Unlocking the Regression Space

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

[1]Queen Mary University of London

[2]King's College London

November 11, 2025

Abstract

This paper introduces and analyzes a framework that accommodates general heterogeneity in regression modeling. It demonstrates that regression models with fixed or time-varying parameters can be estimated using the OLS and time-varying OLS methods, respectively, across a broad class of regressors and noise processes not covered by existing theory. The proposed setting facilitates the development of asymptotic theory and the estimation of robust standard errors. The robust confidence interval estimators accommodate substantial heterogeneity in both regressors and noise. The resulting robust standard error estimates coincide with White's (1980) heteroskedasticity-consistent estimator but are applicable to a broader range of conditions, including models with missing data. They are computationally simple and perform well in Monte Carlo simulations. Their robustness, generality, and ease of implementation make them highly suitable for empirical applications. Finally, the paper provides a brief empirical illustration.

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

JEL Classification: C13, C14, C50

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 largely established for over half of a century (see e.g. Lai and Wei (1982)). The problem of robust estimation has long been a focus of empirical work in economics, beginning with the seminal work by White (1980). Its importance is well understood in applied econometrics. At the same time, several important concerns have been raised by applied researchers. Angrist and Pischke (2010) noted that "Leamer (1983) diagnosed his contemporaries' empirical work as suffering from a distressing lack of robustness to changes in key assumptions", and Leamer (2010) later reflected that "sooner or later, someone articulates the concerns that gnaw away in each of us and asks if the Assumptions are valid." Similarly, Karmakar et al. (2022) observed, that the assumption of parameter constancy, or "stationarity is often an oversimplified assumption that ignores systematic deviations of parameters from constancy". Clearly, this concern extends beyond parameter stability to encompass the stability of regressors, regression noise, and the underlying modelling assumptions.

In this paper, we focus on the inherent capacity of regression modelling to accommodate the effects of structural change in settings with both fixed and time-varying parameters. Many such structural changes influence not only the model parameters but also the regression space itself. This space comprises both the regressors and regression noise, and improper treatment of these components may result in incorrect inferences, misinterpretations, and forecasting distortions. We therefore examine which specifications of regression space can flexibly account for structural change while still enabling estimation of both fixed and time-varying regression parameters, the construction of confidence intervals, and the computation of standard errors.

Among recent developments, Wu (2005), Hall et al. (2012), and others, have proposed advanced theoretical methods for the estimation of the fixed parameters, while Cattaneo et al. (2018), Jochmans (2019) developed procedures to estimate both fixed parameters and standard errors in regression models with an increasing number of covariates and heteroscedasticity. Meanwhile, Li et al. (2020), Sun et al. (2021) and Linton and Xiao (2019) introduced new modelling frameworks that explicitly account for structural change. A common response to concerns about heteroskedasticity in the recent literature is the use of heteroscedasticity-robust variance and standard error estimators for linear regression models, see Eicker (1963), White (1980), MacKinnon (2012) and Cattaneo et al. (2018), among others.

There is also a sizeable and growing literature on the estimation of time-varying coefficient regression models, including works of Fan and Zhang (1999), Vogt (2012), among others. This literature further explores tests for different types of parameter variation, see e.g. Bai and Perron (1998), Zhang and Wu (2012), Zhang and Wu (2015), Hu et al. (2024). In addition, specification tests and tests for parameter instability have received significant attention, with important contributions by Hansen (2000), Georgiev et al. (2018), Hidalgo et al. (2019), Boldea et al. (2019), Fu et al. (2023), and others.

The modelling of deterministic, smooth parameter evolution has a long history in statistics. Early examples include linear processes with time-varying spectral densities, introduced by Priestley (1965). This framework is essentially nonparameteric and it has been further developed by Robinson (1989), Robinson (1991), Dahlhaus (1997), Dahlhaus et al. (2019), Dahlhaus and Richter (2023), some of whom refer to these processes as *locally stationary*. The estimation of time-varying parameters, as well as fixed parameters under heteroskedasticity in time series models, has been studied in Dahlhaus and Giraitis (1998), Xu and Phillips (2008), Giraitis et al. (2020), among other. Nonlinear time-varying time series models have also been developed by Doukhan and Wintenberger (2008), Bardet and Wintenberger (2009), Vogt (2012) and Karmakar et al. (2022). Despite these advances, such approaches have not been not been widely adopted in applied economics, where random coefficient models remain more prevalent.

Various methods have been proposed over the years to identify and handle structural change. Early contributions assumed that changes were deterministic, rare, and abrupt. Testing for parameter breaks dates back to the pioneering work by Chow (1960), with further contributions by Brown et al. (1975), Ploberger and Krämer (1992), among others. More recent approaches allow for random evolution of parameters, where changes may be discrete, as in Markov Switching models by Hamilton (1989) or threshold models by Tong (1990), or continuous as in smooth transition models by Terasvirta (1998), or those driven by unobservable shocks, as in random coefficient models by Nyblom (1989a). For example, Cogley and Sargent (2005) use random coefficient models to study stochastic volatility, while Primiceri (2005) examines whether changes in parameters or in the variance of shocks - policy-induced or otherwise - contributed to the period of macroeconomic calmness known as the "Great Moderation" after 1985. In these frameworks, parameters typically evolve as random walks or autoregressive processes.

Building on this literature, Giraitis et al. (2014), Giraitis et al. (2018), Dendramis et al. (2021), and others have developed a theoretical time series framework for random coefficient models and their estimation using kernel-based methods, which performs well in finite samples. These methods are computationally simple and straightforward to implement in applied research. For example, Chronopoulos et al. (2022) demonstrated the empirical prevalence of persistent volatility, suggesting that GARCH-type volatility structures may be less common than previously thought. Nevertheless, a full treatment of estimation and inference within a general regression framework has, surprisingly, not yet been provided.

In this paper, we provide a rigorous validation of the asymptotic normality of the feasible t-statistic for the estimation of both fixed and time-varying parameters in linear regression models within an extended regresion space of regressors and regression disturbances. Our main objective is to describe, in transparent terms, the extended regression space under which such normality is preserved. The class of admissible regressors and regression noises is broad. Regressors are obtained by rescaling and shifting stationary short-memory sequences, while regression errors are generated by arbitrary rescaling of a stationary martingale difference sequence. The restrictions imposed on the scale factors and mean processes are weak, allowing these to be either deterministic or stochastic, and to vary over time, possibly abruptly or through non-stationary (e.g. unit-root) dynamics. Some assumptions on the scale factors are necessary and resemble the Lindeberg condition in the classical Lindeberg–Feller central limit theorem. Importantly, the robust feasible t-statistic retains the same form and limiting distribution as in the standard setting. The infeasible robust standard errors coincide with the heteroskedasticity-consistent standard error estimator of White (1980). Our assumptions do not rely on mixing or near-epoch dependence conditions, which prevail throughout the existing literature. Given the generality of the regression space, these assumptions typically require no additional empirical verification.

The estimation framework for fixed regression parameters is developed in Section 2, which introduces the extended regression space, the underlying assumptions, and the main theoretical results. Section 3 establishes the estimation theory for time-varying regression parameters within the same framework. The proofs highlight how the results for the fixed-parameter case naturally extend to time-varying settings, with only negligible additional terms.

Our results are complementary to existing frameworks. The novelty lies in providing a methodological foundation that confirms the validity of robust regression estimation in an extended regression space. The fundamental theory in this area traces back to Lai and Wei (1982), who studied regression models with heteroskedastic martingale difference noise under eigenvalue-based assumptions. Alternative methods, such as bootstrap procedures, see Hall et al. (2012); Boldea et al. (2019), are widely used in regression analysis but may not be directly applicable to such a general class of regressors and regression noises. In contrast, we demonstrate that White-type standard errors remain applicable and computationally straightforward.

All theoretical results are supported by detailed, rigorous proofs. Monte Carlo simulations confirm that the proposed robust regression estimators perform well in finite samples. Overall, the framework developed in this paper is particularly suited to modelling economic and financial data, where heterogeneity, structural change, and dependence are inherent features.

The remainder of this paper is organised as follows. Section 2 presents the regression setting with the extended regression space, accommodating heterogeneity and dependence, and outlines the theoretical results for infeasible and feasible t-statistics in the case of fixed parameters. Section 3 extends the analysis to the time-varying regression parameters. Section 4 addresses regression modelling with missing data patterns. Section 5 illustrates the flexibility of our robust estimation method by its application to the estimation of an AR(p) model generated by a stationary martingale difference noise. Sections 6 presents Monte Carlo simulation results. In Section 7, we provide an empirical application of the robust regression framework to modelling asset returns. Finally, Section 8 concludes. Proofs and additional simulations are provided in the Supplemental Material.

2 OLS estimation in general regression space

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

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

where β is a p-dimensional parameter vector, $z_t = (z_{1t}, ..., z_{pt})'$ is a stochastic regressor and u_t is an uncorrelated noise term. To include an intercept, the first component can be set as $z_{1t} = 1$. We refer to the collection of $\{z_t, u_t\}$ jointly as "the regression space".

