.

Figure 8 shows that randomly missing data (Type 2) does not distort the robust time-varying OLS estimation. For all three parameters and time periods t, the coverage rate is close to the nominal. Overall, the robust estimation of time-varying parameter does not seem to be affected by randomly missing data.

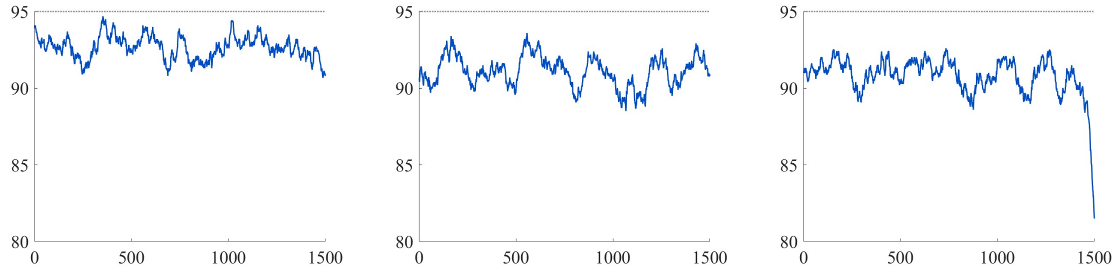
6.4 Estimation of a stationary AR(p) model

We assess the performance of the robust and standard procedures in a case of a stationary AR(2) model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t, \quad \beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.4, 0.3)', \tag{38}$$

where $\varepsilon_t = e_t e_{t-1}$, $e_t \sim i.i.d. \mathcal{N}(0, 1)$ is a stationary martingale difference noise. Here regressors $z_t = (z_{1,t}, z_{2,t}, z_{3,t})' = (1, y_{t-1}, y_{t-2})'$ include an intercept and the two past lags of y_t. By Theorem 5.1, parameter β can be estimated by the robust estimation method.

Table 5 shows that the coverage rate for the robust OLS estimation method is close to the nominal 95% while the standard OLS estimation leads to extensive coverage distortion



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 8: Coverage rates (in %) of robust confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3, 500 randomly missing data, $n = 1500$, bandwidth $H = n^{0.5}$.

Table 5: Robust OLS estimation in $AR(2)$ model (38).

Parameters	Bias	RMSE	CP	CP_{st}	SD
β_1	-0.00808	0.05250	94.9	92.3	0.05187
β_2	0.00104	0.04183	94.5	75.0	0.04182
β_3	0.00356	0.03091	94.8	88.8	0.03070

for β_2 and β_3.

7 Empirical experiment

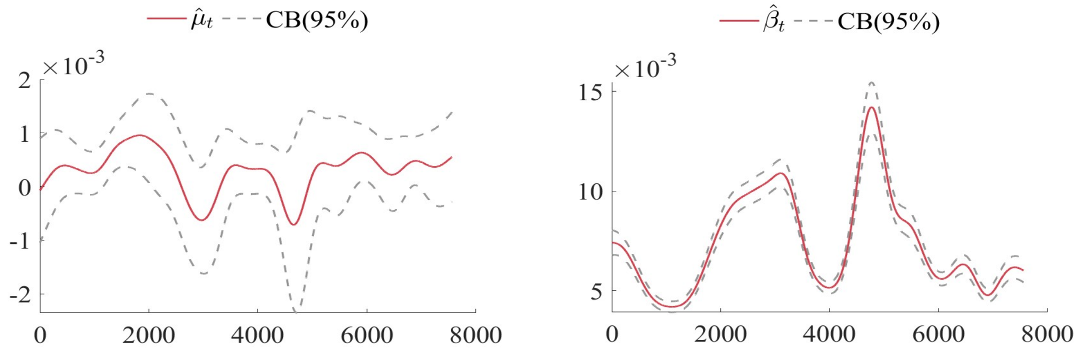
In this section, we consider the problem of the structure and modelling of daily S&P 500 log returns, r_t, from 02/01/1990 to 31/12/2019, (sample size $n = 7558$). We use robust regression estimation to verify whether returns r_t obey the following regression model for uncorrelated noise:

$$r_t = \mu_t + u_t, \quad u_t = h_t \varepsilon_t, \tag{39}$$

where $\{\varepsilon_t\}$ is an i.i.d.$(0, 1)$ noise and the time-varying mean and scale factor μ_t, h_t are independent of $\{\varepsilon_t\}$. Our objective is to estimate μ_t (time-varying mean), the scale factor h_t and test for absence of correlation in $|u_t| = h_t|\varepsilon_t|$ to confirm the fit of the model (39) to the data. Indeed, if r_t follows (39) with i.i.d. noise ε_t, then for $t \neq s$, $|u_t|$'s are uncorrelated:

$$\text{cov}(|u_t|, |u_s|) \;=\; \text{cov}(h_t|\varepsilon_t|, h_s|\varepsilon_s|) = E\big[h_t h_s \text{cov}(|\varepsilon_t|, |\varepsilon_s|)\big] = 0.$$

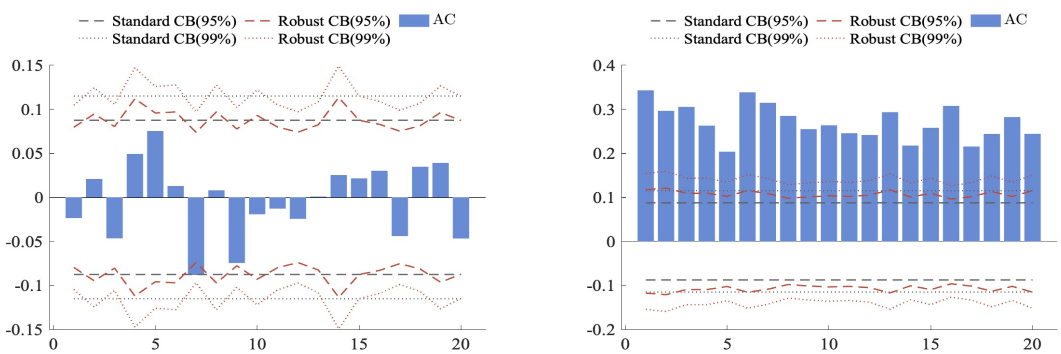
Conversely, under the presence of ARCH effects (stationary conditional heteroskedasticity) in ε_t, $|u_t|$ would be a sequence of correlated variables and the hypothesis of absence of correlation in $|u_t|$ would be rejected. To estimate the intercept μ_t, we have used the time-varying OLS estimator with bandwidths $H = n^{0.4}, n^{0.5}, ..., n^{0.7}$. Figure 9(a) displays the path and corresponding confidence intervals for $\widehat{\mu}_t$ for $H = n^{0.6}$ and reveals that μ_t is very likely to change over time.



(a) Confidence bands for μ_t

(b) Confidence bands for $\beta_{1t} = h_t E|\varepsilon_t|$

Figure 9: Robust 95% confidence bands for μ_t in model (39) and $\beta_{1t} = h_t E|\varepsilon_t|$ in model (40), $n = 7558$, $H = n^{0.6}$.



(a) Testing at individual lag: $\widehat{\widetilde{u}}_j$

(b) Testing at individual lag: $\widehat{u}^*_j$

Figure 10: Robust and standard tests for absence of correlation in subsample of residuals $\widehat{\widetilde{u}}_j$, $\widehat{u}^*_j$, $j \in [500, 1000]$, $H = n^{0.6}$, significance level 5%.

Assumption (39) implies that

$$|u_t| = |r_t - \mu_t| = h_t|\varepsilon_t| = h_t E|\varepsilon_t| + h_t(|\varepsilon_t| - E|\varepsilon_t|).$$

Then, $|\widehat{u}_t| = |r_t - \widehat{\mu}_t| \sim h_t E|\varepsilon_t| + h_t(|\varepsilon_t| - E|\varepsilon_t|)$ and thus $y_t = |\widehat{u}_t|$ follows a time-varying regression model

$$y_t = \beta_{1t} + \widetilde{u}_t, \quad \widetilde{u}_t = g_t \eta_t, \tag{40}$$

where $\beta_{1t} = h_t E|\varepsilon_t|$ is a time-varying intercept, $g_t = h_t$ is a scale factor and $\eta_t = |\varepsilon_t| - E|\varepsilon_t|$ is an i.i.d. noise. Hence β_{1t} can be estimated using the time-varying OLS estimator $\widehat{\beta}_{1t}$. Figure 9(b) displays the estimate $\widehat{\beta}_{1t}$ and confidence intervals for $\beta_{1t} = h_t E|\varepsilon_t|$ for bandwidth $H = n^{0.6}$ which reveals significant time variation in h_t.

Figure 10(a) reports testing results for zero correlation at lags $k = 1, ..., 20$ in the residual sequence $\widehat{\widetilde{u}}_t = y_t - \widehat{\beta}_{1t}$. We employ the standard test and robust test procedure developed in Giraitis et al. (2024). Since the sample size $n = 7558$ is large and β_{1t} is estimated non-parametrically with bandwidth $H = n^{0.6}$, we restrict testing to the subsample $j \in [500, 1000]$. Both tests report no evidence of significant correlation in this subsample, and this suggests a good fit of the model (39) to the returns r_t in this time period.

The same is not likely true if $r_t^* = r_t - \widehat{\mu}_t$ follows a GARCH(1,1) model, and this is confirmed by the following experiment. We match a GARCH(1,1) model to data $r_t^* = r_t - \widehat{\mu}_t$,

$$r_t^* = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1.563 \times 10^{-6} + 0.88913\sigma_{t-1}^2 + 0.096974 r_{t-1}^{*2},$$

generate a GARCH(1,1) sample $r_{g1}^*, ..., r_{gn}^*$, fit to $y_t^* = |r_{gt}^*|$ regression model (40) and compute residuals, $\widehat{u}_t^* = y_t^* - \widehat{\beta}_{1t}$. Figure 10(b) shows that both standard and robust test detect significant correlation in residuals $\widehat{u}_t^*$.

8 Conclusion

The robust OLS and time-varying OLS regression estimation and inference methods, presented in this paper, offer considerable flexibility in specifying regression models for economic and financial empirical applications. It allows for general heterogeneity in regression components and structural change of regression coefficient over time. The generalization of the structure of regressors and error terms further expands the area of scenarios to which robust OLS regression method can be applied. In particular, the present paper develops asymptotic theory of general regression modelling, when regressors are stochastic and may include intercept, and provides data based robust standard errors for building confidence intervals for regression parameters. Our Monte Carlo analysis shows the remarkable performance of the robust estimation approach under complex settings, and verifies the asymptotic normality property and consistency of parameter estimators.

References

Angrist, J. D. and J. S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B 37*(2), 149–163.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591–605.

Chronopoulos, I., L. Giraitis, and G. Kapetanios (2022). Choosing between persistent and stationary volatility. *Annals of Statistics 50*(6), 3466–3483.

Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics 8*(2), 262–302.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics 25*(1), 1–37.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory 37*(6), 1100–1134.

Enders, C. K. (2022). *Applied Missing Data Analysis*. Guilford Publications.

Giraitis, L., G. Kapetanios, and T. Yates (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics 179*(1), 46–65.

Giraitis, L., G. Kapetanios, and T. Yates (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis 39*(2), 129–149.

Giraitis, L., Y. Li, and P. C. B. Phillips (2024). Robust inference on correlation under general heterogeneity. *Journal of Econometrics 240*(1), 105691.

Giraitis, L., M. Taniguchi, and M. S. Taqqu (2018). Estimation pitfalls when the noise is not iid. *Japanese Journal of Statistics and Data Science 1*, 59–80.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Lai, T. L. and C. Z. Wei (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics 10*(1), 154–166.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review 73*(1), 31–43.

Leamer, E. E. (2010). Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives 24*(2), 31–46.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association 84*(405), 223–230.

Ploberger, W. and W. Krämer (1992). The CUSUM test with OLS residuals. *Econometrica 60*, 271–285.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B 27*(2), 204–229.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In P. Hackl (Ed.), *Economic Structural Change: Analysis and Foresting*, pp. 253–264. Springer Berlin.

Robinson, P. M. (1991). Time-varying nonlinear regression. In P. Hackl and A. H. Westland (Eds.), *Economic Structural Change: Analysis and Forecasting*, pp. 179–190. Springer Berlin.

Terasvirta, T. (1998). Modelling economic relationships with smooth transition regressions. In A. Ullah and D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics*, pp. 507–552. Marcel Dekker.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–838.

SUPPLEMENT TO "REGRESSION MODELLING UNDER GENERAL HETEROGENEITY"

Liudas Giraitis[1], George Kapetanios[2], Yufei Li[2]

[1]Queen Mary University of London

[2]King's College London

August 22, 2024

This Supplement provides proofs of the results given in the text of the main paper. It is organised as follows: Section 9, 10, 11 provide proofs of the main theorems. Section 12 contains auxiliary technical lemmas used in the proofs.

Formula numbering in this supplement includes the section number, e.g. (8.1), and references to lemmas are signified as "Lemma 10.#", e.g. Lemma 10.1. Theorem references to the main paper include section number and are signified, e.g. as Theorem 2.1, while equation references do not include section number, e.g. (1), (2).

In the proofs, C stands for a generic positive constant which may assume different values in different contexts.

9 Proofs of Theorems 2.1 and 2.2, Lemma 2.1 and Corollary 2.1

Proof of Theorem 2.1. Notice that in view of (1),

$$\begin{aligned}\widehat{\beta}-\beta &= \big(\sum_{j=1}^{n} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{n} z_j(z_j'\beta+u_j)\big)-\beta \\ &= S_{zz}^{-1}S_{zu}, \quad S_{zz}=\sum_{j=1}^{n} z_j z_j', \quad S_{zu}=\sum_{j=1}^{n} z_j u_j.\end{aligned}$$

Recall definition (5) of D and D_g. Then

$$\begin{aligned}D(\widehat{\beta}-\beta) &= (DS_{zz}^{-1}D)(D^{-1}S_{zu}) \\ &= (DD_g^{-1})(D_gS_{zz}^{-1}D_g)(D_g^{-1}D)(D^{-1}S_{zu})=O_p(1), \qquad (9.1)\end{aligned}$$

since $DD_g^{-1} = O_p(1)$ by (7) of Assumption 2.3, $D^{-1}S_{zu} = O_p(1)$ by (12.7) of Lemma 12.2. Moreover, by (12.6) and (12.3),

$$D_g S_{zz}^{-1} D_g = D_g E[S_{zz}\,|\mathcal{F}_n^*]^{-1} D_g + o_p(1) = O_p(1).$$

This completes the proof of the consistency claim (9) of the theorem. □

Recall that for $p \times p$ symmetric matrices A, B and a $p \times 1$ vector b it holds:

$$||AB||_{sp} \leq ||A||_{sp}||B||_{sp}, \quad ||AB|| \leq ||A||_{sp}||B||, \quad ||A||_{sp} \leq ||A||,$$

where $||A||_{sp}$ denotes the spectral norm and $||A||$ the Euclidean norm of the matrix A.

Proof of Theorem 2.2. *Proof of (13)*. By (9.1),

$$D(\widehat{\beta} - \beta) \;\; = \;\; \{DS_{zz}^{-1}D\}\{D^{-1}S_{zu}\}.$$

Moreover, by the same argument as in the proof of (9.1),

$$\begin{aligned} DS_{zz}^{-1}D &= (DD_g^{-1})(D_g S_{zz}^{-1} D_g)(D_g^{-1}D) \\ &= (DD_g^{-1})(D_g E[S_{zz}\,|\mathcal{F}_n^*]^{-1} D_g + o_p(1))(D_g^{-1}D) \\ &= DE[S_{zz}\,|\mathcal{F}_n^*]^{-1}D + o_p(1), \quad DS_{zz}^{-1}D = O_p(1). \end{aligned} \tag{9.2}$$

Hence,

$$\begin{aligned} a'\,D(\widehat{\beta} - \beta) &= a'\{DE[S_{zz}\,|\mathcal{F}_n^*]^{-1}D + o_p(1)\}\{D^{-1}S_{zu}\} \\ &= d_n S_{zu} + o_p(1), \quad d_n = a'(DE[S_{zz}|\mathcal{F}_n^*]^{-1}). \end{aligned} \tag{9.3}$$

By (12.11) of Lemma 12.2,

$$v_n^2 := (a'D\Omega_n Da) \geq b_n, \quad b_n^{-1} = O_p(1). \tag{9.4}$$

This together with (9.3) implies:

$$\frac{a'D(\widehat{\beta} - \beta)}{\sqrt{a'D\Omega_n Da}} = v_n^{-1} d_n S_{zu} + o_p(1).$$

Write

$$s_n \;\; = \;\; v_n^{-1} d_n S_{zu} = \sum_{t=1}^{n} \xi_t, \quad \xi_t = v_n^{-1} d_n z_t u_t.$$

To prove (13), it remains to show that

$$s_n \to_d \mathcal{N}(0,1). \tag{9.5}$$

Notice that $\{\xi_t\}$ is an m.d. sequence with respect to the σ-field
$\mathcal{F}_{n,t} = \sigma(\varepsilon_1, ..., \varepsilon_t;\ \mu_s, h_s, g_s, s = 1, ..., n)$:

$$E[\xi_t\,|\mathcal{F}_{n,t-1}] = E[v_n^{-1} d_n z_t h_t \varepsilon_t|\mathcal{F}_{n,t-1}] = v_n^{-1} d_n z_t h_t E[\varepsilon_t|\mathcal{F}_{n,t-1}] = 0. \tag{9.6}$$

Therefore, by Corollary 3.1 of Hall and Heyde (1980), to prove (9.5), it suffices to show that

$$\begin{aligned} &(a) \quad \sum_{t=1}^n E[\xi_t^2\,|\mathcal{F}_{n,t-1}] \to_p 1, \\ &(b) \quad \sum_{t=1}^n E[\xi_t^2 I(\xi_t^2 \geq \epsilon)\,|\mathcal{F}_{n,t-1}] = o_p(1) \quad \text{for any } \epsilon > 0. \end{aligned} \tag{9.7}$$

To verify (a), notice that

$$\begin{aligned} \xi_t^2 &= (v_n^{-1} d_n z_t u_t)^2 = v_n^{-2} d_n z_t z_t' d_n' u_t^2, \\ E[\xi_t^2\,|\mathcal{F}_{n,t-1}] &= v_n^{-2} d_n z_t z_t' d_n' E[u_t^2\,|\mathcal{F}_{n,t-1}] = v_n^{-2} d_n z_t z_t' d_n' h_t^2 E[\varepsilon_t^2\,|\mathcal{F}_{t-1}]. \end{aligned}$$

Then, setting $S_{zzuu}^{(c)} = \sum_{t=1}^n z_t z_t' h_t^2 E[\varepsilon_t^2\,|\mathcal{F}_{t-1}]$, we can write,

$$\begin{aligned} \sum_{t=1}^n E[\xi_t^2\,|\mathcal{F}_{n,t-1}] &= v_n^{-2} d_n\, S_{zzuu}^{(c)}\, d_n' \\ &= v_n^{-2}\, a'\{DE[S_{zz}|\mathcal{F}_n^*]^{-1}D\}\{D^{-1}S_{zzuu}^{(c)}D^{-1}\}\{DE[S_{zz}|\mathcal{F}_n^*]^{-1}D\}a. \end{aligned} \tag{9.8}$$

Recall that by (9.2), $DE[S_{zz}|\mathcal{F}_n^*]^{-1}D = O_p(1)$. We show in (12.14) of Lemma 12.2 that

$$D^{-1}S_{zzuu}^{(c)}D^{-1} = D^{-1}E[S_{zzuu}|\mathcal{F}_n^*]D^{-1} + o_p(1).$$

Together with (9.4), this implies

$$\begin{aligned} \sum_{t=1}^n E[\xi_t^2\,|\mathcal{F}_{n,t-1}] &= v_n^{-2}\, a'\{D\, E[S_{zz}|\mathcal{F}_n^*]^{-1}\, E[S_{zzuu}|\mathcal{F}_n^*]\, E[S_{zz}|\mathcal{F}_n^*]^{-1} D\}a + o_p(1) \\ &= v_n^{-2}(a' D\Omega_n D\, a) + o_p(1) = 1 + o_p(1) \end{aligned}$$

which proves (a).

Next we prove (b). We have

$$\begin{aligned} \xi_t &= v_n^{-1} d_n z_t u_t = v_n^{-1}(d_n D)(D^{-1} z_t u_t), \\ \xi_t^2 &\leq v_n^{-2}||d_n D||^2||D^{-1}z_t u_t||^2. \end{aligned}$$

By definition of d_n, $||d_n D||^2 = ||a' D\, E[S_{zz}|\mathcal{F}_n^*]^{-1} D||^2$. On the other hand, by (12.18) of Corollary 12.1, for any a,

$$a' D^{-1} E[S_{zzuu}|\mathcal{F}_n^*] D^{-1} a \quad \geq \quad b_n ||a||^2, \quad b_n^{-1} = O_p(1),$$

where b_n is $\mathcal{F}_n^*$ measurable, and, thus, also $\mathcal{F}_{n,t-1}$ measurable. Then,

$$\begin{aligned} v_n^2 &= a' D \Omega_n D\, a = \{a' D (E[S_{zz}|\mathcal{F}_n^*])^{-1} D\}\{D^{-1} E[S_{zzuu}|\mathcal{F}_n^*] D^{-1}\}\{D (E[S_{zz}|\mathcal{F}_n^*])^{-1} D\, a\} \\ &\geq ||a' D (E[S_{zz}|\mathcal{F}_n^*])^{-1} D||^2 b_n = ||d_n D||^2 b_n, \\ \xi_t^2 &\leq b_n^{-1} ||D^{-1} z_t u_t||^2. \end{aligned}$$

Hence,

$$\sum_{t=1}^n E[\xi_t^2 I(\xi_t^2 \geq \epsilon)\, |\mathcal{F}_{n,t-1}] \leq \sum_{t=1}^n E\big[b_n^{-1} ||D^{-1} z_t u_t||^2 I\big(b_n^{-1} ||D^{-1} z_t u_t||^2 \geq \epsilon\big)\, |\mathcal{F}_{n,t-1}\big] = o_p(1),$$

by (12.54) of Lemma 12.3. This completes the proof (b) and the claim (13) of the theorem.

The claim (14) follows from (13) by setting $a = (a_1, ..., a_p)' = (0, ..., 0, 1, 0....)'$ where $a_k = 1$ and $a_j = 0$ for $j \neq k$. Then $a' D = v_k$ and $a' D \Omega_n D a = v_k^2 \omega_{kk}$, where ω_{kk} is the (k, k)-th diagonal element of Ω_n. Then,

$$\frac{a' D(\widehat{\beta} - \beta)}{\sqrt{a' D \Omega_n D a}} = \frac{(\widehat{\beta} - \beta)}{\sqrt{\omega_{kk}}} \to_d \mathcal{N}(0, 1).$$

This completes the proof of the theorem. $\square$

Proof of Lemma 2.1. *Proof of (6)*. It suffices to show that

$$i_n = v_{gk}^{-2} \max_{1 \leq t \leq n} (g_{kt}^2 + \mu_{kt}^2) = o_p(1). \tag{9.9}$$

Notice also that $z_{kt}^2 = \mu_{kt}^2 + 2\mu_{kt} g_{kt} \eta_{kt} + g_{kt}^2 \eta_{kt}^2$,

$$E[z_{kt}^2\, |\mathcal{F}_n^*] = \mu_{kt}^2 + 2\mu_{kt} g_{kt} E[\eta_{kt}\, |\mathcal{F}_n^*] + g_{kt}^2 E[\eta_{kt}^2\, |\mathcal{F}_n^*] = \mu_{kt}^2 + g_{kt}^2. \tag{9.10}$$

In addition, by assumption of lemma, $v_{gk}^{-2} = (\sum_{t=1}^n g_{kt}^2)^{-1} = O_p(n^{-1})$. Thus,

$$i_n = O_p(1) i_{n,1}, \quad i_{n,1} = n^{-1} \max_{1 \leq t \leq n} E[z_{kt}^2\, |\mathcal{F}_n^*].$$

We will show that $E i_{n,1} = o(1)$ which implies (9.9). Observe that for any $L \geq 1$,

$$z_{kt}^2 \leq L + z_{kt}^2 I(z_{kt}^2 \geq L) \leq L + L^{-1} z_{kt}^4.$$

By assumption (10), $E[z_{kt}^4] \le c < \infty$ where c does not depend on t, n. Hence,

$$\begin{aligned} i_{n,1} &\le n^{-1}L + n^{-1}L^{-1} \max_{t=1,\dots,n} E[z_{kt}^4 \,|\mathcal{F}_n^*], \\ Ei_{n,1} &\le n^{-1}L + n^{-1}L^{-1} \textstyle\sum_{t=1}^n E[z_{kt}^4] \le n^{-1}L + L^{-1}c \to 0, \quad n, L \to \infty \end{aligned}$$

which implies $i_n = o_p(1)$ and proves (6).

Proof of (11). It suffices to verify that

$$i_n = v_k^{-2} \max_{1 \le t \le n} (g_{kt}^2 + \mu_{kt}^2) h_t^2 = o_p(1). \tag{9.11}$$

By assumption of lemma, $v_k^{-2} = O_p(n^{-1})$. This together with (9.10) implies that

$$i_n = O_p(1) i_{n,2}, \quad i_{n,2} = n^{-1} \max_{1 \le t \le n} E[z_{kt}^2 h_t^2 \,|\mathcal{F}_n^*].$$

We will show that $Ei_{n,2} = o(1)$ which implies $i_{n,2} = o_p(1)$ and proves (9.11).

Similarly as above, for any $L \ge 1$, setting $L_0 = \log L$, for $\delta > 0$ we obtain

$$\begin{aligned} z_{kt}^2 h_t^2 &\le L + z_{kt}^2 h_t^2 I(z_{kt}^2 h_t^2 \ge L) \\ &\le L + L_0^{-1} z_{kt}^4 I(h_t^2 \le L_0^{-1} z_{kt}^2) + h_t^4 L_0 I(h_t^2 > L_0^{-1} z_{kt}^2) I(h_t^4 L_0 \ge L) \\ &\le L + L_0^{-1} z_{kt}^4 + h_t^4 L_0 \Big(\frac{h_t^4}{L L_0^{-1}}\Big)^\delta \\ &\le L + L_0^{-1} z_{kt}^4 + h_t^{4+4\delta} A_L, \quad A_L = L^{-\delta} L_0^{1+\delta}. \end{aligned}$$

By assumption (10), $E[z_{kt}^4] \le c$ and there exists $\delta > 0$ such that $E[|u_t|^{4+4\delta}] \le c$, where $c < \infty$ does not depend on t, n. Hence, $E[h_t^{4+4\delta}] = E[(E[u_t^2 \,|\mathcal{F}_n^*])^{2+2\delta}] \le E[|u_t|^{4+4\delta}] \le c$. Notice that $A_L \to 0$ as $L \to \infty$. Therefore, as $n, L \to \infty$,

$$\begin{aligned} Ei_{n,2} &\le n^{-1}L + L_0^{-1} n^{-1} \textstyle\sum_{t=1}^n E[z_{kt}^4] + A_L n^{-1} \sum_{t=1}^n E[h_t^{4+4\delta}] \\ &\le n^{-1}L + L_0^{-1} c + A_L c \to 0, \end{aligned}$$

which implies $i_n = o_p(1)$ and proves (11).

Proof of (7). Recall that $v_k^2 \asymp_p n$, $v_{gk}^2 \asymp_p n$ by assumption of lemma. By (9.10),

$$\begin{aligned} \mu_{kt}^2 &\le E[z_{kt}^2 \,|\mathcal{F}_n^*], \quad E[\mu_{kt}^2] \le E[z_{kt}^2] \le c, \\ E[\mu_{kt}^4] &\le E[(E[z_{kt}^2 \,|\mathcal{F}_n^*])^2] \le E[(E[z_{kt}^4 \,|\mathcal{F}_n^*])] \le E[z_{kt}^4] \le c, \\ E[g_{kt}^4] &\le E[(E[z_{kt}^2 \,|\mathcal{F}_n^*])^2] \le \ c, \\ E[h_t^4] &= E[(E[u_t^2 \,|\mathcal{F}_n^*])^2] \le E[(E[u_t^4 \,|\mathcal{F}_n^*])] \le E[u_t^4] \le c, \\ E[\mu_{kt}^2 h_t^2] &\le (E[\mu_{kt}^4] E[h_t^4])^{1/2} \le c, \\ E[g_{kt}^2 h_t^2] &\le (E[g_{kt}^4] E[h_t^4])^{1/2} \le c, \end{aligned}$$

where $c < \infty$ does not depend on t, n. Hence,

$$n^{-1}E[\sum_{t=1}^n \mu_{kt}^2] \le c, \quad \sum_{t=1}^n \mu_{kt}^2 = O_p(n),$$
$$n^{-1}E[\sum_{t=1}^n \mu_{kt}^2 h_t^2] \le c, \quad \sum_{t=1}^n \mu_{kt}^2 h_t^2 = O_p(n),$$
$$n^{-1}E[\sum_{t=1}^n g_{kt}^2 h_t^2] \le c, \quad \sum_{t=1}^n g_{kt}^2 h_t^2 = O_p(n),$$

which proves (7). This completes the proof of the lemma. □

Proof of Corollary 2.1. We will show that

$$\frac{\widehat{\omega}_{kk}}{\omega_{kk}} = 1 + o_p(1) \tag{9.12}$$

which together with (14) implies (16):

$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} = \Big(\sqrt{\frac{\omega_{kk}}{\widehat{\omega}_{kk}}}\Big)\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\omega_{kk}}} = (1 + o_p(1))\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\omega_{kk}}} \to_d \mathcal{N}(0, 1).$$