An applied researcher may want to work within a regression space that accommodates a wide range of regressors and regression noises, without being hindered by restrictive technical assumptions. Ideally, such a setting should permit regressors exhibiting non-stationarity and undefined generic structural change, while enabling estimation and inference under weak theoretical constraints that do not require empirical verification.

Our goal is to extend the OLS estimation procedure to a broad regression framework defined by baseline assumptions aligned with empirical research practice. These assumptions cover a wide variety of types of potentially non-stationary regression variables encountered in applied work. The framework achieves a level of generality comparable to that in Giraitis et al. (2024), which addresses testing for absence of correlation and cross-correlation under general heterogeneity.

We begin with specifying the structure of an uncorrelated regression noise u_t. 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 independent of $\{\varepsilon_t\}$. The following assumption formalizes these conditions.

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

$$\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}\varepsilon_t^8 < \infty, \quad \mathbb{E}\varepsilon_t^2 = 1.$$

The sequence $\{\varepsilon_t\}$ is independent of $\{h_t\}$. Moreover, variable ε_1 has a probability density function $f(x)$ satisfying $f(x) \le c < \infty$ for all $|x| \le x_0$, for some $x_0 > 0$.

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

A typical example of an m.d. noise in applied work is provided by the ARCH/GARCH family and the class of stochastic volatility processes. The specification (2) therefore allows for conditional heteroskedasticity in u_t.

We next specify the regressors $z_t = (z_{1t}, ..., z_{pt})'$ which form the key structural component of our regression space. For $k = 1, ..., p$, the regressors can be written as

$$z_{kt} = \mu_{kt} + g_{kt}\eta_{kt}, \quad t = 1, ..., n, \tag{3}$$

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 include an intercept in model (1), we set

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

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

In summary, the admissible regressors $\{z_t\}$ in our setting are obtained by shifting and rescaling a short-memory stationary process $\{\eta_t\}$ by the mean process μ_t and the scale factor g_t:

$$z_t = \mu_t + I_{gt}\eta_t, \quad I_{gt} = \mathrm{diag}(g_{1t}, ..., g_{pt})'.$$

The underlying stationary sequence $\{\eta_t\}$ is the fundamental component structuring regressors z_t. Estimation of the regression parameter β requires only mild assumptions on $\{\mu_t, g_t\}$, and short-memory dependence assumption on η_t, satisfied by ARMA and related stationary time series models. This framework eliminates the need for additional empirical validation.

Definition 2.1. *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 an $\mathcal{F}_{t-1}$ measurable sequence with $E[\eta_{kt}^2] = 1$ and $E[\eta_{kt}^8] < \infty$.*

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

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

The novelty of this regression framework lies in the structural specification (3), which accommodates regressors $z_t = (z_{1t}, ..., z_{pt})'$ that may be deterministic or stochastic, and stationary or non-stationary. This flexibility arises from allowing a broad class of scale factors and mean processes $\{h_t, g_t, \mu_t\}$ which brings the OLS estimation closer to empirical practice.

The estimation framework also accommodates triangular arrays of means and scale factors: $\big(\mu_t, g_t, h_t,\ \ t = 1, ..., n\big) = \big(\mu_{nt}, g_{nt}, h_{nt},\ \ t = 1, ..., n\big)$. Throughout the paper, we assume that these quantities may depend on the sample size n. For brevity of notation, the subscript n is omitted.

The underlying stationary noise component η_t in the regressors z_t in (3) is weakly exogenous with respect to the stationary m.d. noise ε_t in $u_t = h_t\varepsilon_t$. The mean and scale factors $\{\mu_t, g_t\}$ are independent of $\{\varepsilon_t\}$, though they may be dependent on $\{h_t\}$. Overall, $\{\mu_t, g_t, h_t\}$ are mutually independent of $\{\eta_t, \varepsilon_t\}$, while potential dependence among $\{\mu_t\}$, $\{g_t\}$ and $\{h_t\}$ is unrestricted.

The processes μ_{kt} and g_{kt} can be interpreted as conditional mean and variance, $\mu_{kt} = E[z_{kt}\,|\mathcal{F}_n^*]$, and $g_{kt}^2 = \mathrm{var}(z_{kt}|\mathcal{F}_n^*)$ of z_{kt}, where $\mathcal{F}_n^* = \sigma\big(\mu_t, g_t, h_t, t = 1, ..., n\big)$ denotes the information set generated by the means and scale factors.

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

$$\begin{aligned} v_k^2 &= \sum_{t=1}^n g_{kt}^2 h_t^2, \quad v_{gk}^2 = \sum_{t=1}^n g_{kt}^2, \\ D &= \mathrm{diag}(v_1, ..., v_p), \quad D_g = \mathrm{diag}(v_{g1}, ..., v_{gp}). \end{aligned} \tag{5}$$

We write $a_n \asymp_p b_n$ 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 $k = 1, ..., p$,*

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

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

Assumptions (6)–(7) impose only mild restrictions on the means μ_t and scale factors g_t. In particular, condition (6) resembles the Lindeberg condition in the classical Lindeberg–Feller central limit theorem, as it excludes the possibility that the OLS estimation is dominated by a single extreme observation of z_t.

The first restriction on g_{kt} in (6) is necessary. For example, consider the regressor $z_t = g_t\eta_t, t = 1, ..., n$, with scale factors $g_1 = 1$ and $g_2 = g_3 = ... = g_n = 0$, so that $z_2 = z_3 = ... = z_n = 0$. In this case, the OLS estimator of β is inconsistent, and such a scale factor g_t does not satisfy (6).

The second condition (7) ensures that OLS estimation is driven by the stochastic component $g_t\eta_t$ of the regressor z_t, rather than by deterministic or stochastic drift in μ_t.

In the presence of an intercept, condition (7) further implies that $\sum_{t=1}^n h_t^2 \asymp_p n$, since $v_1^2 \asymp_p v_{g1}^2$, $g_{1t} = 1$, $v_{g1}^2 = n$, and $v_1^2 = \sum_{t=1}^n h_t^2$.

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

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

computed from the sample y_j, z_j, $j = 1, ..., n$.

Consistency. We first establish the consistency of the OLS estimator $\widehat{\beta}$.

Theorem 2.1. *Suppose that $(y_1, ..., y_n)$ is a sample from the regression model (1) and Assumptions 2.1, 2.2 and 2.3 are satisfied. Then, the OLS estimator $\widehat{\beta}$ is consistent, i.e.*

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

This result implies that the k-th component $\widehat{\beta}_k$ of the OLS estimator is v_k-consistent, that is, $\widehat{\beta}_k - \beta_k = O_p(v_k^{-1})$. The convergence rate v_k may deviate from the conventional $\sqrt{n}$ rate and may differ across components. From the definition of v_k and v_{gk}, it follows that

$$\text{if } g_{kt}, h_t \geq c > 0 \text{ for all } t, n, \text{ then } v_k, v_{gk} \geq c\sqrt{n}. \tag{10}$$

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

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

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

Assumption 2.4 is not required when ε_t is i.i.d. Together, Assumptions 2.3 and 2.4(ii) exclude cases in which the mean process μ_t or a few extreme observations of z_t or u_t, dominate the estimation of the regression parameter. Overall, these assumptions are mild. They accommodate both deterministic and stochastic means μ_t and scale factors h_t, g_t that may change abruptly over time unlike other theoretically rigorous treatments which restrict structural change to be deterministic and smooth. This flexibility makes the framework particularly suitable for modelling financial data, as it allows for volatility jumps, commonly observed in empirical finance (see, e.g., Eraker et al. (2003)). In modern macroeconomic VAR models, the scale factor h_t in the uncorrelated noise representation $u_t = h_t\varepsilon_t$ is typically assumed to be stochastic (see, e.g., Chan et al. (2024), Carriero et al. (2024)), which our framework naturally encompasses.

Lemma 2.1 below shows that Assumptions 2.3 and 2.4(ii) holds for regressors z_t and noises u_t with bounded $4+\delta$ moments satisfying (10). The following example provides additional sufficient conditions.

Example 2.1. Assumptions 2.3 and 2.4(ii) are satisfied by regressors z_t and noises u_t whose scale factors h_t, g_t and means μ_t satisfy $0 < c \leq h_t, g_{kt} \leq C$, $||\mu_t|| \leq C$, where $0 < c$, $C < \infty$ do not depend on t, n or $k = 1, ..., p$.

When $0 < c \leq h_t \leq C$, $||\mu_t|| \leq C$ for all t, n, Assumptions 2.3 and 2.4(ii) hold for scale factors g_{kt} satisfying

$$\frac{\min_{t=1,...,n} g_{kt}^2}{\sum_{t=1}^n g_{kt}^2} = o_p(1), \quad k = 1, ..., p.$$

This condition is, for example, satisfied when g_{kt} follows a unit root process defined by $g_{kt} = \sum_{j=1}^n \xi_j$, where $\{\xi_j\}$ is a sequence of i.i.d $(0, \sigma^2)$ random variables with finite moments of order $\theta > 2$. The idea of modelling parameters as unit root processes was discussed, for example, in Nyblom (1989b).