To prove (9.12), we will verify that

$$D\widehat{\Omega}_n D = D\Omega_n D + o_p(1) \tag{9.13}$$

which implies the following property for diagonal elements:

$$v_k^2 \widehat{\omega}_{kk} = v_k^2 \omega_{kk} + o_p(1).$$

In (12.11) of Lemma 12.2 it is shown that

$$a' D\Omega_n Da \ge b_n, \quad a' D\Omega_n Da \le b_{n2} \tag{9.14}$$

for any $a = (a_1, ..., a_p)'$, $||a|| = 1$ where b_n, $b_{n2} > 0$ do not depend on a, n and $b_n^{-1} = O_p(1)$, $b_{n2} = O_p(1)$. Set $a = (0, ..., 1, ...0)'$, where $a_j = 0$ for $j \ne k$ and $a_k = 1$. Then $a' D\Omega_n Da = v_k^2 \omega_{kk}$, and by (9.14), $v_k^2 \omega_{kk} \ge b_n > 0$. This proves (9.12):

$$\frac{\widehat{\omega}_{kk}}{\omega_{kk}} = \frac{v_k^2 \widehat{\omega}_{kk}}{v_k^2 \omega_{kk}} = \frac{v_k^2 \omega_{kk} + o_p(1)}{v_k^2 \omega_{kk}} = 1 + o_p(1).$$

In addition, the bounds (9.14) imply that $\sqrt{\omega_{kk}} \asymp_p v_k^{-1}$:

$$v_k^{-1} \le b_n^{-1/2}\sqrt{\omega_{kk}} = O_p(\sqrt{\omega_{kk}}), \quad v_k\sqrt{\omega_{kk}} = O_p(1), \quad \sqrt{\omega_{kk}} = O_p(v_k^{-1}).$$

Proof of (9.13). Set $V_n = DD_g^{-1}$. By (7) of Assumption 2.3, $V_n = O_p(1)$. We have

$$D\widehat{\Omega}_n D = V_n\{D_g S_{zz}^{-1} D_g\} V_n \{D^{-1} S_{zz\widehat{u}\widehat{u}} D^{-1}\} V_n \{D_g S_{zz}^{-1} D_g\} V_n,$$

$$\begin{aligned} D\Omega_n D &= V_n W_{zz}^{-1} V_n W_{zzuu} V_n W_{zz}^{-1} V_n, \\ & \quad W_{zz}^{-1} = D_g E[S_{zz}\,|\mathcal{F}_n^*]^{-1} D_g, \;\; W_{zzuu} = D^{-1} E[S_{zzuu}\,|\mathcal{F}_n^*] D^{-1}. \end{aligned}$$

By (12.6), (12.3), (12.12) and (12.10) of Lemma 12.2,

$$\begin{aligned} D_g S_{zz}^{-1} D_g &= W_{zz}^{-1} + o_p(1), \;\; W_{zz}^{-1} = O_p(1), \\ D^{-1} S_{zzuu} D^{-1} &= W_{zzuu} + o_p(1), \;\; W_{zzuu} = O_p(1). \end{aligned}$$

We will show that

$$D^{-1} S_{zz\widehat{u}\widehat{u}} D^{-1} = D^{-1} S_{zzuu} D^{-1} + o_p(1). \tag{9.15}$$

This implies (9.13):

$$\begin{aligned} D\widehat{\Omega}_n D &= V_n\{W_{zz}^{-1} + o_p(1)\} V_n \{W_{zzuu} + o_p(1)\} V_n \{W_{zz}^{-1} + o_p(1)\} V_n \\ &= V_n W_{zz}^{-1} V_n W_{zzuu} V_n W_{zz}^{-1} V_n + o_p(1) = D\Omega_n D + o_p(1). \end{aligned}$$

Proof of (9.15). By definition,

$$\begin{aligned} & ||D^{-1}(S_{zz\widehat{u}\widehat{u}} - S_{zzuu}) D^{-1}|| = ||\sum_{t=1}^n D^{-1} z_t z_t' D^{-1} (\widehat{u}_t^2 - u_t^2)|| \\ & \quad \le \sum_{t=1}^n ||D^{-1} z_t||^2 \, |\widehat{u}_t^2 - u_t^2| \le i_n \times (\sum_{t=1}^n ||D^{-1} z_t||^2), \quad i_n = \max_{t=1,\dots,n} |\widehat{u}_t^2 - u_t^2|. \end{aligned}$$

Notice that

$$\sum_{t=1}^n ||D^{-1} z_t||^2 \le ||D^{-1} D_g||^2 \sum_{t=1}^n ||D_g^{-1} z_t||^2 = O_p(1),$$

since $||D_g D^{-1}|| = O_p(1)$ by assumption (7) and $\sum_{t=1}^n ||D_g^{-1} z_t||^2 = O_p(1)$ by (12.8) of Lemma 12.2. Hence, to verify (9.15), it suffices to show that

$$i_n = o_p(1). \tag{9.16}$$

Recall the equality $\widehat{u}_t^2 - u_t^2 = (\widehat{u}_t - u_t)^2 + 2(\widehat{u}_t - u_t) u_t$. Denote $q_n = ||D(\beta - \widehat{\beta})||$. Then,

$$\begin{aligned} \widehat{u}_t - u_t &= (\beta - \widehat{\beta})' z_t = \{(\beta - \widehat{\beta})' D\}\{D^{-1} z_t\}, \\ |\widehat{u}_t - u_t| &\le ||D^{-1} z_t||\, q_n, \\ |\widehat{u}_t^2 - u_t^2| &\le (\widehat{u}_t - u_t)^2 + 2|(\widehat{u}_t - u_t) u_t| \le ||D^{-1} z_t||^2\, q_n^2 + 2||D^{-1} z_t||\, |u_t|\, q_n. \end{aligned}$$

Hence,

$$i_n \le (\max_{t=1,\dots,n} ||D^{-1} z_t||^2)\, q_n^2 + 2(\max_{t=1,\dots,n} ||D^{-1} z_t u_t||)\, q_n = o_p(1),$$

where $q_n = O_p(1)$ by Theorem 2.1, and

$$\max_{t=1,...,n} ||D^{-1}z_t||^2 = o_p(1), \quad \max_{t=1,...,n} ||D^{-1}z_t u_t|| = o_p(1)$$

by (12.53) of Lemma 12.3. This implies (9.16) and completes the proof of the corollary. □

10 Proofs of Theorem 3.1 and Corollary 3.1

Proof of Theorem 3.1. By (19), $\widetilde{y}_j = \widetilde{z}_j'\beta_t + \widetilde{u}_j + r_j$. Thus,

$$\widehat{\beta}_t = \Big(\sum_{j=1}^n \tilde{z}_j\tilde{z}_j'\Big)^{-1}\Big(\sum_{j=1}^n \tilde{z}_j\tilde{y}_j\Big) = \widetilde{\beta}_t + B_t, \tag{10.1}$$

$$\widetilde{\beta}_t = \Big(\sum_{j=1}^n \tilde{z}_j\tilde{z}_j'\Big)^{-1}\sum_{j=1}^n \tilde{z}_j(\tilde{z}_j'\beta_t + \widetilde{u}_j), \quad B_t = \Big(\sum_{j=1}^n \tilde{z}_j\tilde{z}_j'\Big)^{-1}\sum_{j=1}^n \tilde{z}_j r_j.$$

The term $\widetilde{\beta}_t$ is the OLS estimator of the *fixed* parameter β_t in the regression model

$$y_j^* = \beta_t'\tilde{z}_j + \tilde{u}_j, \quad j = 1, ..., n.$$

As is shown below, the consistency rate and the asymptotic normality for $\widetilde{\beta}_t - \beta_t$ can be derived using results of Section 2. The term $B_t = (B_{1t}, ..., B_{pt})'$ arises due to time variation in the parameter β_j. It can be treated as a negligible term or a "bias" term. First we will show that for $k = 1, ..., p$,

$$\widetilde{\beta}_{kt} - \beta_{kt} = O_p(H^{-1/2}), \quad \frac{\widetilde{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} \to_d \mathcal{N}(0,1), \quad \sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}, \tag{10.2}$$

$$B_{kt} = O_p\big((H/n)^\gamma\big). \tag{10.3}$$

Proof of (10.2). Recall that $\widetilde{z}_j = (\widetilde{z}_{1j}, ..., \widetilde{z}_{pj})'$, and

$$\widetilde{z}_{kj} = \widetilde{\mu}_{kj} + \widetilde{g}_{kj}\eta_{kj}, \quad \widetilde{u}_j = \widetilde{h}_j\varepsilon_j, \tag{10.4}$$

$$\widetilde{\mu}_{kj} = b_{n,tj}^{1/2}\mu_{kj}, \quad \widetilde{g}_{kj} = b_{n,tj}^{1/2}g_{kj}, \quad \widetilde{h}_j = b_{n,tj}^{1/2}h_j.$$

By Lemma 10.1, under assumptions of theorem, $\widetilde{\mu}_{kj}$ and the scale factors $\{\widetilde{g}_{kj}, \widetilde{h}_j\}$ satisfy Assumptions 2.3 and 2.4(ii). Thus, by Theorem 2.1,

$$\widetilde{\beta}_{kt} - \beta_{kt} = O_p(v_k^{-1}) = O_p(H^{-1/2}),$$

where $v_k^2 \equiv v_{kt}^2 = \sum_{j=1}^n \widetilde{g}_{kj}^2\widetilde{h}_j^2 = \sum_{j=1}^n b_{n,tj}g_{kj}^2h_j^2 \asymp_p H$ by Assumption 3.2. This proves the first claim in (10.2), while the second claim holds by (14) of Theorem 2.2. The third claim holds since by (16) of Corollary 2.1 and Assumption 3.2,

$$\sqrt{\omega_{kk,t}} \asymp_p (\sum_{j=1}^{n} \widetilde{g}_{kj}^2 \widetilde{h}_j^2)^{-1/2} \asymp_p H^{-1/2}. \tag{10.5}$$

Proof of (10.3). Write

$$B_t = S_{\tilde{z}\tilde{z},t}^{-1} S_{\tilde{z}\tilde{z}\beta,t}, \quad \text{where } S_{\tilde{z}\tilde{z}\beta,t} = \textstyle\sum_{j=1}^{n} \tilde{z}_j \tilde{z}_j' (\beta_j - \beta_t).$$

We will show that

$$||S_{\tilde{z}\tilde{z},t}^{-1}|| = O_p(H^{-1}), \quad ||S_{\tilde{z}\tilde{z}\beta,t}|| = O_p\big(H(H/n)^{\gamma}\big), \tag{10.6}$$

which implies $||B_t|| \le ||S_{\tilde{z}\tilde{z},t}^{-1}||\,||S_{\tilde{z}\tilde{z}\beta,t}|| = O_p\big((H/n)^{\gamma}\big)$. Then, $|B_{kt}| \le ||B_t|| = O_p\big((H/n)^{\gamma}\big)$ which implies (10.3).

To verify (10.6), recall notation of the $p \times p$ diagonal matrix

$$D_{\widetilde{g}} = \text{diag}(v_{\widetilde{g}1}, ..., v_{\widetilde{g}p}), \quad v_{\widetilde{g}} = \textstyle\sum_{j=1}^{n} \widetilde{g}_{kj}^2, \quad k = 1, ..., p.$$

Notice that

$$||S_{\tilde{z}\tilde{z},t}^{-1}|| = ||D_{\widetilde{g}}^{-1}(D_{\widetilde{g}} S_{\tilde{z}\tilde{z},t}^{-1} D_{\widetilde{g}})|D_{\widetilde{g}}^{-1}|| \le ||D_{\widetilde{g}}^{-1}||^2 ||D_{\widetilde{g}} S_{\tilde{z}\tilde{z},t}^{-1} D_{\widetilde{g}}|| = O_p(H^{-1})$$

because $||D_{\widetilde{g}}^{-1}||^2 = \sum_{k=1}^{p} v_{\widetilde{g}k}^2 = O_p(H^{-1})$ by Assumption 3.2. On the other hand, $D_{\widetilde{g}} S_{\tilde{z}\tilde{z},t}^{-1} D_{\widetilde{g}} = O_p(1)$ by (12.6) and (12.3) of Lemma 12.2. This proves the first claim in (10.6).

Next, bound

$$E||S_{\tilde{z}\tilde{z}\beta,t}|| \le E[\sum_{j=1}^{n} ||\tilde{z}_j||^2 ||\beta_j - \beta_t||] \le \sum_{j=1}^{n} (E||\tilde{z}_j||^4)^{1/2} (E||\beta_j - \beta_t||^2)^{1/2}.$$

We have $||\widetilde{z}_j||^4 = b_{n,tj}^2 ||z_j||^4$. Recall that $E||z_j||^4 \le c$ by Assumption 3.2, $E||\beta_j - \beta_t||^2 \le c(|t-j|/n)^{2\gamma}$ by Assumption 3.1, and it is trivial to show that under (22), $\sum_{j=1}^{n} b_{n,tj}(|t - j|/H)^{\gamma} = O(H)$. This implies

$$E||S_{\tilde{z}\tilde{z}\beta,t}|| \le CH(H/n)^{\gamma} \Big(H^{-1} \sum_{j=1}^{n} b_{n,tj}(|t-j|/H)^{\gamma}\Big) \le CH(H/n)^{\gamma}$$

which proves the second claim in (10.6).

We now use (10.2) and (10.3) to prove the results of the theorem. By (10.1), $\widehat{\beta}_t = \widetilde{\beta}_t + B_t$. Thus, (10.2) and (10.3) imply the consistency result (25):

$$\widehat{\beta}_t - \beta_t = (\widetilde{\beta}_t - \beta_t) + B_t = O_p\big(H^{-1/2} + (H/n)^{\gamma}\big).$$

Further, suppose that $H = o(n^{2\gamma/(2\gamma+1)})$. We have

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} = \frac{\widetilde{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} + \omega_{kk,t}^{-1/2} B_t.$$

By (10.5), $\omega_{kk,t}^{-1/2} = O_p\big(H^{1/2}\big)$. Together with (10.3) this implies that $\omega_{kk,t}^{-1/2} B_t = O_p\big(H^{1/2}(H/n)^{\gamma}\big) = o_p(1)$. Then,

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} = \frac{\widetilde{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} + o_p(1) \to_d \mathcal{N}(0,1)$$

by (10.2) which proves the asymptotic normality property (26) of the theorem. Noting that we already proved (10.5), this completes the proof of the theorem. □

Proof of Corollary 3.1. Write the time-varying regression model as a regression model $\widetilde{y}_j = \widetilde{z}_j' \beta_t + \widetilde{u}_j + r_j$ with a fixed parameter (19) where $r_j = (\beta_j - \beta_t)' \widetilde{z}_j$. We showed in the proof of Theorem 3.1 that regressors $\widetilde{z}_j$ and the noise $\widetilde{u}_j$ satisfy assumptions of Theorem 2.2 and that the term r_j is asymptotically negligible. That allowed us to establish the asymptotic normality property (26) of Theorem 3.1 for $\widehat{\beta}_{kt}$ using results of Section 2.

Clearly, to prove Corollary 3.1, it suffices to verify the second claim in (28),

$$\frac{\widehat{\omega}_{kk,t}}{\omega_{kk,t}} = 1 + o_p(1).$$

Proof of the corresponding result in the case of fixed parameter in Corollary 2.1 shows that we need to verify the validity of (9.15) for our regression model (19), i.e. to show that

$$j_n = D^{-1} S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} D^{-1} = D^{-1} S_{\widetilde{z}\widetilde{z}\widetilde{u}\widetilde{u}} D^{-1} + o_p(1), \tag{10.7}$$

where $\widehat{u}_j = \widetilde{y}_j - \widehat{\beta}_t' \widetilde{z}_j$ and $D = \mathrm{diag}(v_1, ..., v_k)'$, $v_k^2 = \sum_{j=1}^n \widetilde{g}_{kj}^2 \widetilde{h}_j^2$.

Set $\widehat{u}_j^* = (\beta_t - \widehat{\beta}_t)' \widetilde{z}_j + \widetilde{u}_j$. Write

$$j_n = D^{-1} S_{\widetilde{z}\widetilde{z}\widehat{u}^*\widehat{u}^*} D^{-1} + D^{-1}(S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} - S_{\widetilde{z}\widetilde{z}\widehat{u}^*\widehat{u}^*}) D^{-1} = j_{n1} + j_{n2}.$$

By (9.15), $j_{n1} = D^{-1} S_{\widetilde{z}\widetilde{z}\widetilde{u}\widetilde{u}} D^{-1} + o_p(1)$. Hence, to prove (10.7), we need to show that

$$j_{n2} = o_p(1). \tag{10.8}$$

By Assumption 3.2, $||D^{-1}|| = O_p(H^{-1/2})$. Hence,

$$||j_{n2}|| \le ||D^{-1}||^2 ||S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} - S_{\widetilde{z}\widetilde{z}\widehat{u}^*\widehat{u}^*}|| = O_p(1)||j_{n3}||, \quad j_{n3} = H^{-1}(S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} - S_{\widetilde{z}\widetilde{z}\widehat{u}^*\widehat{u}^*}).$$

We will show that $j_{n3} = o_p(1)$ which implies (10.8). Notice that

$$\begin{aligned}
\widehat{u}_j &= \widetilde{y}_j - \widehat{\beta}_t'\widetilde{z}_j = (\beta_t - \widehat{\beta}_t)'\widetilde{z}_j + \widetilde{u}_j + r_j = \widehat{u}_j^* + r_j, \\
\widehat{u}_j^2 - \widehat{u}_j^{*\,2} &= (\widehat{u}_j - \widehat{u}_j^*)^2 + 2(\widehat{u}_j - \widehat{u}_j^*)\widehat{u}_j^* \\
&= r_j^2 + 2r_j\widehat{u}_j^* = r_j^2 + 2r_j(\beta_t - \widehat{\beta}_t)'\widetilde{z}_j + 2r_j\widetilde{u}_j. \qquad (10.9)
\end{aligned}$$

Using the inequality $2|ab| \le a^2 + b^2$, we can bound in (10.9),

$$2|r_j(\beta_t - \widehat{\beta}_t)'\widetilde{z}_j| \le r_j^2 + \big((\beta_t - \widehat{\beta}_t)'\widetilde{z}_j\big)^2 \le r_j^2 + ||\beta_t - \widehat{\beta}_t||^2||\widetilde{z}_j||^2.$$

Next we evaluate $|r_j\widetilde{u}_j|$ in (10.9). Let $L > 1$ be large number. Then,

$$\begin{aligned}
&|r_j| \le L^{-1}||\widetilde{z}_j||\, I\big(|r_j| \le L^{-1}||\widetilde{z}_j||\big) + |r_j|I\big(|r_j| > L^{-1}||\widetilde{z}_j||\big) \le L^{-1}||\widetilde{z}_j|| + Lr_j^2||\widetilde{z}_j||^{-1}, \\
&|r_j\widetilde{u}_j| \le L^{-1}||\widetilde{z}_j||\,|\widetilde{u}_j| + Lr_j^2||\widetilde{z}_j||^{-1}\,|\widetilde{u}_j|.
\end{aligned}$$

Hence,

$$|\widehat{u}_j^2 - \widehat{u}_j^{*\,2}| \le 2r_j^2 + ||\beta_t - \widehat{\beta}_t||^2||\widetilde{z}_j||^2 + 2L^{-1}||\widetilde{z}_j||\,|\widetilde{u}_j| + 2L||\widetilde{z}_j||^{-1}\,|\widetilde{u}_j|r_j^2.$$

Since $r_j^2 \le ||\beta_j - \beta_t||^2||\widetilde{z}_j||^2$, this yields

$$\begin{aligned}
||\widetilde{z}_j||^2|\widehat{u}_j^2 - \widehat{u}_j^{*\,2}| &\le 2||\beta_j - \beta_t||^2||\widetilde{z}_j||^4 + ||\beta_t - \widehat{\beta}_t||^2||\widetilde{z}_j||^4 + 2L^{-1}||\widetilde{z}_j||^3\,|\widetilde{u}_j| \\
&\quad + 2L||\widetilde{z}_j||^3\,|\widetilde{u}_j|\,||\beta_j - \beta_t||^2.
\end{aligned}$$

Recall that $\widetilde{z}_j = b_{n,tj}^{1/2}z_j$ and $\widetilde{u}_j = b_{n,tj}^{1/2}u_j$. Denote $\theta_j = 2||z_j||^4 + 2||z_j||^3|u_j|$. Then,

$$||\widetilde{z}_j||^2|\widehat{u}_j^2 - \widehat{u}_j^{*\,2}| \le Lb_{n,tj}^2||\beta_j - \beta_t||^2\theta_j + \big(||\beta_t - \widehat{\beta}_t||^2 + L^{-1}\big)b_{n,tj}^2\theta_j.$$

Hence,

$$\begin{aligned}
|j_{n3}| &= H^{-1}\big|\textstyle\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'(\widehat{u}_j^2 - \widehat{u}_j^{*\,2})\big| \le H^{-1}\sum_{j=1}^n ||\widetilde{z}_j||^2|\widehat{u}_j^2 - \widehat{u}_j^{*\,2}| \\
&\le L\{H^{-1}\textstyle\sum_{j=1}^n b_{n,tj}^2||\beta_j - \beta_t||^2\theta_j\} + (||\beta_t - \widehat{\beta}_t||^2 + L^{-1})\{H^{-1}\sum_{j=1}^n b_{n,tj}^2\theta_j\} \\
&\le L\{\textstyle\sum_{j=1}^n b_{n,tj}||\beta_j - \beta_t||^2\}\{H^{-1}\sum_{j=1}^n b_{n,tj}\theta_j\} \\
&\qquad + (||\beta_t - \widehat{\beta}_t||^2 + L^{-1})\{H^{-1}\textstyle\sum_{j=1}^n b_{n,tj}^2\theta_j\} \\
&= Lq_{n1}q_{n2} + (||\beta_t - \widehat{\beta}_t||^2 + L^{-1})q_{n3}. \qquad (10.10)
\end{aligned}$$

By (25) of Theorem 3.1, $||\beta_t - \widehat{\beta}_t||^2 = o_p(1)$, and L^{-1} can be made arbitrarily small by selecting large L. We will show that

$$Eq_{n1} = o(1), \quad Eq_{n2} = O(1), \quad Eq_{n3} = O(1). \qquad (10.11)$$

Combining this with (10.10), we obtain

$$|j_{n3}| = Lo_p(1) + \big(o_p(1) + L^{-1}\big)O_p(1),$$

so that the right hand side can be made arbitrarily small by selecting a large enough L and letting $n \to \infty$. This proves (10.8).

To bound Eq_{n1} observe that by Assumption 3.1, $E||\beta_t - \beta_j||^2 \le C(|t-j|/n)^{2\gamma}$, where $0 < \gamma \le 1$ and under assumption (22), $H^{-1}\sum_{j=1}^{n} b_{n,tj}(|t-j|/H)^{2\gamma} = O(1)$. Hence,

$$\begin{aligned} Eq_{n1} \le \sum_{j=1}^{n} b_{n,tj} E||\beta_j - \beta_t||^2 &\le C\big(H(\frac{H}{n})^{2\gamma}\big)\{H^{-1}\sum_{j=1}^{n} b_{n,tj}(\frac{|t-j|}{H})^{2\gamma}\} \\ &\le C(H/n)^{2\gamma} = o(1) \end{aligned}$$

when $H = o(n^{2\gamma/(2\gamma+1)})$. This proves (10.11) for Eq_{n1}.

To bound Eq_{n2} and Eq_{n3}, recall that by Assumption 3.2, $Ez_{kj}^4 \le C$ and $Eu_j^4 \le C$ which implies that $E\theta_j \le C$. Moreover, under (22) it holds $H^{-1}\sum_{j=1}^{n} b_{n,tj} = O(1)$ and $b_{n,tj}^2 \le C b_{n,tj}$. Hence,

$$\begin{aligned} Eq_{n2} &\le H^{-1}\textstyle\sum_{j=1}^{n} b_{n,tj} E\theta_j \le CH^{-1}\sum_{j=1}^{n} b_{n,tj} = O(1), \\ Eq_{n3} &\le H^{-1}\textstyle\sum_{j=1}^{n} b_{n,tj}^2 E\theta_j \le CH^{-1}\sum_{j=1}^{n} b_{n,tj} = O(1). \end{aligned}$$

This completes the proof of (10.11) and the corollary. □

Lemma 10.1. *Suppose that Assumption 3.2 holds and Assumptions 2.1, 2.2 are satisfied. Then $\{\widetilde{\mu}_{kj}, \widetilde{g}_{kj}, \widetilde{h}_j\}$ in (10.4) satisfy Assumption 2.3 and Assumption 2.4(ii).*

Proof of Lemma 10.1. Notice that assumptions (22) imply $\sum_{j=1}^{n} b_{n,tj} \asymp H$. Thus, the claim of Lemma 10.1 follows using the same argument as in the proof of Lemma 2.1. □

11 Proofs of Propositions 4.1, 4.2 and Theorem 5.1

Proof of Proposition 4.1. Under Assumptions 4.1 and 4.2 and $N \asymp_p n$, regressors $\widetilde{z}_t$, the noise $\widetilde{u}_t$ and the scale factors $\widetilde{g}_t, \widetilde{h}_t$ have property $v_k \asymp_p n^{1/2}$ and satisfy assumptions of Lemma 2.1. This Lemma implies that regression model (29) satisfies Assumptions 2.3 and 2.4(ii), and overall Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Hence assumptions of Theorem 2.2 are satisfied, (16) of Corollary 2.1 remains true, and $\sqrt{\omega_{kk}} \asymp_p v_k^{-1} \asymp_p n^{-1/2}$. □

Proof of Proposition 4.2. Under Assumptions 4.1, 4.2, 3.1 and (32), regression model (31) satisfies Assumptions 2.1, 2.2, 3.1 and 3.2 of Theorem 3.1 and Corollary 3.1 which implies the claims of the proposition. □

Proof of Theorem 5.1. The AR(p) model (33) can be written as regressions model, $y_t = \beta' z_t + \varepsilon_t$ with fixed parameter $\beta = (\beta_1, ..., \beta_{p+1})' = (\phi_0, ..., \phi_p)'$ and stationary regressors $z_t = (z_{1t}, z_{2t}, ..., z_{p+1,t})' = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})'$. They include the intercept 1 and the past lags y_{t-k} which are $\mathcal{F}_t = \sigma(\varepsilon_s, s \le t)$ measurable. For $k \ge 2$, regressors can be written as in (3), $z_{kt} = \mu_{kt} + g_{kt}\eta_{kt} = \mu + (y_{t-k+1} - \mu)$, where the scale factors g_{kt} are equal to 1 and the means $\mu_{kt} = \mu = Ey_1$.

Under assumptions of theorem, $y_t - \mu = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} |a_j| < \infty$. If a stationary $m.d.$ sequence ε_t satisfies $E|\varepsilon_t|^p < \infty$, for some $p > 2$, then

$$E\big|\sum_{j=0}^{\infty} a_j \varepsilon_{t-j}\big|^p \le C\big(\sum_{j=0}^{\infty} a_j^2\big)^{p/2},$$

where $C < \infty$ does not depend on n, see e.g., Lemma 2.5.2 in Giraitis et al. (2012). By assumption of theorem, $E\varepsilon_t^8 < \infty$. Hence $E(y_t - \mu)^8 < \infty$ and $E\eta_{kt}^8 < \infty$.

Thus, regressors z_t and regression noise $u_t = \varepsilon_t$ satisfy Assumptions 2.1, 2.2, 2.3 and 2.4 of Section 2. Therefore, the robust OLS estimator $\widehat{\beta}$ of β has properties derived in Corollary 2.1 which implies Theorem 5.1. □

12 Proofs of Section 2: Auxiliary lemmas

This section contains auxiliary lemmas used in the proofs of the main results for Section 2. For the ease of referencing, we include the statement of Lemma 12.1(i) established in Giraitis et al. (2024).