We now describe the infeasible standard errors $\sqrt{\omega_{kk}}$ using the notation:

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

where ω_{jk} denotes the (j,k)-th element of the matrix Ω_n. The infeasible standard error of $\widehat{\beta}_k$ is defined as $\sqrt{\omega_{kk}}$, i.e., the square root of the corresponding diagonal element of Ω_n.

The generality of our regression setting limits the multivariate asymptotic theory that can be established for $\widehat{\beta}_t$. While a full joint distribution of $\widehat{\beta}_t$ is not available, we can derive asymptotic normality for linear combinations $a'\widehat{\beta}$ and then construct feasible inference for individual component β_k.

Existing literature typically imposes stronger assumptions on regressors and errors such as mixing regressors (White, 2014, Theorem 3.78), locally stationary regressors in (Zhang and Wu, 2012, eq. (2.3)), or near-epoch dependent errors in (Hall et al., 2012, Assumption 8).

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

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

In particular, for $k = 1, ..., p$, the t-statistic for β_k satisfies

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

Property (13) is difficult to implement in practice because it requires estimation of the unknown matrices D, Ω_n, except in the special case $a' = (0, ..., 1, ...0)$ with only the k-th element nonzero. In this case, (13) reduces to (14), and the infeasible standard error $\sqrt{\omega_{kk}}$ can be consistently estimated by

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

The feasible standard error $\sqrt{\widehat{\omega}_{kk}}$ is the square root of the diagonal element $\widehat{\omega}_{kk}$ of $\widehat{\Omega}_n$.

Corollary 2.1. *Under the assumptions 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 straightforward computation of standard errors and the construction of confidence intervals for β_k in the extended regression framework. Notably, the estimator $\widehat{\Omega}_n$ coincides with the heteroskedasticity-consistent standard error estimator of White (1980).

Remark 2.1. The consistency rate $v_k = (\sum_{k=1}^n g_{kt}^2 h_t^2)^{1/2}$ for the parameter β_k may take the form $v_k \sim cn^\alpha$ for any $\alpha > 0$, ranging from super-slow $(0 < \alpha < 1)$ to super-fast $(\alpha > 1)$ convergence. To illustrate this, consider the regression model

$$\begin{aligned} y_t &= \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + u_t, \quad u_t = h_t \varepsilon_t \text{ with } h_t = 1, \\ & \quad z_{kt} = g_{kt} \eta_{kt}, \quad g_{kt} = t^{(\alpha_k - 1)/2} \text{ for } k = 2, 3, \end{aligned}$$

where $\alpha_2 > 1$, $0 < \alpha_3 < 1$, and $\{\eta_{2t}\}$, $\{\eta_{3t}\}$, $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}(0, 1)$. Then $v_1 = \sqrt{n}$ and $v_k \sim \alpha_k^{-1/2} n^{\alpha_k/2}$ for $k = 2, 3$, producing different convergence rates across parameters. Even in this simple case, the usual multivariate asymptotic normality for $\sqrt{n}(\widehat{\beta} - \beta)$ does not hold.

Corollary 2.1 allows us to establish the asymptotic power and consistency of a test for testing the hypothesis

$$H_0 : \beta_k = \beta_k^0, \quad \text{vs.} \quad H_1 : \beta_k \neq \beta_k^0,$$

i.e., whether the k-th element of the regression parameter $\beta = (\beta_1, \ldots, \beta_p)'$ is equal to a specific value β_k^0.

Corollary 2.2. *Suppose that $\beta_k^0 \neq \beta_k$. Then, under the assumptions of Corollary 2.1,*

$$t = \frac{\widehat{\beta}_k - \beta_k^0}{\sqrt{\widehat{\omega}_{kk}}} \asymp_p v_k \to_p \infty. \tag{17}$$

We conclude this section with a lemma that provides simple sufficient moment-type conditions for the validity of Assumptions 2.3 and 2.4(ii). In particular, condition (10) implies (19).

Lemma 2.1. *Suppose that for* $k = 1, ..., p$,

$$\begin{aligned} Ez_{kt}^4 &\leq c, \quad E|u_t|^{4+\delta} \leq c \ \textit{ for some } \delta > 0, &(18)\\ n/v_k^2 &= O_p(1), \quad n/v_{gk}^2 = O_p(1), &(19)\end{aligned}$$

where $c < \infty$ *does not depend on* t, n. *Then Assumptions 2.3 and 2.4(ii) hold.* *In particular, (19) is satisfied if* $\min_{t=1,...,n} h_t^{-1} = O_p(1)$, $\min_{t=1,...,n} g_{kt}^{-1} = O_p(1)$.

The regular estimator of standard errors in OLS regression estimation is given by

$$\widehat{\Omega}_n^{(st)} = S_{zz}^{-1} \widehat{\sigma}_u^2, \quad \widehat{\sigma}_u^2 = n^{-1} \sum_{j=1}^n \widehat{u}_j^2. \tag{20}$$

Unlike the robust standard errors $\sqrt{\widehat{\omega}_{kk}}$, these conventional standard errors may produce coverage distortions, particularly when heteroskedasticity or heterogeneity in g_t, h_t, or μ_t is present, see Section 6. This underscores the robustness and strong empirical performance of the normal approximation in (16).

In this section, we have provided a rigorous validation of the asymptotic normality of feasible t-statistics for the components of the OLS estimator in linear regression models with general heterogeneity. The assumptions imposed are mild yet flexible, allowing a wide class of (possibly nonstationary) regressors and noise processes beyond those typically considered in the existing literature. Some conditions on scale factors are analogous to the Lindeberg condition and remain necessary. Our framework complements, rather than replaces, prior approaches; for instance, near-unit-root regressors Georgiev et al. (2018) require a distinct theoretical treatment. Although bootstrap methods, see, e.g., Hall et al. (2012), Boldea et al. (2019), are widely applied in regression analysis, they may not extend to the heterogeneous structures considered here. By contrast, we demonstrate that the heteroskedasticity-consistent standard errors of White (1980) remain applicable and computationally straightforward.

In this paper we focus on the regression model (1), where the regression noise u_t in (1) is uncorrelated. Extending the asymptotic theory to account for dependence in u_t is a natural next step and is currently under consideration.

Detailed proofs of all results are provided in the Online Supplement.

3 Time-varying OLS estimation in extended regression space

This section demonstrates further advantages of the theory of regression estimation with a fixed parameter, developed in Section 2. Thanks to the flexible setting, estimation of time-varying parameters naturally follows from our theory for fixed-parameter regression in the extended space, along with bounding of some negligible terms.

In the previous section, we discussed the estimation of the regression model (1), $y_j = \beta' z_j + u_j$, with a fixed parameter β. We now extend the model by allowing the regression parameter to vary over time. Specifically, we consider the model

$$y_j = \beta_j' z_j + u_j, \quad j = 1, ..., n, \tag{21}$$

where the regressors z_j and the regression noise u_j, as defined in (3) and (2), remain unchanged. That is, they belong to the same regression space as in Section 2.

The primary objective is to develop a point-wise estimation procedure for the path $\beta_1, ..., \beta_n$ of the time-varying parameter β_j in model (21), while preserving the same regression space introduced in Section 2.

The literature on estimation of time-varying regression parameters β_j is extensive. It primarily focuses on estimation and testing for parameter stability under relatively strong assumptions on the regressors and regression noise. For instance, regressors are assumed to be locally stationary in (Vogt (2012), model (3)), stationary and strongly mixing in (Fu et al. (2023), Assumption A.1) and strictly stationary in (Hu et al. (2024), Assumption P(d)). It is clear that the class of regressors considered in our setting is broader, and they may be neither mixing nor stationary.

The objective of this section is to describe the extended regression space of regressors z_t and disturbances u_t that ensures the asymptotic normality of the feasible t-statistic estimating the components of the time-varying parameter β_t. We show that, as long as the regressors and the disturbance follow the structure $z_t = \mu_t + I_{gt}\eta_t$ and $u_t = h_t \varepsilon_t$, the class of admissible means μ_t and scale factors g_t, h_t is very broad and characterized by weak restrictions that may not require empirical verification.

Further extensions of the regression space are possible. For example, the weakly exogenous component η_t of the regressors z_t in our paper is assumed to be a short-memory process. In contrast, Hu et al. (2024) demonstrate that estimation of the time-varying parameter β_t also permits weakly exogenous, strictly stationary regressors z_t that exhibit long-memory behavior.

While most assumptions on the regressors z_j and regression noise u_j remain unchanged from Section 2, the estimator requires some modifications. Under an additional smoothness assumption on $\{\beta_j\}$, the time-varying OLS estimator $\widehat{\beta}_t$ of parameter β_t at time t is the standard OLS estimator for a fixed regression parameter, obtained by regressing $\widetilde{y}_j = b_{n,tj}^{1/2} y_j$ on $\widetilde{z}_j = b_{n,tj}^{1/2} z_j$:

$$\widehat{\beta}_t = \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{z}_j'\big)^{-1} \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{y}_j\big) = \big(\sum_{j=1}^n b_{n,tj} z_j z_j'\big)^{-1} \big(\sum_{j=1}^n b_{n,tj} z_j y_j\big). \tag{22}$$

The weights $b_{n,tj}$ are generated as follows:

$$b_{n,tj} = K(\frac{|t-j|}{H}),\ t,j = 1,...,n, \tag{23}$$

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

$$\begin{aligned} K(x) &\geq a_0 > 0,\ \ 0 \leq x \leq \delta, \\ K(x) &\leq Cx^{-\theta},\ \ x > \delta. \end{aligned} \tag{24}$$

For example, (24) is satisfied by functions $K(x) = I(x \in [0,1])$ and $K(x) = p(x)$ where $p(x)$ is the probability density function of the standard normal distribution.

We impose a smoothness assumption on the time-varying parameter β_j, which may be either deterministic or stochastic.

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

$$E||\beta_t - \beta_j||^2 \leq c\big(\frac{|t-j|}{n}\big)^{2\gamma}, \tag{25}$$

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

Next, we briefly outline how our asymptotic theory for the time-varying robust estimator builds on the results from Section 2 on fixed-parameter regression estimation and the smoothness assumption (25). To demonstrate this, we introduce the following regression model with a fixed parameter $\beta = \beta_t$:

$$y_j^* \ = \ \beta'\widetilde{z}_j + \widetilde{u}_j,\ \ \widetilde{u}_j = b_{n,tj}^{1/2} u_j, \quad j = 1,...,n. \tag{26}$$

Notice that the OLS estimator $\widehat{\beta}$ of the fixed parameter β satisfies:

$$\widehat{\beta} \ = \ \big(\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'\big)^{-1}\big(\sum_{j=1}^n \widetilde{z}_j y_j^*\big) = \beta + \big(\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'\big)^{-1}\big(\sum_{j=1}^n \widetilde{z}_j\widetilde{u}_j\big). \tag{27}$$

Since $\widetilde{y}_j = y_j^* + (\beta_j - \beta_t)'\widetilde{z}_j$, the time-varying estimator $\widehat{\beta}_t$ given in (22) satisfies:

$$\begin{aligned} \widehat{\beta}_t - \beta_t \ &= \ \big(\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'\big)^{-1}\big(\sum_{j=1}^n \widetilde{z}_j\{y_j^* + (\beta_j - \beta_t)'\widetilde{z}_j\}\big) - \beta_t \\ &= \ \widehat{\beta} - \beta + R_t, \quad R_t = \big(\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'\big)^{-1}\big(\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'(\beta_j - \beta_t)\big). \end{aligned} \tag{28}$$

Notice that $\widehat{\beta} - \beta$ in (28) does not depend on β. Additionally, the regression space in estimation of the fixed parameter in Section 2 permits rescaling, so premultiplying by the kernel weights $b_{n,tj}^{1/2}$ does not change the structure of regressors $\widetilde{z}_j = (\widetilde{z}_{1j}, ..., \widetilde{z}_{pj})'$ and $\widetilde{u}_j$: they still satisfy the settings (3) and (2). Consequently, the model (26) is covered by the regression model (1) with a fixed parameter, and the asymptotic results for $\widehat{\beta} - \beta$ follow from Section 2. The main technical task in this section is to show that the remainder term R_t in (28) is negligible, which follows from the smoothness assumption (25).

The regressors z_j and regression noise u_j belong to the same regression space as defined in as in Section 2. While the assumptions on the stationary process $\{\eta_j\}$ and the martingale difference noise $\{\varepsilon_j\}$ remain unchanged, for simplicity, we replace the previous conditions on the scale factors g_j, h_j and the means μ_j with simple sufficient assumptions similar to those used in Lemma 2.1. As before, the scale factors $\{h_j, g_j, \mu_j\}$ can be deterministic or stochastic, may vary with n, and are independent of $\{\eta_j, \varepsilon_j\}$.

Denote

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

Assumption 3.2. *z_t and u_t are such that, for $k = 1, ..., p$,*

$$\begin{aligned} Ez_{kt}^4 &\leq c, \quad E|u_t|^{4+\delta} \leq c \text{ for some } \delta > 0, & (29)\\ H/v_{kt}^2 &= O_p(1), \qquad H/v_{gk,t}^2 = O_p(1), & (30)\end{aligned}$$

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

It is straightforward to show that (30) is valid if $g_{kt}, h_t \geq c > 0$ for all t, n.

To describe the infeasible standard errors $\sqrt{\omega_{kk,t}}$, we use:

$$\begin{aligned} &S_{zz,t} = \textstyle\sum_{j=1}^n b_{n,tj} z_j z_j', \quad S_{zzuu,t} = \sum_{j=1}^n b_{n,tj}^2 z_j z_j' u_j^2,\\ &\Omega_{nt} = E[S_{zz,t}|\mathcal{F}_n^*]^{-1} E[S_{zzuu,t}|\mathcal{F}_n^*] E[S_{zz,t}|\mathcal{F}_n^*]^{-1} = (\omega_{jk,t}),\end{aligned}$$

where $\omega_{jk,t}$ denotes the (j,k)-th element of the matrix Ω_{nt}. The infeasible standard error $\sqrt{\omega_{kk,t}}$ is defined by the diagonal element $\omega_{kk,t}$ of the matrix Ω_{nt}.

The next theorem establishes the consistency rate and asymptotic normality property for the components of the time-varying OLS estimator $\widehat{\beta}_t = (\widehat{\beta}_{1t}, ..., \widehat{\beta}_{pt})'$, and allows for arrays of integers $t = t_n \in [1, ..., n]$, which may depend on n.

Theorem 3.1. *Suppose that $(y_1, ..., y_n)$ is a sample from a regression model (21). Assume that Assumptions 2.1, 2.2, 2.4(i), 3.1 and 3.2 hold. Then, for $1 \leq t = t_n \leq n$ and $k = 1, ..., p$:*

$$\widehat{\beta}_{kt} - \beta_{kt} = O_p\big(H^{-1/2} + (H/n)^\gamma\big), \tag{31}$$

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\omega_{kk,t}}} \to_d \mathcal{N}(0,1) \quad \text{if } H = o(n^{2\gamma/(2\gamma+1)}), \tag{32}$$

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

The consistency rate in (31) is determined by the bandwidth parameter H and the smoothness parameter $\gamma \in (0,1)$ in (25). The condition $H = o(n^{2\gamma/(2\gamma+1)})$ ensures that in (32) the bias term remains negligible.

As in the fixed-parameter case, for (z_j, u_j) from the extended regression space, the asymptotic normality can be established in point-wise estimation for each individual component $\widehat{\beta}_{kt}$ of $\widehat{\beta}_t$.

The unknown standard error $\sqrt{\omega_{kk,t}}$ can be consistently estimated by:

$$\widehat{\Omega}_{nt} = S_{zz,t}^{-1} S_{zz\widehat{u}\widehat{u},t} S_{zz,t}^{-1} = (\widehat{\omega}_{jk,t}), \quad \widehat{u}_j = y_j - \widehat{\beta}_j' z_j. \tag{33}$$

The feasible standard error $\sqrt{\widehat{\omega}_{kk,t}}$ is defined by the diagonal element $\widehat{\omega}_{kk,t}$ of $\widehat{\Omega}_{nt}$.

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

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} \to_d \mathcal{N}(0,1), \quad \frac{\widehat{\omega}_{kk,t}}{\omega_{kk,t}} = 1 + o_p(1). \tag{34}$$

Corollary 3.1 allows us to establish the asymptotic power of the test of the hypothesis

$$H_0 : \beta_{kt} = \beta_{kt}^0, \quad \text{vs.} \quad H_1 : \beta_{kt} \neq \beta_{kt}^0,$$

based on the t-statistics $(\widehat{\beta}_{kt} - \beta_{kt}^0)/\sqrt{\widehat{\omega}_{kk,t}}$.

Corollary 3.2. *Suppose that $|\beta_{kt}^0 - \beta_{kt}| \geq a > 0$ for $t = t_n \in [1, ..., n]$ as $n \to \infty$. Then, under assumption of Corollary 3.1,*

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}^0}{\sqrt{\widehat{\omega}_{kk,t}}} \asymp_p H^{1/2} \to_p \infty. \tag{35}$$

The estimator $\widehat{\Omega}_{nt}$ used to obtain robust standard errors in (33) is a time-varying version of heteroskedasticity-consistent estimator of standard errors by White (1980). Simulation results confirm that it does not produce coverage distortions in the estimation of β_t under the settings considered in this section.