Lemma 12.1. *Assume that sequences $\{\beta_t\}$ and $\{z_t\}$ are mutually independent.*

(i) *If $\{z_t\}$ is a covariance stationary short memory sequence, then*

$$\sum_{t=1}^{n} \beta_t z_t = \Big(\sum_{t=1}^{n} \beta_t\Big) E z_1 + O_p\Big(\big(\sum_{t=1}^{n} \beta_t^2\big)^{1/2}\Big). \tag{12.1}$$

(ii) *If $E|z_t| < \infty$, then*

$$\big|\sum_{t=1}^{n} \beta_t z_t\big| = O_p\Big(\sum_{t=1}^{n} |\beta_t|\Big)\big(\max_{t=1,...,n} E|z_t|\big). \tag{12.2}$$

Proof of Lemma 12.1. The claim (i) of Lemma 12.1 was derived in (Giraitis et al. (2024), Lemma A5). To prove (ii), denote $s_n = \sum_{t=1}^{n} |\beta_t|$. Then,

$$\begin{aligned} E\big[s_n^{-1} \textstyle\sum_{t=1}^{n} |\beta_t|\,|z_t|\big] &= \textstyle\sum_{t=1}^{n} E[s_n^{-1}|\beta_t|]\, E[|z_t|] \\ &\le (\max_{t=1,...,n} E|z_t|) E[s_n^{-1} \textstyle\sum_{t=1}^{n} |\beta_t|] = \max_{t=1,...,n} E|z_t|, \end{aligned}$$

$$s_n^{-1} \sum_{t=1}^n |\beta_t|\,|z_t| \;=\; O_p\Big(\max_{t=1,...,n} E|z_t|\Big).$$

This implies

$$\big|\sum_{t=1}^n \beta_t z_t\big| \le s_n\big\{s_n^{-1} \sum_{t=1}^n |\beta_t|\,|z_t|\big\} = s_n O_p\big(\max_{t=1,...,n} E|z_t|\big).$$

This completes the proof of (12.2) and the lemma. □

Recall notation

$$\begin{aligned} S_{zz} &= \sum_{t=1}^n z_t z_t', \quad S_{zzuu} = \sum_{t=1}^n z_t z_t' u_t^2, \quad S_{zu} = \sum_{t=1}^n z_t u_t, \\ & D = \mathrm{diag}(v_1, ..., v_p), \quad v_k = (\sum_{t=1}^n g_{kt}^2 h_t^2)^{1/2}, \\ & D_g = \mathrm{diag}(v_{g1}, ..., v_{gp}), \quad v_{gk} = (\sum_{t=1}^n g_{kt}^2)^{1/2}. \end{aligned}$$

Recall definition $\mathcal{F}_n^* = \sigma(\mu_t, g_t, t = 1, ..., n)$ and $\mathcal{F}_{n,t-1}$ in (9.6). Denote

$$\begin{aligned} W_{zz} &= D_g^{-1} E[S_{zz}|\mathcal{F}_n^*] D_g^{-1}, \quad W_{zzuu} = D^{-1} E[S_{zzuu}|\mathcal{F}_n^*] D^{-1}, \\ & \Omega_n = (E[S_{zz}|\mathcal{F}_n^*])^{-1}(E[S_{zzuu}|\mathcal{F}_n^*])(E[S_{zz}|\mathcal{F}_n^*])^{-1}. \end{aligned}$$

Lemma 12.2. *Suppose that z_t and u_t satisfy Assumptions 2.1, 2.2 and 2.3. Then the following holds.*

(i) *There exists $b_n > 0$ such that $b_n^{-1} = O_p(1)$ and such that for any $a = (a_1, ..., a_p)'$, $||a|| = 1$,*

$$a' W_{zz} a \;\ge\; b_n, \quad ||W_{zz}^{-1}||_{sp} \le b_n^{-1}, \tag{12.3}$$

$$||W_{zz}|| \;\le\; b_{2n} = O_p(1). \tag{12.4}$$

Moreover,

$$D_g^{-1} S_{zz} D_g^{-1} \;=\; W_{zz} + o_p(1), \tag{12.5}$$

$$D_g S_{zz}^{-1} D_g \;=\; W_{zz}^{-1} + o_p(1), \tag{12.6}$$

$$D^{-1} S_{zu} \;=\; O_p(1), \tag{12.7}$$

$$\sum_{t=1}^n ||D_g^{-1} z_t||^2 \;=\; O_p(1). \tag{12.8}$$

(ii) In addition, if Assumption 2.4 holds, then there exists $b_n > 0$ such that $b_n^{-1} = O_p(1)$ and such that for any $a = (a_1, ..., a_p)'$, $||a|| = 1$,

$$a' W_{zzuu} a \;\ge\; b_n, \quad ||W_{zzuu}^{-1}||_{sp} \le b_n^{-1}, \tag{12.9}$$

$$||W_{zzuu}|| \;\le\; b_{2n} = O_p(1), \tag{12.10}$$

$$a' D\Omega_n D a \;\ge\; b_n, \quad a' D\Omega_n D a \le b_{2n} = O_p(1). \tag{12.11}$$

Moreover,

$$D^{-1}S_{zzuu}D^{-1} = W_{zzuu} + o_p(1), \tag{12.12}$$

$$DS_{zzuu}^{-1}D = W_{zzuu}^{-1} + o_p(1), \tag{12.13}$$

$$D^{-1}S_{zzuu}^{(c)}D^{-1} = W_{zzuu} + o_p(1), \qquad S_{zzuu}^{(c)} = \sum_{t=1}^{n} z_t z_t' E[u_t^2 \,|\mathcal{F}_{n,t-1}]. \tag{12.14}$$

Before the proof of lemma, we will state the following corollary. Denote

$$c_{*,n} = \sum_{t=1}^{n} ||D_g^{-1}\mu_t||^2, \qquad c_{**,n} = \sum_{t=1}^{n} ||D^{-1}\mu_t h_t||^2. \tag{12.15}$$

Notice that under (7) of Assumption 2.3,

$$c_{*,n} = \sum_{k=1}^{p}\{v_{gk}^{-2}\sum_{t=1}^{n}\mu_{kt}^2\} = O_p(1), \quad c_{**,n} = \sum_{k=1}^{p}\{v_k^{-2}\sum_{t=1}^{n}\mu_{kt}^2 h_t^2\} = O_p(1). \tag{12.16}$$

Corollary 12.1. *In Lemma 12.2, the claims (12.3) and (12.9) hold with b_n as below:*

$$a'W_{zz}a \;\geq\; b_n = \begin{cases} c^{-1}: & \textit{Case 1 (intercept not included),} \\ c^{-1}(1+c_{*,n})^{-1}: & \textit{Case 2 (intercept included),} \end{cases} \tag{12.17}$$

$$a'W_{zzuu}a \;\geq\; b_n = \begin{cases} c^{-1}(1+c_{**,n})^{-4}: & \textit{Case 1 (intercept not included),} \\ c^{-1}(1+c_{**,n})^{-9}: & \textit{Case 2 (intercept included),} \end{cases} \tag{12.18}$$

where $c > 0$ does not depend on n, $b_n^{-1} = O_p(1)$ and b_n is $\mathcal{F}_n^*$ measurable.

Proof of Lemma 12.2(i). *Proof of (12.3).* Set $I_{gt} = \mathrm{diag}(g_{1t}, ..., g_{pt})$. By definition,

$$z_t = \mu_t + I_{gt}\eta_t = \mu_t + \widetilde{z}_t, \quad \widetilde{z}_t = I_{gt}\eta_t. \tag{12.19}$$

Then

$$\begin{aligned} z_t z_t' &= (\mu_t + \widetilde{z}_t)(\mu_t + \widetilde{z}_t)' = \widetilde{z}_t\widetilde{z}_t' + \mu_t\mu_t' + \mu_t\widetilde{z}_t' + \widetilde{z}_t\mu_t', \\ E[z_t z_t'|\mathcal{F}_n^*] &= E[\widetilde{z}_t\widetilde{z}_t'|\mathcal{F}_n^*] + \mu_t\mu_t' + \mu_t E[\widetilde{z}_t'|\mathcal{F}_n^*] + E[\widetilde{z}_t|\mathcal{F}_n^*]\mu_t' \\ &= E[\widetilde{z}_t\widetilde{z}_t'|\mathcal{F}_n^*] + \mu_t\mu_t' + \mu_t e_t' + e_t\mu_t' \\ &= E[\widetilde{z}_t\widetilde{z}_t'|\mathcal{F}_n^*] + (\mu_t + e_t)(\mu_t + e_t)' - e_t e_t', \end{aligned} \tag{12.20}$$

where $e_t = E[\widetilde{z}_t|\mathcal{F}_n^*] = I_{gt}E[\eta_t]$. Using (12.20), we can write

$$a'W_{zz}a = \sum_{t=1}^{n} a'D_g^{-1}E[z_t z_t'|\mathcal{F}_n^*]D_g^{-1}a$$

$$= \sum_{t=1}^{n} a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a + \sum_{t=1}^{n} (a' D_g^{-1} \mu_t)^2 + 2 \sum_{t=1}^{n} (a' D_g^{-1} \mu_t)(e_t' D_g^{-1} a) \quad (12.21)$$

$$= \sum_{t=1}^{n} a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a + \sum_{t=1}^{n} (a' D_g^{-1} (\mu_t + e_t))^2 - \sum_{t=1}^{n} (a' D_g^{-1} e_t)^2. \quad (12.22)$$

We split the proof into two cases when regression model (1) does not include intercept and when intercept is included.

Case 1 (no intercept): $e_t = I_{gt} E[\eta_t] = (0, ..., 0)'$.

Case 2 (intercept included): $e_t = I_{gt} E[\eta_t] = I_{gt}(1, 0, ..., 0)' = (g_{1t}, 0, ..., 0)'$, $g_{1t} = 1$.

Case 1. Let $e_t = 0$. Then (12.21) implies

$$a' W_{zz} a \quad \geq \quad \sum_{t=1}^{n} a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a. \quad (12.23)$$

In this instance,

$$E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] = I_{gt} E[\eta_t \eta_t'] I_{gt} = I_{gt} \Sigma I_{gt},$$

where $E[\eta_t \eta_t'] = \Sigma = (\sigma_{jk})_{j,k=1,...,p}$. By Assumption 2.2(ii), the matrix Σ is positive definite. Therefore, there exists $b > 0$ such that for any $\alpha = (\alpha_1, ..., \alpha_p)'$,

$$\alpha' \Sigma \alpha \geq b ||\alpha||^2.$$

Hence, setting $\gamma_{kt} = v_{gk}^{-1} g_{kt}$, we derive

$$\begin{aligned}
\sum_{t=1}^{n} a' D_g^{-1} E[z_t z_t' | \mathcal{F}_n^*] D_g^{-1} a &= \sum_{t=1}^{n} \{a' D_g^{-1} I_{gt}\} \Sigma \{I_{gt} D_g^{-1} a\} \\
&\geq b \sum_{t=1}^{n} ||a' D_g^{-1} I_{gt}||^2 = b \sum_{t=1}^{n} \big[\sum_{k=1}^{p} a_k^2 \gamma_{kt}^2 \big] \\
&= b \sum_{k=1}^{p} a_k^2 \big(\sum_{t=1}^{n} \gamma_{kt}^2 \big) = b \sum_{k=1}^{p} a_k^2 = b ||a||^2 = b,
\end{aligned}$$

since $\sum_{t=1}^{n} \gamma_{kt}^2 = 1$ and $||a|| = 1$. With (12.23) this proves the first claim in (12.3):

$$a' W_{zz} a \quad \geq \quad b. \quad (12.24)$$

Matrix W_{zz} is symmetric and, thus, it has real eigenvalues. The bound (12.24) implies that the smallest eigenvalue of W_{zz} has property $\lambda_{min} \geq b_n > 0$. Therefore W_{zz} is positive definite, and the largest eigenvalue $\theta_{\max}$ of W_{zz}^{-1} has property $\theta_{\max} = \lambda_{min}^{-1} \leq 1/b_n$, which implies that $||W_{zz}^{-1}||_{sp} \leq 1/b_n$. This proves the second claim in (12.3).

Case 2 (intercept included): $e_t = I_{gt} E[\eta_t] = I_{gt}(1, 0, ..., 0)' = (g_{1t}, 0, ..., 0)'$. Recall that in presence of intercept, $g_{1t} = 1$ and $\eta_{1t} = 1$.

Proof of (12.3). Set $a = (a_1, ..., a_p)'$, $\widetilde{a} = (a_2, ..., a_p)'$. Recall that

$$1 = ||a||^2 = a_1^2 + ... + a_p^2 = a_1^2 + ||\widetilde{a}||^2. \tag{12.25}$$

We will show that there exists $b > 0$ such that for any a and $n \geq 1$,

$$a' W_{zz} a \geq b||\widetilde{a}||^2, \tag{12.26}$$
$$a' W_{zz} a \geq b||\widetilde{a}||^2 + \{a_1^2 - 2|a_1|\,||\widetilde{a}||c_{*,n}^{1/2}\}, \tag{12.27}$$

where $c_{*,n}$ is defined as in (12.15). These bounds imply (12.3). Indeed, suppose that $||\widetilde{a}|| > (1-b)|a_1|/(2c_{*,n}^{1/2})$. By (12.25), this is equivalent to

$$||\widetilde{a}||^2 > \frac{(1-b)^2 a_1^2}{4c_{*,n}} = \frac{(1-b)^2(1-||\widetilde{a}||^2)}{4c_{*,n}}, \quad ||\widetilde{a}||^2 > \frac{(1-b)^2}{(1-b)^2 + 4c_{*,n}}.$$

Then, by (12.26),

$$a' W_{zz} a \geq b||\widetilde{a}||^2 = \frac{b(1-b)^2}{(1-b)^2 + 4c_{*,n}}.$$

On the other hand, if $||\widetilde{a}|| \leq (1-b)|a_1|/(2c_{*,n}^{1/2})$, then in (12.27),

$$a_1^2 - 2|a_1|\,||\widetilde{a}||c_{*,n}^{1/2} \geq a_1^2 - (1-b)a_1^2 = b\,a_1^2$$

which together with (12.27) implies

$$a' W_{zz} a \geq b||\widetilde{a}||^2 + a_1^2 b = b(||\widetilde{a}||^2 + a_1^2) = b||a||^2 = b.$$

Therefore,

$$a' W_{zz} a \geq \min\Big(\frac{b(1-b)^2}{(1-b)^2 + 4c_{*,n}}, b\Big) = \frac{b(1-b)^2}{(1-b)^2 + 4c_{*,n}}.$$

This implies that there exists $c > 0$ such that

$$a' W_{zz} a \geq b_n = c^{-1}(1 + c_{*,n})^{-1}, \tag{12.28}$$

where $b_n^{-1} = c(1 + c_{*,n}) = O_p(1)$ by (12.16). This verifies the first claim in (12.3).