In conclusion, we provide examples of smoothly varying deterministic and stochastic parameters β_t that satisfy Assumption 3.1.

Example 3.1. *A standard 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 $|g(x) - g(y)| \leq C|x - y|$. Such β_t satisfies (25) 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 (25) with* $\gamma = 1/2$, *that is for* $t > s$,

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

The above results are equipped with thorough and mathematically rigorous proofs, which can be found in the Online Supplement.

The key new features in the estimation of time-varying parameter β_t are similar to those highlighted in the estimation of the fixed parameter in Section 2. Although the computation is straightforward, establishing the validity of the robust standard errors $\sqrt{\widehat{\omega}_{kk,t}}$ in the extended regression space of (z_t, u_t) is challenging because the scale factors h_t, g_t, μ_t in model (21) are unknown and potentially random, and highly general, while the asymptotic behaviour of the $\omega_{kk,t}$ may not be well-defined. The asymptotic normality of a single component of the estimator can still be established, even though a full multivariate asymptotic theory is not available. Unlike most existing literature, β_t is permitted to evolve as a smoothly varying stochastic process.

4 Regression with missing data

In the previous sections, we showed that the extended regression space enables the estimation of both fixed and time-varying regression parameters. It offers several theoretical advantage, in particular, the ability to estimate regression models in the presence of missing data. Given the importance in empirical regression analysis in situations where some observations y_t or regressors z_t are missing, see, e.g., Enders (2022), we now present new and somewhat unexpected results on regression estimation with missing data. We show that the foundational assumptions underlying the constriction of regression space also allow us to accommodate an a broad range of missing data patterns.

In this section we suppose that instead of the full sample $(y_1, z_1), ..., (y_n, z_n)$, we observe a subsample

$$(y_{k_1}, z_{k_1}), ..., (y_{k_N}, z_{k_N}), \quad N \le n, \tag{36}$$

of dependent variable y_t and regressor z_t. Our primary interest is to estimate both fixed and time-varying regression parameters using the subsample (36).

To that end, we represent the observed data as partially observed sample

$$(\widetilde{y}_j, \widetilde{z}_j) = (\tau_j y_j,\, \tau_j z_j), \quad j = 1, ..., n \tag{37}$$

where τ_j is missing-data indicator. In (36) it is defined as

$$\tau_j = \begin{cases} 1 & \text{for} \quad j = k_1, k_2, \dots, k_N, \quad \text{where} \quad k_1 < k_2 < \dots < k_N \le n, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

We set $\tau_j = 1$ if both y_j and z_j are observed, otherwise $\tau_j = 0$. Throughout this section, τ_j is treated as a sequence of random or deterministic variables, allowing for regularly missing, block-wise missing, or randomly missing data patterns.

In order for the theoretical results of the previous section to apply, we impose the following assumptions on the missing data indicator τ_t, the regressors $z_{kt} = \mu_{kt} + g_{kt}\eta_{kt}$ in (3) and the regression noise $u_t = h_t\varepsilon_t$ in (2).

Assumption 4.1. *The missing-data indicator $\{\tau_t\}$ is assumed to be independent of $\{\varepsilon_t, \eta_t\}$ in (2) and (3).*

Assumption 4.2. (i) *$Ez_{kt}^4 \le c$ and $E|u_t|^{4+\delta} \le c$ for some $\delta > 0$, where $c > 0$ does not depend on k, t, n.*

(ii) *$g_{kt} \ge c > 0$ and $h_t \ge c > 0$, where c does not depend on k, t, n.*

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

Estimation of a fixed parameter. Suppose that $y_t = \beta' z_t + u_t$ follows the regression model (1) with a fixed parameter β as in Section 2. Our primary interest is to estimate the parameter β using subsample (36). In view of (1), we can write the partially observed regression model as

$$\begin{aligned} \widetilde{y}_t &= \tau_t y_t = \tau_t(\beta' z_t + u_t), \\ \widetilde{y}_t &= \beta'\widetilde{z}_t + \widetilde{u}_t, \quad \widetilde{u}_t = \tau_t u_t = \{\tau_t h_t\}\varepsilon_t. \end{aligned} \tag{39}$$

In (39), the regressors $\widetilde{z}_t$ and the noise $\widetilde{u}_t$ can be represented as

$$\begin{aligned} \widetilde{z}_{kt} &= \widetilde{\mu}_{kt} + \widetilde{g}_{kt}\eta_{kt}, \quad \widetilde{\mu}_{kt} = \tau_t\mu_{kt}, \quad \widetilde{g}_{kt} = \tau_t g_{kt}, \\ \widetilde{u}_t &= \widetilde{h}_t\varepsilon_t, \quad \widetilde{h}_t = \tau_t h_t. \end{aligned} \tag{40}$$

They belong to the regression space described in (2) and (3). Therefore, parameter β and the correspondent standard errors in model (39) can be estimated using the OLS estimator $\widehat{\beta}$ and $\widehat{\omega}_{kk}$:

$$\begin{aligned} \widehat{\beta} &= \Big(\sum_{t=1}^{n} \widetilde{z}_t\widetilde{z}_t'\Big)^{-1}\Big(\sum_{t=1}^{n} \widetilde{z}_t\widetilde{y}_t\Big), \quad \widehat{\Omega}_n = S_{\widetilde{z}\widetilde{z}}^{-1} S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} S_{\widetilde{z}\widetilde{z}}^{-1} = (\widehat{\omega}_{jk}), \\ & S_{\widetilde{z}\widetilde{z}} = \textstyle\sum_{t=1}^{n} \widetilde{z}_t\widetilde{z}_t', \quad S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} = \sum_{t=1}^{n} \widetilde{z}_t\widetilde{z}_t'\widehat{u}_t^2, \quad \widehat{u}_t = \widetilde{y}_t - \widehat{\beta}'\widetilde{z}_t. \end{aligned} \tag{41}$$

Theorem 4.1. *The OLS estimator $\widehat{\beta}$ of parameter β in regression model (40) with missing data has the following asymptotic properties. If Assumptions 4.1 and 4.2 hold and $n/N = O_p(1)$, then, for $k = 1, ..., p$, as $n \to \infty$,*

$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} \to_d \mathcal{N}(0,1), \qquad \sqrt{\widehat{\omega}_{kk}} \asymp_p n^{-1/2}. \tag{42}$$

Remark 4.1. Theorem 4.1 shows that ignoring missing data does not affect the estimation of the fixed parameter. That is, the researcher can compute the estimators $\widehat{\beta}$ and $\sqrt{\widehat{\omega}_{kk}}$ directly using subsample y_{k_j}, z_{k_j}, $j = 1, ..., N$:

$$\begin{aligned}
\widehat{\beta} \ &= \ \Big(\sum_{j=1}^{N} z_{k_j} z'_{k_j}\Big)^{-1}\Big(\sum_{j=1}^{N} z_{k_j} y_{k_j}\Big), \quad \widehat{\Omega}_n = S_{*,zz}^{-1} S_{*,zz\widehat{u}\widehat{u}} S_{*,zz}^{-1} = (\widehat{\omega}_{jk}), \\
& S_{*,zz} = \textstyle\sum_{j=1}^{N} z_{k_j} z'_{k_j}, \quad S_{*,zz\widehat{u}\widehat{u}} = \sum_{j=1}^{N} z_{k_j} z'_{k_j} \widehat{u}^2_{k_j}, \quad \widehat{u}_{k_j} = y_{k_j} - \widehat{\beta}' z_{k_j}.
\end{aligned}$$

Estimation of a time-varying parameter. Assume now that $y_t = \beta'_t z_t + u_t$ follows the regression model (21) with time-varying parameter β_t, where regressors z_t and regression noise u_t are as in (3) and (2). We are interested in estimating the parameter β_t in the presence of missing data using the subsample (36). Similarly to (39), we base the estimation on the partially observed regression model with a time-varying parameter,

$$\widetilde{y}_j = \beta'_j \widetilde{z}_j + \widetilde{u}_j, \quad j = 1, ..., n, \tag{43}$$

where regressors $\widetilde{z}_j$ and the noise $\widetilde{u}_j$ are defined as in (40). They belong to the regression space described by (2), and (3) and thus results of Section 3 on the estimation of time-varying parameter β_j apply.

We show in the following theorem that under Assumptions 4.1 and 4.2, parameter β_t and standard errors can be estimated point-wise at each time $t = 1, ..., n$ provided that the missing data pattern satisfies the following condition:

$$H/N_t = O_p(1), \quad N_t = \sum_{j=1}^{n} \tau_j b_{n,tj}. \tag{44}$$

This condition holds, for example, if $\tau_j = 1$ for $|j - t| \le \epsilon H$ for some $\epsilon > 0$.

The estimator $\widehat{\beta}_t$ and the estimator of the robust standard errors $\widehat{\omega}_{kk,t}$ given in (22) and (33) are defined as

$$\begin{aligned}
\widehat{\beta}_t \ &= \ \Big(\sum_{j=1}^{n} b_{n,tj} \widetilde{z}_j \widetilde{z}'_j\Big)^{-1}\Big(\sum_{j=1}^{n} b_{n,tj} \widetilde{z}_j \widetilde{y}_j\Big), \\
\widehat{\Omega}_{nt} \ &= \ S^{-1}_{\widetilde{z}\widetilde{z},t} S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u},t} S^{-1}_{\widetilde{z}\widetilde{z},t} = (\widehat{\omega}_{jk,t}), \quad \widehat{u}_j = \widetilde{y}_j - \widehat{\beta}'_j \widetilde{z}_j.
\end{aligned} \tag{45}$$

Theorem 4.2. *The OLS estimator $\widehat{\beta}_t$ of the time-varying parameter β_t in regression model (43) with missing data has the following properties. Assume that $1 \le t = t_n \le n$, Assumptions 4.1, 3.1 and 4.2 are satisfied and that the condition $H/N_t = O_p(1)$ holds. Then, for $k = 1, ..., p$, as $n \to \infty$,*

$$\widehat{\beta}_{kt} - \beta_{kt} = O_p\big(H^{-1/2} + (H/n)^{\gamma}\big), \tag{46}$$

$$\frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} \to_d \mathcal{N}(0,1) \quad \text{if } H = o(n^{2\gamma/(2\gamma+1)}), \tag{47}$$

$$\widehat{\omega}_{kk,t} \asymp_p H^{-1}. \tag{48}$$

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

In this section we focus on another practical application of our regression framework developed in Section 2. We show that it covers the estimation of parameters of a stationary AR(p) model driven by a stationary martingale difference noise ε_t:

$$y_t = \phi_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t, \tag{49}$$

where parameters $\phi_0, ..., \phi_p$ are such that the model (49) has a stationary solution. Xu and Phillips (2008) developed estimation theory for AR(p) model $y_t = \phi_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t$, when $u_t = h_t \varepsilon_t$ where h_t is smoothly varying deterministic sequence and a m.d. sequence ε_t has property $E[\varepsilon_t^2 | \mathcal{F}_{t-1}] = 1$ a.s. Giraitis et al. (2018) were among the first to analyze the distortions of standard errors caused by m.d. noise in estimation of ARMA models. This paper shows that the variance of the parameter vector ϕ converges to a well-defined limit; however, its complex structure complicates the estimation of the limiting variance and the corresponding standard errors in empirical applications. They restricted the estimation of standard errors to AR(1) and MA(1) models. In the case of AR(p) model, using our method we are able to estimate standard errors for any p without relying on asymptotic approximations which is the main novelty and contribution of this section. Notice that the model (49) can be written as a special case of the regression model (1),

$$y_t = \beta' z_t + u_t, \quad u_t = \varepsilon_t. \tag{50}$$

Here, the parameter $\beta = (\beta_1, ..., \beta_{p+1})' = (\phi_0, ..., \phi_p)'$ is fixed, and the regressors $z_t = (z_{1t}, z_{2t}, ..., z_{p+1,t})' = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})'$ are stationary random variables. It is straightforward to verify that the regressors

$$z_{kt} = \mu_{kt} + g_{kt}\eta_{kt}, \quad \mu_{kt} = E[y_{t-k}] = Ey_1, \quad g_{kt} = 1, \quad \eta_{kt} = y_{t-k} - E[y_{t-k}]$$

for $k = 2, ..., p+1$ satisfy the regression assumption (3). In the theorem below, we assume that the standard stationarity conditions on parameters of the AR(p) model (49) are satisfied, see e.g. Theorem 3.1.1 in Brockwell and Davis (1991), which ensure the existence of a stationary solution

$$y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \text{where } \sum_{j=0}^{\infty} |a_j| < \infty, \ \mu = E y_t. \tag{51}$$

We assume that ε_t satisfies Assumption 2.1 and $\eta_t = (y_{t-1}, y_{t-2}, ..., y_{t-p})'$ satisfy Assumptions 2.2 and 2.4(i). These assumptions impose only mild restrictions on the m.d. noise ε_t, and their validity can be verified for typical examples of uncorrelated m.d. noise, such as ARCH-type processes.

The OLS estimator $\widehat{\beta}$ of β in regression model (50) is defined as in (8) and $\widehat{\omega}_{kk}$ as in (15).

Theorem 5.1. *Suppose that AR(p) model (49) with m.d. noise ε_t has a stationary solution as in (51), that $E\varepsilon_t^8 < \infty$ and that (ε_t, η_t) satisfy Assumptions 2.1, 2.2, and 2.4(i). Then the OLS estimator $\widehat{\beta}$ of parameter β in regression model (50) has the following properties: for $k = 1, ..., p+1$, as $n \to \infty$,*

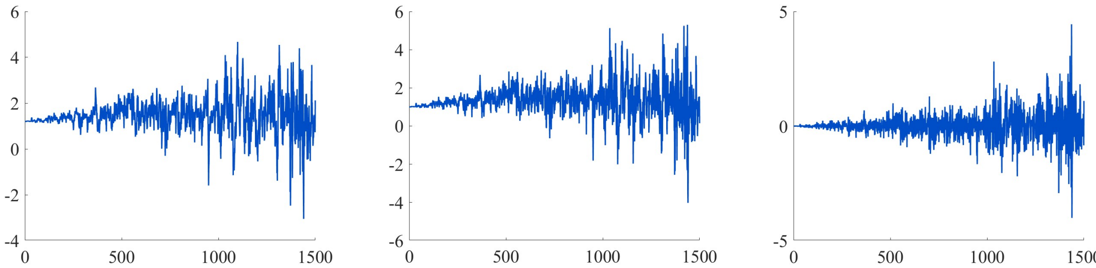
$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} \to_d \mathcal{N}(0,1), \qquad \sqrt{\widehat{\omega}_{kk}} \asymp_p n^{-1/2}. \tag{52}$$

The Monte Carlo results presented in Section 6.4 demonstrate that the robust OLS estimation produces correct 95% confidence intervals for β_k, whereas the standard OLS method exhibits coverage distortions, when the noise ε_t is not i.i.d. This finding indicates that the robust OLS estimator has a broader range of applicability than merely addressing heteroscedasticity, and that it can also be effectively used in regression settings not covered by the standard OLS estimation and inference theory.

It is worth noting that the papers by Doukhan and Wintenberger (2008), Bardet and Wintenberger (2009) and Karmakar et al. (2022) provide advanced theoretical results on the modelling and estimation of general nonlinear time-varying time series models; however, they address the linear AR(p) model (49) only in the trivial case of an i.i.d. noise ε_t.

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 the 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 and β_t, while application of the standard confidence intervals leads to severe distortion of coverage rates.



(a) y_t (b) z_{2t} (c) u_t

Figure 1: Plots of y_t, z_{2t}, u_t in Model 6.1.

6.1 Estimation of a fixed parameter

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

$$y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + u_t, \quad u_t = h_t \varepsilon_t, \quad \beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.4, 0.3)'. \tag{53}$$

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). We also include a more complex example in the online supplement.

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

$$\begin{aligned} z_{kt} &= \mu_{kt} + g_{kt}\eta_{kt}, \ \ k = 2, 3, \\ &\mu_{kt} = 0.5\sin(\pi t/n) + 1, \quad \eta_{kt} = 0.5\eta_{k,t-1} + \xi_{kt}, \end{aligned} \tag{54}$$

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 a GARCH$(1,1)$ process

$$\varepsilon_t = \sigma_t e_t, \quad \sigma_t^2 = 1 + 0.7\sigma_{t-1}^2 + 0.2\varepsilon_{t-1}^2, \quad e_t \sim i.i.d.\,\mathcal{N}(0,1). \tag{55}$$

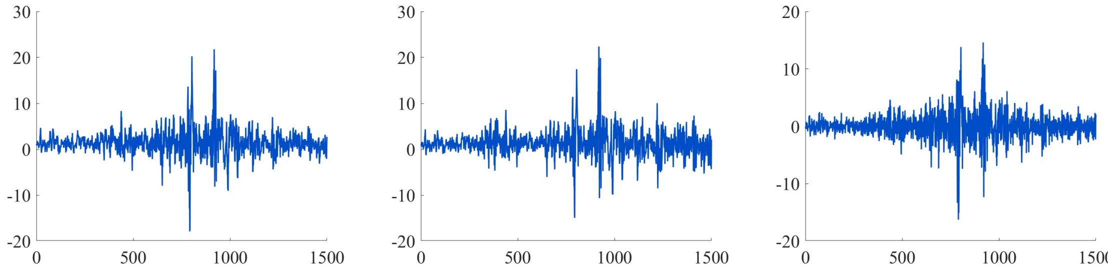
Model 6.1. *y_t follows (53) with deterministic scale factors. We set: $h_t = 0.3(t/n)$ and $g_{2,t} = g_{3,t} = 0.4(t/n)$.*

Model 6.2. *y_t follows (53) with stochastic scale factors. We set*

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

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. Examples of plots of the simulated dependent variable, regressor and regression noise are shown in Figure 1 and 2 (z_{2t} and z_{3t} have similar patterns). To verify the validity of the asymptotic normal



(a) y_t (b) z_{2t} (c) u_t

Figure 2: Plots of y_t, z_{2t}, u_t in Model 6.2.