Proof of (12.26). Below we will show that there exists $b > 0$ such that

$$i_n = \sum_{t=1}^{n} a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a \geq a_1^2 + b||\widetilde{a}||^2. \tag{12.29}$$

In addition, observe that in Case 2,

$$e_t' D_g^{-1} a = a_1 v_{g1}^{-1} g_{1t}, \quad \sum_{t=1}^n (a' D_g^{-1} e_t)^2 = a_1^2 v_{g1}^{-2} \sum_{t=1}^n g_{1t}^2 = a_1^2. \tag{12.30}$$

Then from (12.22), using (12.29) and (12.30) we arrive at (12.26):

$$\begin{aligned} a' W_{zz} a &\geq \sum_{t=1}^n a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a - \sum_{t=1}^n (a' D_g^{-1} e_t)^2 \\ &\geq \{a_1^2 + b||\widetilde{a}||^2\} - a_1^2 = b||\widetilde{a}||^2. \end{aligned}$$

Proof of (12.27). By (12.21) and (12.29),

$$\begin{aligned} a' W_{zz} a &\geq \sum_{t=1}^n a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a - 2\big| \sum_{t=1}^n (a' D_g^{-1} \mu_t)(e_t' D_g^{-1} a) \big| \\ &\geq \{a_1^2 + b||\widetilde{a}||^2\} - 2|q_n|, \quad q_n = \sum_{t=1}^n (a' D_g^{-1} \mu_t)(e_t' D_g^{-1} a). \end{aligned} \tag{12.31}$$

By Cauchy inequality and (12.30),

$$|q_n| \leq \big\{ \sum_{t=1}^n (a' D_g^{-1} \mu_t)^2 \sum_{t=1}^n (e_t' D_g^{-1} a)^2 \big\}^{1/2} = |a_1| \big(\sum_{t=1}^n (a' D_g^{-1} \mu_t)^2 \big)^{1/2}.$$

Since $\mu_{1t} = 0$, then $|a' D_g^{-1} \mu_t| \leq ||\widetilde{a}|| \, ||D_g^{-1} \mu_t||$. Hence, using notation $c_{*,n}$ introduced in (12.15), we obtain

$$\sum_{t=1}^n (a' D_g^{-1} \mu_t)^2 \leq ||\widetilde{a}||^2 (\sum_{t=1}^n ||D_g^{-1} \mu_t||^2) = ||\widetilde{a}||^2 c_{*,n},$$

which together with (12.31) and (12.29) proves (12.27):

$$a' W_{zz} a \geq \{a_1^2 + b||\widetilde{a}||^2\} - 2|a_1| ||\widetilde{a}|| c_{*,n}^{1/2} = b||\widetilde{a}||^2 + \{a_1^2 - 2|a_1| \, ||\widetilde{a}|| c_{*,n}^{1/2}\}.$$

Proof of (12.29). Recall, that in presence of intercept, $\eta_t = (1, \eta_{2t}, ..., \eta_{pt})'$ and $E[\eta_{kt}] = 0$. Denote $\tilde{\eta} = (\eta_{2t}, ..., \eta_{pt})'$ and $\widetilde{\Sigma} = E[\tilde{\eta} \tilde{\eta}']$. Then

$$E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] = I_{gt} E[\eta_t \eta_t'] I_{gt} = I_{gt} \mathrm{diag}(1, \widetilde{\Sigma}) I_{gt} = \mathrm{diag}\big(g_{1t}^2, \tilde{I}_{gt} \widetilde{\Sigma} \tilde{I}_{gt}\big),$$

where $\mathrm{diag}(1, \widetilde{\Sigma})$ is a block diagonal matrix and $\widetilde{I}_{gt} = \mathrm{diag}(g_{2t}, ..., g_{pt})$. By assumption, the matrix $\widetilde{\Sigma}$ is positive definite. Denote $\widetilde{D}_g = \mathrm{diag}(v_{g2}, ..., v_{gp})$. Then,

$$i_n = \sum_{t=1}^n a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a$$

$$
\begin{aligned}
&= a_1^2\{v_{g1}^{-2}\sum_{t=1}^n g_{1t}^2\} + \sum_{t=1}^n \widetilde{a}'\widetilde{D}_g^{-1}\widetilde{I}_{gt}\widetilde{\Sigma}\widetilde{I}_{gt}\widetilde{D}_g^{-1}\widetilde{a}\\
&= i_{n,1} + i_{n,2}.
\end{aligned}
$$

Observe that $i_{n,1} = a_1^2$ since $v_{g1}^{-2}\sum_{t=1}^n g_{1t}^2 = 1$. Recall that $||\widetilde{a}|| \le 1$. Hence, by (12.24),

$$
i_{n,2} \ge b||\widetilde{a}||^2, \quad i_n \ge a_1^2 + b||\widetilde{a}||^2
$$

for some $b > 0$ which does not depend on n and a. This implies (12.29).

Summarizing, note that by (12.24) and (12.28),

$$
a'W_{zz}a \ge b_n = \begin{cases} c^{-1}: & \text{Case 1 (intercept not included)},\\ c^{-1}(1+c_{*,n})^{-1}: & \text{Case 2 (intercept included)}, \end{cases} \tag{12.32}
$$

where $c > 0$ does not depend on n. Notice that $b_n^{-1} \le c(1+c_{*,n}) = O_p(1)$ by (12.16). This proves the first claim in (12.3).

Proof of the second claim in (12.3) is the same as in Case 1.

Proof of (12.4). Observe that

$$
\begin{aligned}
||W_{zz}|| &\le ||E\big[(\sum_{t=1}^n D_g^{-1}z_tz_t'D_g^{-1}\big|\mathcal{F}_n^*]|| \le E\big[||\sum_{t=1}^n D_g^{-1}z_tz_t'D_g^{-1})||\big|\mathcal{F}_n^*]\\
&\le \sum_{t=1}^n E[||D_g^{-1}z_t||^2\,|\mathcal{F}_n^*] \le c(1+c_{*,n}) = O_p(1)
\end{aligned}
$$

by (12.51) of Lemma 12.3. This proves (12.4).

Proof of (12.5), (12.6), (12.7) and (12.8). Denote by δ_{jk} the jk-th element of the matrix

$$
D_g^{-1}S_{zz}D_g^{-1} - W_{zz} = \sum_{t=1}^n D_g^{-1}\{z_tz_t' - E[z_tz_t'|\mathcal{F}_n^*]\}D_g^{-1} = \big(\delta_{jk}\big). \tag{12.33}
$$

To prove (12.5), it remains to show that

$$
\delta_{jk} = o_p(1). \tag{12.34}
$$

Case 1: $e_t = 0$. Then, by (12.20), we have

$$
z_tz_t' - E[z_tz_t'|\mathcal{F}_n^*] = I_{gt}(\eta_t\eta_t' - E[\eta_t\eta_t'])I_{gt} + \mu_t\eta_t'I_{gt} + I_{gt}\eta_t\mu_t'.
$$

Therefore, setting $\gamma_{jt} = v_{gj}^{-1} g_{jt}$, we can write

$$\begin{aligned}
\delta_{jk} &= \sum_{t=1}^{n} \gamma_{jt}\gamma_{kt}(\eta_{jt}\eta_{kt} - E[\eta_{jt}\eta_{kt}]) + \sum_{t=1}^{n} \{v_{gj}^{-1}\mu_{jt}\gamma_{kt}\}\eta_{kt} + \sum_{t=1}^{n} \{v_{gk}^{-1}\mu_{kt}\gamma_{jt}\}\eta_{jt} \\
&= S_{n,1} + S_{n,2} + S_{n,3}, \qquad (12.35) \\
\delta_{jk}^2 &\leq 3(S_{n,1}^2 + S_{n,2}^2 + S_{n,3}^2).
\end{aligned}$$

By assumption, sequences $\{w_{1t} = \eta_{jt}\eta_{kt} - E[\eta_{jt}\eta_{kt}]\}$, $\{w_{2t} = \eta_{kt}\}$ and $\{w_{3t} = \eta_{jt}\}$ are covariance stationary short memory sequences with zero mean, and the weights $\{b_{1t} = \gamma_{jt}\gamma_{kt}\}$ are independent of $\{w_{1t}\}$, $\{b_{2t} = v_{gj}^{-1}\mu_{jt}\gamma_{kt}\}$ are independent of $\{w_{2t}\}$ and $\{b_{3t} = v_{gk}^{-1}\mu_{kt}\gamma_{jt}\}$ are independent of $\{w_{3t}\}$, Thus, applying Lemma 12.1 to $S_{n,i}$, $i = 1, 2, 3$, we obtain

$$\delta_{jk}^2 = O_P\Big(\sum_{t=1}^{n}(b_{1t}^2 + b_{2t}^2 + b_{3t}^2)\Big).$$

Denote $r_{jn} = \max_{t=1,...,n} \gamma_{jt}^2$. Then,

$$\sum_{t=1}^{n}(b_{1t}^2 + b_{2t}^2 + b_{3t}^2) \leq r_{jn}\sum_{t=1}^{n}\gamma_{kt}^2 + r_{kn}(v_{gj}^{-2}\sum_{t=1}^{n}\mu_{jt}^2) + r_{jn}(v_{gk}^{-2}\sum_{t=1}^{n}\mu_{kt}^2).$$

Notice that $\sum_{t=1}^{n}\gamma_{kt}^2 = 1$. Observe that $r_{jn} = o_p(1)$ by (6) and $v_{gj}^{-2}\sum_{t=1}^{n}\mu_{jt}^2 = O_p(1)$ by (7) of Assumption 2.3. This implies $\delta_{jk}^2 = o_p(1)$ which proves (12.34).

Case 2. Let $e_t = (1, 0, ..., 0)'$.

To prove (12.5), it suffices to show that δ_{jk}, $j, k = 1, ..., p$ in (12.33) have property (12.34): $\delta_{jk} = o_p(1)$. Recall that in presence of intercept we have $z_t = (1, z_{2t}, ..., z_{pt})'$.

First, observe that for $j, k = 2, ..., p$, δ_{jk} are the same as in (12.35) and whence $\delta_{jk} = o_p(1)$ by (12.34). Second, $\delta_{11} = 0$ since $z_{1t} = 1$. Finally, for $k = 2, ..., p$, we have

$$z_{1t}z_{kt} = z_{kt} = \mu_{kt} + g_{kt}\eta_{kt},$$
$$E[z_{1t}z_{kt}|\mathcal{F}_n^*] = E[z_{kt}|\mathcal{F}_n^*] = \mu_{kt}.$$

Then,

$$\begin{aligned}
\delta_{1k} &= \sum_{t=1}^{n} v_{g1}^{-1}\{z_{1t}z_{kt} - E[z_{1t}z_{kt}|\mathcal{F}_n^*]\}v_{gk}^{-1} \\
&= v_{g1}^{-1}\sum_{t=1}^{n}\{v_{gk}^{-1}g_{kt}\}\eta_{kt} = n^{-1/2}\sum_{t=1}^{n}\gamma_{kt}\eta_{kt}.
\end{aligned}$$

By assumption, $\{\eta_{kt}\}$ is a covariance stationary short memory sequence with $E[\eta_{kt}] = 0$, and

$\{\eta_{kt}\}$ and $\{\gamma_{kt}\}$ are mutually independent. Therefore, by Lemma 12.1,

$$\delta_{1k} = n^{-1/2} O_p\Big((\sum_{t=1}^{n} \gamma_{kt}^2)^{1/2} \Big) = n^{-1/2} O_p(1) = o_p(1)$$

which proves (12.34). This completes the proof of (12.5) in Case 2.

Proof of (12.6). It follows using the same argument as in Case 1.

Proof of (12.7). To prove that $D^{-1} S_{zu} = O_p(1)$, write

$$D^{-1} S_{zu} = \sum_{t=1}^{n} D^{-1} z_t u_t = \sum_{t=1}^{n} D^{-1} (\mu_t + I_{gt} \eta_t) h_t \varepsilon_t = (\nu_1, ..., \nu_p)'.$$

It suffices to show that

$$\nu_k = O_p(1). \tag{12.36}$$

We have

$$\begin{aligned}
\nu_k &= \sum_{t=1}^{n} \{v_k^{-1} \mu_{kt} h_t\} \varepsilon_t + \sum_{t=1}^{n} \{v_k^{-1} g_{kt} h_t\} \eta_{kt} \varepsilon_t \\
&= S_{n,1} + S_{n,2}, \\
\nu_k^2 &\leq 2S_{n,1}^2 + 2S_{n,2}^2.
\end{aligned}$$

By Assumptions, 2.1 and 2.2, the sequences $\{w_{1t} = \varepsilon_t\}$, $\{w_{2t} = \eta_{kt}\varepsilon_t\}$ are covariance stationary short memory sequences with zero mean, the weights $\{b_{1t} = v_k^{-1} \mu_{kt} h_t\}$ are independent of $\{w_{1t}\}$, and $\{b_{2t} = v_k^{-1} g_{kt} h_t\}$ are independent of $\{w_{2t}\}$.

Thus, applying Lemma 12.1 to each of the sum $S_{n,1}, S_{n,2}$, we obtain

$$\nu_k^2 = O_P\Big(\sum_{t=1}^{n} (b_{1t}^2 + b_{2t}^2) \Big).$$

Notice that,

$$\sum_{t=1}^{n} (b_{1t}^2 + b_{2t}^2) = v_k^{-2} \sum_{t=1}^{n} \mu_{kt}^2 h_t^2 + v_k^{-2} \sum_{t=1}^{n} g_{kt}^2 h_t^2 = v_k^{-2} \sum_{t=1}^{n} \mu_{kt}^2 h_t^2 + 1 = O_p(1)$$

by (7) of Assumption 2.3 which proves (12.36).

Proof of (12.8). Observe that by (12.5) and (12.4) of Lemma 12.2, $D_g^{-1} (\sum_{t=1}^{n} z_t z_t') D_g^{-1} = O_p(1)$. Therefore,

$$\sum_{t=1}^{n} ||D_g^{-1} z_t||^2 = \text{trace}\Big(D_g^{-1} (\sum_{t=1}^{n} z_t z_t') D_g^{-1} \Big) = O_p(1).$$

This proves (12.8) and completes the proof of the part (i) of the lemma.

Proof of Lemma 12.2 (ii). *Proof of (12.9).* We can write

$$\begin{aligned} a'W_{zzuu}a &= \sum_{t=1}^n a'D^{-1}E[z_t z_t' u_t^2|\mathcal{F}_n^*]D^{-1}a \\ &= E\big[\big(\sum_{t=1}^n ||a'D^{-1}z_t h_t||^2\varepsilon_t^2\big)|\mathcal{F}_n^*\big]. \end{aligned}$$

Let $\delta > 0$ be a small number which will be selected below. Then,

$$\begin{aligned} \varepsilon_t^2 &= \{\varepsilon_t^2 I(\varepsilon_t^2 \geq \delta) + \delta I(\varepsilon_t^2 < \delta)\} + (\varepsilon_t^2 - \delta)I(\varepsilon_t^2 < \delta) \\ &\geq \delta - \delta I(\varepsilon_t^2 < \delta). \end{aligned}$$

Thus,

$$\begin{aligned} a'W_{zzuu}a &\geq \delta\big\{E\big[\big(\sum_{t=1}^n ||a'D^{-1}z_t h_t||^2\big)|\mathcal{F}_n^*\big] - E\big[\big(\sum_{t=1}^n ||a'D^{-1}z_t h_t||^2 I(\varepsilon_t^2 < \delta)|\mathcal{F}_n^*\big]\big\} \\ &= \delta\{q_{1,n} - q_{2,n}\}. \end{aligned} \tag{12.37}$$

We will show that there exist $b_n > 0$ and $\delta = \delta_n > 0$ such that $b_n^{-1} = O_p(1)$, $\delta_n^{-1} = O_p(1)$ and for any $a = (a_1, ..., a_p)'$, $||a|| = 1$ and $n \geq 1$,

$$q_{1,n} \geq b_n, \tag{12.38}$$

$$q_{2,n} \leq b_n/2. \tag{12.39}$$

Using these bounds in (12.37), we obtain

$$a'W_{zzuu}a \geq b_n^* = \delta_n\{b_n - (b_n/2)\} = \delta_n b_n/2, \quad 1/b_n^* = O_p(1). \tag{12.40}$$

First we prove (12.38). Setting

$$\begin{aligned} Z_t &= \{h_t\mu_t\} + \{h_t I_{gt}\}\eta_t = \mu_t^* + I_{g^*t}\eta_t, \quad \text{where } \mu_t^* = h_t\mu_t,\ g_t^* = h_t g_t, \\ D_{g*} &= (v_{g^*1}, ..., v_{g^*p})', \quad v_{g^*k} = (\textstyle\sum_{t=1}^n g_{kt}^{*\,2})^{1/2}, \end{aligned}$$

we can write

$$q_{1,n} = \sum_{t=1}^n a'D_{g^*}^{-1}E[Z_t Z_t'|\mathcal{F}_n^*]D_{g^*}^{-1}a = a'\,W_{ZZ}\,a.$$

Observe that the variables $Z_t = \mu_t^* + I_{g^*t}\eta_t$ satisfy assumptions of Lemma 12.2(i). Hence by

(12.32),

$$a'W_{ZZ}a \geq b_n = \begin{cases} c^{-1}: & \text{Case 1 (intercept not included)}, \\ c^{-1}(1+c_{**,n})^{-1}: & \text{Case 2 (intercept included)}, \end{cases} \tag{12.41}$$

where $c > 0$ does not depend on n. Notice that $b_n^{-1} \leq c(1 + c_{**,n}) = O_p(1)$ by (12.16). This proves (12.38).