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

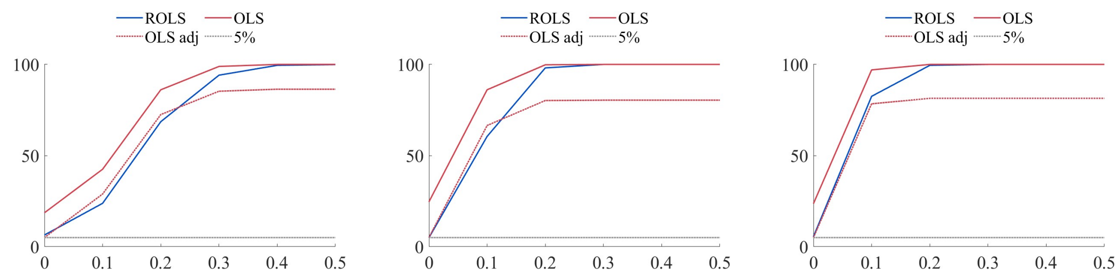
approximation of Corollary 2.1 in finite samples, we compute empirical coverage rates (CP) for 95% confidence intervals used in robust OLS estimation, for parameter β. For comparison, we compute the coverage rates CP$_{st}$ for standard confidence intervals based on the standard errors (20) used in standard 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 differ because the variances (and standard errors) in their normal approximations 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 the 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 the coverage rate CP for robust confidence intervals is close to the nominal 95%, whereas the 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



(a) $n = 200$ (b) $n = 800$ (c) $n = 1500$

Figure 3: Size, power, and adjusted power (%) for test $H_0 : \beta_3 = 0$ in Model 6.1: $\beta_3 = 0, \cdots, 0.5$, $n = 200$, 800, 1500.

To assess power, we vary β_3 in Model 6.1 from 0 to 0.5 and record how often the test rejects $H_0 : \beta_3 = 0$. Figure 3 reports results for ROLS and OLS at sample sizes $n = 200$, 800, 1500. When $\beta_3 = 0$, ROLS achieves a good size close to the nominal 5%, while the size based on OLS results starts around 20% and remains heavily oversized even as n increases. For $\beta_3 \neq 0$, power rises monotonically with β_3 for both methods. In Figure 3, the blue solid lines represent power based on ROLS, and the red solid lines correspond to standard OLS. Considering the OLS estimation has large size distortion, we compute its adjusted power, shown by the red dotted lines. With small sample size $n = 200$, OLS appears more powerful for $\beta_3 \leq 0.2$, whereas ROLS catches up and achives good power when $\beta_3 \geq 0.3$. For $n = 800$ and 1500, both methods already achieve good power around $\beta_3 = 0.2$. Overall, ROLS provides reliable size and competitive power across different sample sizes. Similary results are observed for Model 6.2.

6.2 Estimation of a time-varying parameter

In this section we examine the validity of the normal approximation for the estimator $\widehat{\beta}_t$, (22), of time-varying parameter β_t, as established in Corollary 3.1 of Section 3. We replace the fixed regression parameter β in the model (53) by a time-varying parameter $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$:

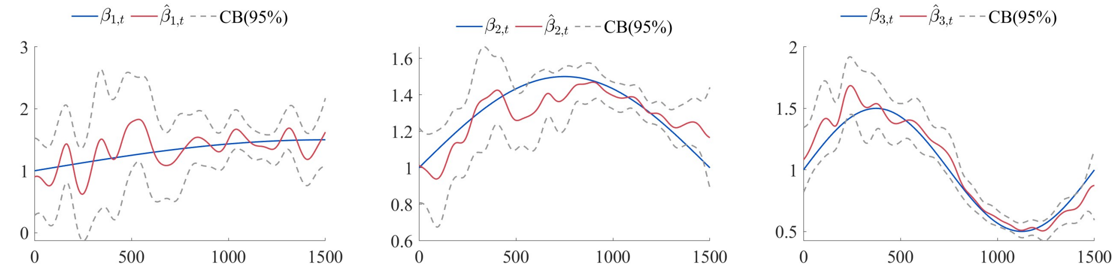
$$y_t = \beta_{1t} + \beta_{2t} z_{2t} + \beta_{3t} z_{3t} + u_t, \quad u_t = h_t \varepsilon_t, \tag{56}$$

where $z_{1t} = 1$ and z_{2t}, z_{3t} are defined using μ_{2t}, μ_{3t} and η_{2t}, η_{3t} as in (54).

We consider two simulation models. Model 6.3 assumes deterministic parameters and scale factors, while Model 6.4 combines deterministic and stochastic parameters and scale factors.

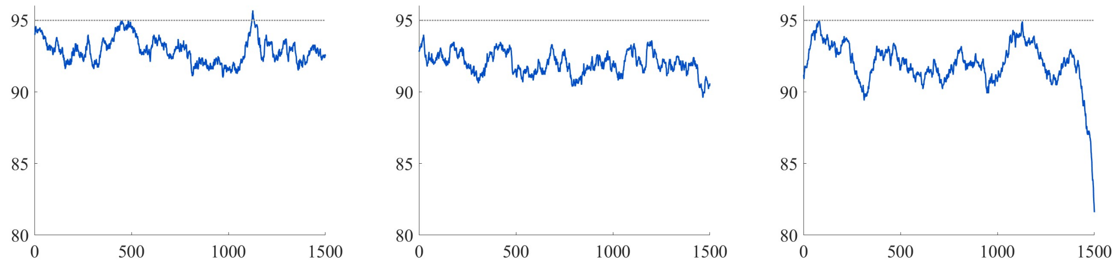
Model 6.3. *y_t follows (56) with ε_t as in (55). The scale factors h_t, g_{2t}, g_{3t} and parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ are deterministic:*

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



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

Figure 4: 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 5: Coverage rates (in %) of robust confidence intervals for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3: $n = 1500$, bandwidth $H = n^{0.5}$.

$$\beta_{1t} \;=\; 0.5\sin(0.5\pi t/n) + 1, \quad \beta_{2t} = 0.5\sin(\pi t/n) + 1, \quad \beta_{3t} = 0.5\sin(2\pi t/n) + 1.$$

Model 6.4. *y_t follows (56) with $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$ and scale factors:*

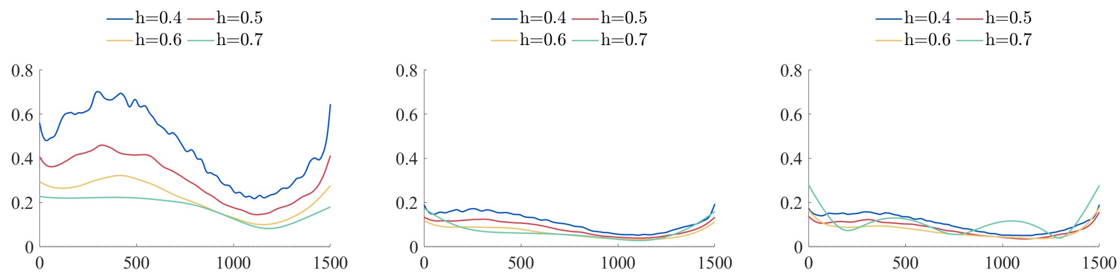
$$h_t = 0.5\sin(2\pi t/n) + 1, \quad g_{2t} = \Big| n^{-\gamma} \sum_{i=1}^{t} \zeta_j \Big| + 0.2, \quad g_{3t} = 0.5\sin(\pi t/n) + 1.$$

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

$$\beta_{3t} = \Big| n^{-\gamma} \sum_{i=1}^{t} \nu_j \Big| + 0.3(t/n),$$

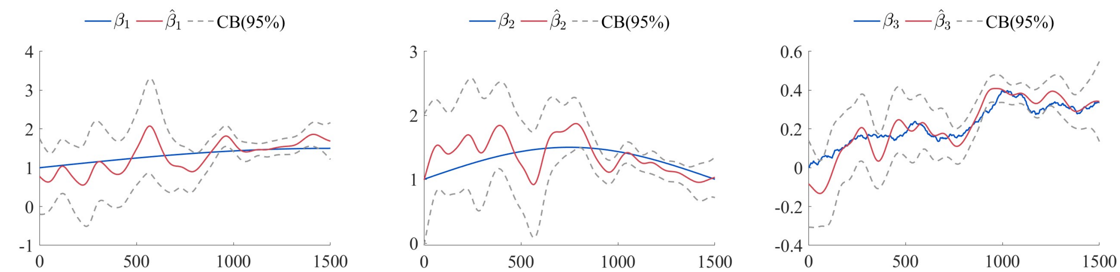
where $\{\zeta_j\}, \{\nu_j\}$ are stationary ARFIMA$(0,d,0)$ processes with memory parameter $d = 0.4$. We estimate β_t using the estimator $\widehat{\beta}_t$, (22), where the weights $b_{n,tj} = K(|t-j|/H)$ are computed with the Gaussian kernel function $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ with bandwidth $H = n^h$, $h = 0.4, 0.5, 0.6, 0.7$.