To prove (12.39), recall that $||a|| = 1$. Bound

$$q_{n,2} \leq ||a||^2 q^*_{n,2} = q^*_{n,2}, \quad q^*_{n,2} = \sum_{t=1}^{n} E[||D^{-1}z_t h_t||^2 I(\varepsilon_t^2 < \delta)|\mathcal{F}^*_n].$$

In (12.52) of Lemma 12.3 we show that $q^*_{n,2} \leq c_1(1+c_{**,n})\delta^{1/4}$, where $c_1 > 0$ does not depend on n and $c_{**,n}$ is defined in (12.15). Thus, selecting

$$\delta_n = \Big(\frac{b_n/2}{c_1(1+c_{**,n})}\Big)^4,$$

we obtain $q_{n,2} \leq c_1(1 + c_{**,n})\delta_n^{1/4} = b_n/2$, which proves the bound (12.39). Notice that $\delta_n \leq (2cc_1)^{-4}$ can be made small by selecting large c in (12.41).

In turn, by (12.40),

$$a'W_{zzuu}a \geq (b_n/2)\delta_n = (b_n/2)\Big(\frac{(b_n/2)}{c_1(1+c_{**,n})}\Big)^4$$

where b_n is defined in (12.41). This implies

$$a'W_{zzuu}a \geq b_n^* = \begin{cases} c^{-1}(1+c_{**,n})^{-4}: & \text{Case 1 (intercept not included)}, \\ c^{-1}(1+c_{**,n})^{-9}: & \text{Case 2 (intercept included)} \end{cases} \tag{12.42}$$

for some $c > 0$ which does not depend on n. Notice that b_n^* is $\mathcal{F}^*_n$ measurable, and $(b_n^*)^{-1} \leq c(1 + c_{**,n})^9 = O_p(1)$ by (12.16). This proves the first claim in (12.9). The second claim follows using the same argument as in the proof of (12.3).

Proof of (12.10). Observe that

$$\begin{aligned} ||W_{zzuu}|| &\leq ||E\big[(\sum_{t=1}^{n} D^{-1}z_t z_t' u_t^2 D^{-1})\big|\mathcal{F}^*_n]|| \leq E\big[||\sum_{t=1}^{n} D^{-1}z_t u_t^2 z_t' D^{-1}||\big|\mathcal{F}^*_n] \\ &\leq \sum_{t=1}^{n} E[||D^{-1}z_t u_t||^2 \,|\mathcal{F}^*_n] \leq b_{n3} = c(1+c_{**,n}) = O_p(1) \end{aligned}$$

by (12.51) of Lemma 12.3 which implies (12.10).

Proof of (12.11). Write $D\Omega_n D = W_{zz}^{-1} W_{zzuu} W_{zz}^{-1}$, $(D\Omega_n D)^{-1} = W_{zz} W_{zzuu}^{-1} W_{zz}$. By (12.3), (12.4), (12.9) and (12.10),

$$||D\Omega_n D||_{sp} \leq ||D\Omega_n D|| \leq ||W_{zz}^{-1}||\,||W_{zzuu}||\,||W_{zz}^{-1}|| \leq b_{n4} = O_p(1), \quad (12.43)$$
$$||(D\Omega_n D)^{-1}||_{sp} \leq ||(D\Omega_n D)^{-1}|| \leq ||W_{zz}||\,||W_{zzuu}^{-1}||\,||W_{zz}|| \leq b_{n5} = O_p(1). \quad (12.44)$$

We will show that

$$a' D\Omega_n D a \geq b_n := b_{n5}^{-1}. \quad (12.45)$$

Since $b_n^{-1} = b_{n5} = O_p(1)$ this proves the first claim in (12.11). To verify (12.45), notice that the smallest eigenvalue λ_{min} of the matrix $D\Omega D$ and the largest eigenvalue θ_{max} of the inverse matrix $(D\Omega_n D)^{-1}$ are related by the equality $\theta_{max} = \lambda_{min}^{-1}$. By (12.44), $\theta_{max} \leq b_{n5}$. Thus, for $||a|| = 1$,

$$a' D\Omega_n D a \geq \lambda_{min} = \theta_{max}^{-1} \geq b_n := b_{n5}^{-1},$$

where $b_n^{-1} = b_{n5} = O_p(1)$ which proves (12.45). Finally, by (12.43), for $||a|| = 1$, $a' D\Omega_n D a \leq ||D\Omega_n D||_{sp} \leq b_{n4} = O_p(1)$ which proves the second bound in (12.11).

Proof of (12.12) and (12.13). Write

$$D^{-1} S_{zzuu} D^{-1} - W_{zzuu} = \sum_{t=1}^{n} D^{-1}\{z_t z_t' u_t^2 - E[z_t z_t' u_t^2 | \mathcal{F}_n^*]\} D^{-1} = \big(\delta_{jk}\big).$$

To prove (12.12), it suffices to verify that

$$\delta_{jk} = o_p(1). \quad (12.46)$$

Recall that $z_t = \mu_t + \widetilde{z}_t$ and $u_t = h_t \varepsilon_t$, where $E\varepsilon_t^2 = 1$. Hence,

$$E[u_t^2 | \mathcal{F}_n^*] = h_t^2,$$
$$E[\widetilde{z}_t u_t^2 | \mathcal{F}_n^*] = h_t^2 I_{gt} E[\eta_t \varepsilon_t^2] = I_{gt} h_t^2 \bar{e}, \quad \bar{e} = E[\eta_1 \varepsilon_1^2].$$

By (12.20),

$$z_t z_t' u_t^2 = \widetilde{z}_t \widetilde{z}_t' u_t^2 + \mu_t \mu_t' u_t^2 + \mu_t \widetilde{z}_t' u_t^2 + \widetilde{z}_t \mu_t' u_t^2,$$
$$E[z_t z_t' u_t^2 | \mathcal{F}_n^*] = E[\widetilde{z}_t \widetilde{z}_t' u_t^2 | \mathcal{F}_n^*] + \mu_t \mu_t' E[u_t^2 | \mathcal{F}_n^*] + \mu_t E[\widetilde{z}_t' u_t^2 | \mathcal{F}_n^*] + E[\widetilde{z}_t u_t^2 | \mathcal{F}_n^*] \mu_t'$$
$$= E[\widetilde{z}_t \widetilde{z}_t' u_t^2 | \mathcal{F}_n^*] + \mu_t \mu_t' h_t^2 E[\varepsilon_t^2] + \{h_t \mu_t\} \bar{e}' \{h_t I_{gt}\} + \{h_t I_{gt}\} \bar{e} \{h_t \mu_t'\}.$$

Then,

$$z_t z_t' u_t^2 - E[z_t z_t' u_t^2 | \mathcal{F}_n^*] = h_t I_{gt} (\eta_t \eta_t' \varepsilon_t^2 - E[\eta_t \eta_t' \varepsilon_t^2]) h_t I_{gt} + \mu_t \mu_t' h_t^2 (\varepsilon_t^2 - E[\varepsilon_t^2])$$
$$+ h_t \mu_t (\eta_t' \varepsilon_t^2 - E[\eta_t' \varepsilon_t^2]) h_t I_{gt} + h_t I_{gt} (\eta_t \varepsilon_t^2 - E[\eta_t' \varepsilon_t^2]) h_t \mu_t'.$$

Therefore, setting $\gamma_{jt} = v_j^{-1} g_{jt} h_t$, it follows that

$$\begin{aligned}\delta_{jk} &= \sum_{t=1}^{n} \gamma_{jt}\gamma_{kt}(\eta_{jt}\eta_{kt}\varepsilon_t^2 - E[\eta_{jt}\eta_{kt}\varepsilon_t^2]) + \sum_{t=1}^{n} \{v_j^{-1}\mu_{jt}h_t\}\gamma_{kt}(\eta_{kt}\varepsilon_t^2 - E[\eta_{kt}\varepsilon_t^2]) \\ &\quad + \sum_{t=1}^{n} \{v_k^{-1}\mu_{kt}h_t\}\gamma_{jt}(\eta_{jt}\varepsilon_t^2 - E[\eta_{jt}\varepsilon_t^2]) + \sum_{t=1}^{n} \{v_j^{-1}\mu_{jt}h_t\}\{v_k^{-1}\mu_{kt}h_t\}(\varepsilon_t^2 - E[\varepsilon_t^2]) \\ &= r_{n,jk}^{(1)} + r_{n,jk}^{(2)} + r_{n,jk}^{(3)} + r_{n,jk}^{(4)}.\end{aligned}$$

To prove (12.46), it suffices to show that

$$r_{n,jk}^{(i)} = o_p(1), \quad i = 1, ..., 4. \tag{12.47}$$

By Assumption 2.4, $\{\eta_{jt}\eta_{kt}\varepsilon_t^2\}$, $\{\eta_{kt}\varepsilon_t^2\}$ and $\{\varepsilon_t^2\}$ are covariance stationary short memory zero mean sequences, and these sequences are mutually independent of the weights $\{\gamma_{jt}\gamma_{kt}\}$, $\{v_j^{-1}\mu_{jt}h_t\gamma_{kt}\}$ and $\{(v_j^{-1}\mu_{jt}h_t)(v_k^{-1}\mu_{kt}h_t)\}$. Moreover, definition of v_k and γ_{kt} and (7) of Assumption 2.3 imply that

$$\sum_{t=1}^{n} \gamma_{kt}^2 = 1, \quad v_k^{-2}\sum_{t=1}^{n} \mu_{kt}^2 h_t^2 = O_p(1)$$

and by (11) of Assumption 2.4,

$$\max_{t=1,\dots,n} \gamma_{kt}^2 = o_p(1), \quad v_k^{-2} \max_{t=1,\dots,n} \mu_{kt}^2 h_t^2 = o_p(1).$$

Thus, (12.47) follows by using Lemma 12.1 and applying a similar argument as in the proof of (12.5). This completes the proof of (12.12).

The claim (12.13) follows using (12.12) and property $W_{zzuu}^{-1} = O_p(1)$ of (12.9):

$$\begin{aligned}DS_{zzuu}^{-1}D &= \big(D^{-1}S_{zzuu}D^{-1}\big)^{-1} = \big(W_{zzuu} + o_p(1)\big)^{-1} = W_{zzuu}^{-1}\big(1 + W_{zzuu}^{-1} \times o_p(1)\big)^{-1} \\ &= W_{zzuu}^{-1}\big(1 + o_p(1)\big)^{-1} = W_{zzuu}^{-1} + o_p(1).\end{aligned}$$

Proof of (12.14). Write

$$D^{-1}S_{zzuu}^{(c)}D^{-1} = D^{-1}S_{zzuu}D^{-1} + D^{-1}(S_{zzuu}^{(c)} - S_{zzuu})D^{-1}. \tag{12.48}$$

By (12.12), $D^{-1}S_{zzuu}D^{-1} = W_{zzuu} + o_p(1)$. We will show that

$$D^{-1}(S_{zzuu}^{(c)} - S_{zzuu})D^{-1} = o_p(1), \tag{12.49}$$

which together with (12.48) implies (12.14): $D^{-1}S_{zzuu}^{(c)}D^{-1} = W_{zzuu} + o_p(1)$. We have, $u_t^2 -$

$E[u_t^2|\mathcal{F}_{n,t-1}] = h_t^2(\varepsilon_t^2 - \sigma_t^2)$, where $\sigma_t^2 = E[\varepsilon_t^2\,|\mathcal{F}_{t-1}]$. Write

$$D^{-1}(S_{zzuu} - S_{zzuu}^{(c)})D^{-1} \;=\; \sum_{t=1}^{n} D^{-1}z_t z_t'(u_t^2 - E[u_t^2|\mathcal{F}_{n,t-1}])D^{-1} = \big(\delta_{jk}\big).$$

Then (12.49) follows if we show that

$$\delta_{jk} = o_p(1). \tag{12.50}$$

We have $z_t = \mu_t + \widetilde{z}_t$ and $u_t = h_t\varepsilon_t$. So,

$$\begin{aligned}
&z_t z_t' = \widetilde{z}_t\widetilde{z}_t' + \mu_t\mu_t' + \mu_t\widetilde{z}_t' + \widetilde{z}_t\mu_t',\\
&z_t z_t'(u_t^2 - E[u_t^2|\mathcal{F}_{n,t-1}]) = z_t z_t' h_t^2(\varepsilon_t^2 - \sigma_t^2)\\
&\quad = h_t I_{gt}\eta_t\eta_t' I_{gt} h_t(\varepsilon_t^2 - \sigma_t^2) + \mu_t\mu_t' h_t^2(\varepsilon_t^2 - \sigma_t^2)\\
&\qquad\qquad\qquad\qquad + h_t\mu_t\eta_t' I_{gt} h_t(\varepsilon_t^2 - \sigma_t^2) + I_{gt}\eta_t\mu_t' h_t^2(\varepsilon_t^2 - \sigma_t^2).
\end{aligned}$$

Hence, denoting $\gamma_{jt} = v_j^{-1}g_{jt}h_t$, we obtain

$$\begin{aligned}
\delta_{jk} \;&=\; \sum_{t=1}^{n}\gamma_{jt}\gamma_{kt}\{\eta_{jt}\eta_{kt}(\varepsilon_t^2 - \sigma_t^2)\} + \sum_{t=1}^{n}\{v_j^{-1}\mu_{jt}h_t\}\gamma_{kt}\{\eta_{kt}(\varepsilon_t^2 - \sigma_t^2)\}\\
&\quad + \sum_{t=1}^{n}\{v_k^{-1}\mu_{kt}h_t\}\gamma_{jt}\{\eta_{jt}(\varepsilon_t^2 - \sigma_t^2)\} + \sum_{t=1}^{n}\{v_j^{-1}\mu_{jt}h_t\}\{v_k^{-1}\mu_{kt}h_t\}\{\varepsilon_t^2 - \sigma_t^2\}\\
&=\; r_{n,jk}^{(1)} + r_{n,jk}^{(2)} + r_{n,jk}^{(3)} + r_{n,jk}^{(4)}.
\end{aligned}$$

Observe, that sequences $\{w_{1t} = \eta_{jt}\eta_{kt}(\varepsilon_t^2 - \sigma_t^2)\}$, $\{w_{2t} = \eta_{kt}(\varepsilon_t^2 - \sigma_t^2)\}$, $\{w_{3t} = \eta_{jt}(\varepsilon_t^2 - \sigma_t^2)\}$, $\{w_{4t} = \varepsilon_t^2 - \sigma_t^2\}$ are sequences of uncorrelated random variables with zero mean and constant variance. For example, by assumption, $\eta_{jt}\eta_{kt}$ are $\mathcal{F}_{t-1}$ measurable. Then, for $t \geq s$,

$$\begin{aligned}
E[w_{1t}] \;&=\; E\big[E[w_{1t}|\mathcal{F}_{t-1}]\big] = E\big[\eta_{jt}\eta_{kt}E[(\varepsilon_t^2 - \sigma_t^2)|\mathcal{F}_{t-1}]\big] = 0,\\
E[w_{1t}w_{1s}] \;&=\; E\big[\eta_{jt}\eta_{kt}\eta_{js}\eta_{ks}(\varepsilon_s^2 - \sigma_s^2)E[(\varepsilon_t^2 - \sigma_t^2)|\mathcal{F}_{t-1}]\big] = 0,\\
E[w_{1t}^2] \;&=\; E\big[\eta_{jt}^2\eta_{kt}^2E[(\varepsilon_t^2 - \sigma_t^2)^2|\mathcal{F}_{t-1}]\big]\\
&\leq\; E\big[\eta_{jt}^2\eta_{kt}^2E[\varepsilon_t^4|\mathcal{F}_{t-1}]\big] = E\big[E[\eta_{jt}^2\eta_{kt}^2\varepsilon_t^4|\mathcal{F}_{t-1}]\big] = E\big[\eta_{j1}^2\eta_{k1}^2\varepsilon_1^4\big] < \infty.
\end{aligned}$$

Then using the same argument as in the proof of (12.47) it follows

$$r_{n,jk}^{(i)} = o_p(1), \quad i = 1, ..., 4.$$

which proves (12.50) and completes the proof of (12.14).

This completes the proof of the part (ii) and of the lemma. □

Proof of Corollary 12.1. The claim (12.17) is shown in (12.32), and the claim (12.18) is shown in (12.42). $\square$

Lemma 12.3. *Under Assumptions of Theorem 2.1, the exists $c > 0$ such that*

$$\sum_{t=1}^{n} E\big[||D_g^{-1} z_t||^2 \; |\mathcal{F}_n^*\big] \le c(1 + c_{*,n}), \qquad \sum_{t=1}^{n} E\big[||D^{-1} z_t u_t||^2 \; |\mathcal{F}_n^*\big] \le c(1 + c_{**,n}), \tag{12.51}$$

$$\sum_{t=1}^{n} E\big[||D^{-1} z_t h_t||^2 I(\varepsilon_t^2 < \delta)|\mathcal{F}_n^*\big] \le c(1 + c_{**,n})\delta^{1/4}, \tag{12.52}$$

for sufficiently small $\delta > 0$, where c does not depend on n and δ and $c_{*,n} = O_p(1)$, $c_{**,n} = O_p(1)$.

In addition, under assumptions of Theorem 2.2,

$$\max_{t=1,\dots,n} ||D^{-1} z_t u_t||^2 = o_p(1), \qquad \max_{t=1,\dots,n} ||D_g^{-1} z_t||^2 = o_p(1), \tag{12.53}$$

$$\sum_{t=1}^{n} E\big[b_n^{-1} ||D^{-1} z_t u_t||^2 I\big(b_n^{-1} ||D^{-1} z_t u_t||^2 \ge \epsilon\big) \, |\mathcal{F}_{n,t-1}\big] = o_p(1) \text{ for any } \epsilon > 0, \tag{12.54}$$

where b_n is $\mathcal{F}_n^*$ measurable, $b_n^{-1} = O_p(1)$ and $\mathcal{F}_{n,t-1}$ is defined as in (9.6).

Proof of Lemma 12.3. *Proof of (12.51).* Denote

$$b_{1t} = ||D_g^{-1} \mu_t||^2 + ||D_g^{-1} I_{gt}||^2, \quad \theta_{1t} = 1 + ||\eta_t||^2,$$
$$b_{2t} = ||D_g^{-1} \mu_t h_t||^2 + ||D_g^{-1} I_{gt} h_t||^2, \quad \theta_{2t} = \varepsilon_t^2 + ||\eta_t||^2 \varepsilon_t^2.$$

By (12.19),

$$\begin{aligned} ||D_g^{-1} z_t||^2 = ||D_g^{-1} \mu_t + D_g^{-1} I_{gt} \eta_t||^2 &\le 2(||D_g^{-1} \mu_t||^2 + ||D_g^{-1} I_{gt}||^2 ||\eta_t||^2) \\ &\le 2 b_{1t} \theta_{1t}, \\ ||D_g^{-1} z_t u_t||^2 = ||D_g^{-1} \mu_t h_t \varepsilon_t + D_g^{-1} I_{gt} \eta_t h_t \varepsilon_t||^2 &\le 2 b_{2t} \theta_{2t}. \end{aligned} \tag{12.55}$$

By Assumption 2.2(i) and Assumption 2.4(i),

$$E[\theta_{1t} \; |\mathcal{F}_n^*] = E[\theta_{1t}] = E[\theta_{11}], \quad E[\theta_{2t} \; |\mathcal{F}_n^*] = E[\theta_{2t}] = E[\theta_{21}].$$

This implies

$$\begin{aligned} E\big[||D_g^{-1} z_t||^2 \; |\mathcal{F}_n^*\big] &\le 2 b_{1t} E[\theta_{11}], \\ E\big[||D^{-1} z_t u_t||^2 \; |\mathcal{F}_n^*\big] &\le 2 b_{2t} E[\theta_{21}], \\ \textstyle\sum_{t=1}^{n} E\big[||D_g^{-1} z_t||^2 \; |\mathcal{F}_n^*\big] &= 2E[\theta_{11}](\textstyle\sum_{t=1}^{n} b_{1t}), \\ \textstyle\sum_{t=1}^{n} E\big[||D_g^{-1} z_t u_t||^2 \; |\mathcal{F}_n^*\big] &= 2E[\theta_{21}](\textstyle\sum_{t=1}^{n} b_{2t}). \end{aligned} \tag{12.56}$$

Notice that

$$\begin{aligned} \textstyle\sum_{t=1}^n b_{1t} = \sum_{t=1}^n ||D_g^{-1}\mu_t||^2 + \sum_{t=1}^n ||D_g^{-1}I_{gt}||^2 = c_{*,n} + p, \\ \textstyle\sum_{t=1}^n b_{2t} = \sum_{t=1}^n ||D^{-1}\mu_t h_t||^2 + \sum_{t=1}^n ||D^{-1}I_{gt}h_t||^2 = c_{**,n} + p, \end{aligned} \tag{12.57}$$

by definition (12.15) of $c_{*,n}$ and $c_{**,n}$ and because

$$\begin{aligned} \textstyle\sum_{t=1}^n ||D_g^{-1}I_{gt}||^2 = \sum_{k=1}^p v_{gk}^{-2}(\sum_{t=1}^n g_{kt}^2) = p, \\ \textstyle\sum_{t=1}^n ||D^{-1}I_{gt}h_t||^2 = \sum_{k=1}^p v_k^{-2}(\sum_{t=1}^n g_{kt}^2 h_t^2) = p. \end{aligned}$$

Moreover, $c_{*,n} = O_p(1)$, $c_{**,n} = O_p(1)$ by (12.16). Clearly, (12.56) and (12.57) prove (12.51).

Proof of (12.52). Denote

$$\theta_{2t}(\delta) = I(\varepsilon_t^2 < \delta) + ||\eta_t||^2 I(\varepsilon_t^2 < \delta).$$

Recall, that by assumption, ε_t is a stationary sequence, and by Assumption 2.2(i), $E[||\eta_t||^4] = E[||\eta_1||^4]$. Then,

$$\begin{aligned} E[\theta_{2t}(\delta)] &\le E[I(\varepsilon_t^2 < \delta)] + (E[||\eta_t||^4)^{1/2}(E[I(\varepsilon_t^2 < \delta)])^{1/2} \\ &= E[I(\varepsilon_1^2 < \delta)] + (E[||\eta_1||^4)^{1/2}(E[I(\varepsilon_1^2 < \delta)])^{1/2}. \end{aligned}$$

We will show that for sufficiently small $\delta > 0$,

$$E[I(\varepsilon_1^2 < \delta)] \le C\delta^{1/2}.$$

Indeed, by Assumption 2.1, the variable ε_1 has probability distribution density $f(x)$ and $f(x) \le c < \infty$ when $|x| \le x_0$ for some $x_0 > 0$. Without restriction of generality assume that $\delta \le x_0$. Then,

$$E[I(\varepsilon_1^2 < \delta)] = \textstyle\int I(|x| \le \delta^{1/2})f(x)dx \le c\int I(|x| \le \delta^{1/2})dx \le C\delta^{1/2}.$$

Therefore, $E[\theta_{2t}(\delta)] \le C\delta^{1/4}$, and as in (12.56), we obtain

$$\begin{aligned} &E\big[||D^{-1}z_t h_t||^2 I(\varepsilon_t^2 < \delta)\ |\mathcal{F}_n^*\big] \le 2b_{2t}E[\theta_{2t}(\delta)] \le C\delta^{1/4}b_{2t}, \\ &\sum_{t=1}^n E\big[||D^{-1}z_t h_t||^2 I(\varepsilon_t^2 < \delta)\,|\mathcal{F}_n^*\big] \le C\delta^{1/4}(\sum_{t=1}^n b_{2t}) \le C\delta^{1/4}(p + c_{**,n}), \end{aligned}$$

which proves (12.52).

Proof of (12.53). We will prove the first claim (the proof of the second claim is similar). By (12.55), $||D^{-1}z_t u_t||^2 \le 2b_{2t}\theta_{2t}$. Let $K > 0$ be a large number. Then, $\theta_{2t} \le K + \theta_{2t}I(\theta_{2t} \ge K)$. Therefore,

$$\max_{t=1,\dots,n} ||D^{-1}z_t u_t||^2 \le 2K(\max_{t=1,\dots,n} b_{2t}) + 2\sum_{t=1}^n b_{2t}\,\theta_{2t}I(\theta_{2t} \ge K). \tag{12.58}$$

By (11) of Assumption 2.4 and (12.57),

$$\max_{t=1,\dots,n} b_{2t} = o_p(1), \qquad \sum_{t=1}^{n} b_{2t} = O_p(1). \tag{12.59}$$

Since $\{b_t\}$ and $\{\theta_{2t}\}$ are mutually independent, then by (12.2) of Lemma 12.1,

$$\sum_{t=1}^{n} b_{2t}\, \theta_{2t} I(\theta_{2t} \geq K) = O_p\big(\sum_{t=1}^{n} b_{2t}\big)\Delta_{n,K}, \quad \Delta_{n,K} = \max_{t=1,\dots,n} E[\theta_{2t} I(\theta_{2t} \geq K)]. \tag{12.60}$$

We will show that

$$\Delta_{n,K} \leq \Delta_K, \tag{12.61}$$

where $\Delta_K \to 0$, $K \to \infty$ and Δ_K does not depend on n. Together with (12.58) this implies

$$\max_{t=1,\dots,n} ||D^{-1} z_t u_t||^2 \leq K o_p(1) + O_p(1)\Delta_K = o_p(1), \quad n, K \to \infty.$$

Next we prove (12.61). Set $L = K^{1/4}$. Then, letting $\varepsilon^{2+}_{L,t} = \varepsilon^2_t I(\varepsilon^2_t > L)$, we obtain

$$\begin{aligned}
\theta_{2t} &= \varepsilon^2_t(||\eta_t||^2 + 1) \leq \{\varepsilon^{2+}_{L,t} + L I(\varepsilon^2_t \leq L)\}(||\eta_t||^2 + 1),\\
\theta_{2t} I(\theta_{2t} \geq K) &\leq \varepsilon^{2+}_{L,t}(||\eta_t||^2 + 1) + L(||\eta_t||^2 + 1) I\big(L(||\eta_t||^2 + 1) \geq K\big),\\
E[\theta_{2t} I(\theta_{2t} \geq K)] &\leq (E[(\varepsilon^{2+}_{L,t})^2])^{1/2} (E[(||\eta_t||^2 + 1)^2])^{1/2} + L E[(||\eta_t||^2 + 1)^4](K/L)^{-1}\\
&\leq (E[(\varepsilon^{2+}_{L,1})^2])^{1/2} (E[(||\eta_1||^2 + 1)^2])^{1/2} + (L^2/K) E[(||\eta_1||^2 + 1)^2]\\
&=: \Delta_K \to 0, \quad K \to \infty
\end{aligned}$$

since, as $K \to \infty$, $L^2/K = K^{-1/2} \to 0$, $E[(\varepsilon^{2+}_{L,1})^2] \to 0$ and $E[||\eta_1||^4 < \infty$. This implies (12.61).

Proof of (12.54). Denote by i_n the left hand side of (12.54). By (12.55), $||D^{-1} z_t u_t||^2 \leq 2 b_{2t} \theta_{2t}$. Let $K > 0$ be a large number. Then,

$$\begin{aligned}
&b_n^{-1} ||D^{-1} z_t u_t||^2 I(b_n^{-1} ||D^{-1} z_t u_t||^2 \geq \epsilon) \leq 2 b_n^{-1} b_{2t} \theta_{2t} I\big(2 b_n^{-1} b_{2t} \theta_{2t} \geq \epsilon\big)\\
&\quad \leq 2 b_n^{-1} b_{2t} K I\big(2 b_n^{-1} b_{2t} K \geq \epsilon\big) I(\theta_{2t} \leq K) + 2 b_n^{-1} b_{2t} \theta_{2t} I(\theta_{2t} > K)\\
&\quad \leq K(\epsilon/K)^{-1} (2 b_n^{-1} b_{2t})^2 + 2 b_n^{-1} b_{2t} \theta_{2t} I(\theta_{2t} > K).
\end{aligned}$$

Observe, that $b_n^{-1} b_{2t}$ is $\mathcal{F}_{n,t-1}$ measurable. Then,

$$i_n \leq K(\epsilon/K)^{-1} (2 b_n^{-1})^2 \textstyle\sum_{t=1}^{n} b_{2t}^2 + 2 b_n^{-1} \sum_{t=1}^{n} b_{2t} \theta_{2t} I(\theta_{2t} > K).$$

Together with (12.60), (12.61) and (12.59), this implies:

$$i_n \quad \leq \quad K(\epsilon/K)^{-1} (2 b_n^{-1})^2 (\max\nolimits_{t=1,\dots,n} b_{2t}) (\textstyle\sum_{t=1}^{n} b_{2t}) + 2 b_n^{-1} (\sum_{t=1}^{n} b_{2t}) \Delta_K$$

$$\leq \quad K(\epsilon/K)^{-1}O_p(1)o_p(1) + O_p(1)\Delta_K = o_p(1), \quad n, K \to \infty.$$

This proves (12.54) and completes the proof of the lemma. □

References

Giraitis, L., H. L. Koul, and D. Surgailis (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press.

Giraitis, L., Y. Li, and P. C. B. Phillips (2024). Robust inference on correlation under general heterogeneity. *Journal of Econometrics 240*(1), 105691.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic press.

School of Economics and Finance

Queen Mary
University of London

This working paper has been produced by
the School of Economics and Finance at
Queen Mary University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/