Figure 4 displays parameter estimation results for a single simulation from Model 6.3. It depicts the estimates $\widehat{\beta}_{k1}, ..., \widehat{\beta}_{kn}$ (red line) against 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



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

Figure 6: RMSE for time-varying parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ in Model 6.3: $n = 1500$, bandwidth $H = n^h$, $h = 0.4, 0.5, 0.6, 0.7$.



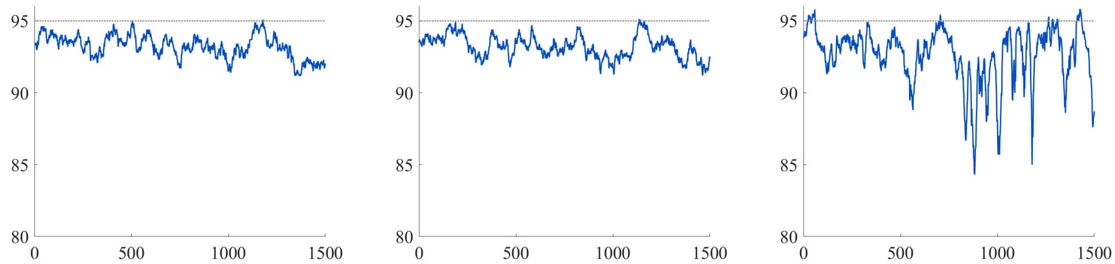
(a) $\beta_{1,t}$ (b) $\beta_{2,t}$ (c) $\beta_{3,t}$

Figure 7: 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.

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

Figure 5 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 time points. Figure 6 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, for example, $H = n^{0.6}$.

Figure 7 reports estimation results for a single simulation from Model 6.4, and Figure 8 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 time points, see Figure 7(c). Figure 8(c) shows that coverage rates of robust time-varying confidence intervals for β_{3t} might be slightly affected by stochastic variation in the parameter and scale factors. Nevertheless, they are still satisfactory and reasonably close to the nominal 95% coverage.



(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 6.4: $n = 1500$, bandwidth $H = n^{0.5}$.

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

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.

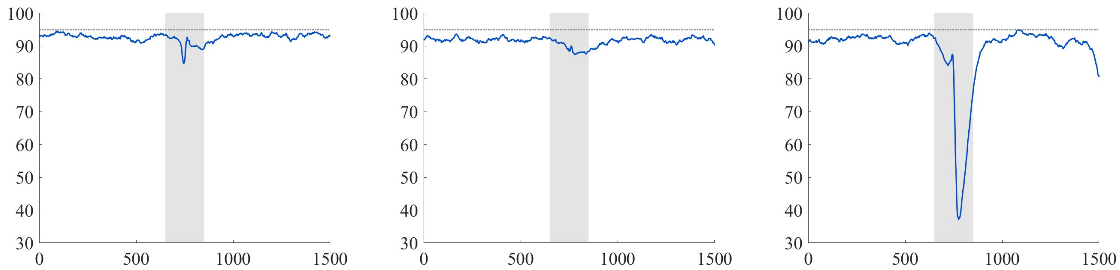
Type 2. 500 single observations are missing at randomly selected times.

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 noticeable changes in Bias, RMSE and SD, and the coverage rate for robust confidence intervals remains around 95%. At the same time, the coverage rate CP_{st} of the standard confidence intervals is substantially distorted.

Table 4 shows that randomly missing data do not affect the coverage rate of robust confidence intervals which remains to the nominal 95%, while the coverage rate of the standard confidence intervals drops to around 65%. This emphasises the flexibility of the robust OLS

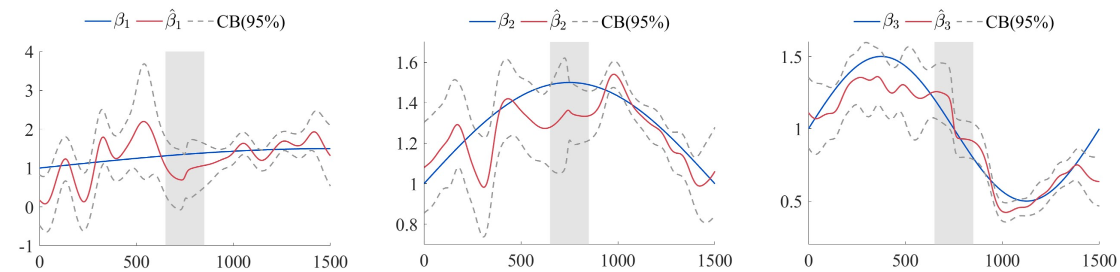
Table 4: Robust OLS estimation in Model 6.1 with randomly missing data (Type 2).

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



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

Figure 9: 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 10: 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.

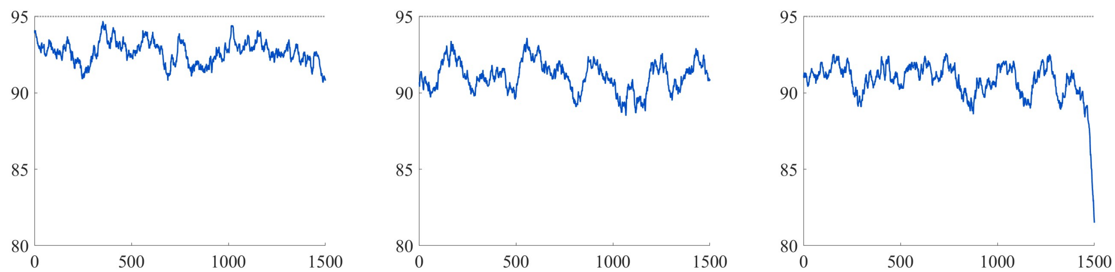
estimation of the fixed parameter in the presence of block or randomly missing data.

Figures 9 – 11 report estimation results for Model 6.3 with time-varying parameter β_t.

Figure 9 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 a larger distortion for parameter $\beta_{3,t}$ within 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 estimating β_t, the coverage distortion seems to be offset by the smooth down-weighting of the data, and the performance of the robust time-varying OLS estimation exceeds expectations.

Figure 10 reports the path of the estimator $\widehat{\beta}_{kt}$ and the point-wise robust confidence intervals, for a single simulation. The robust confidence intervals become wider in the shaded region, which likely explains the satisfactory coverage performance during that period.

Figure 11 shows that randomly missing data (Type 2) do 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, robust estimation of time-varying parameter does not appear be affected by randomly missing data.



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

Figure 11: 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 (57).

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

6.4 Estimation of a stationary AR(p) model

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

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t, \quad \beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.4, 0.3)', \tag{57}$$

where $\varepsilon_t = e_t e_{t-1}$, $e_t \sim i.i.d.\mathcal{N}(0,1)$ is a stationary martingale difference noise. The 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, the parameter β can be estimated by using the robust estimation method.

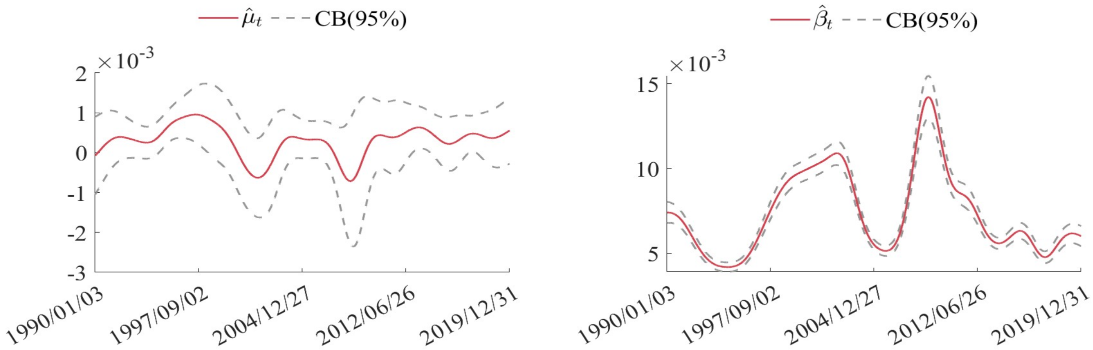
Table 5 shows that the coverage rate for the robust OLS estimation is close to the nominal 95%, while the standard OLS estimation exhibits extensive coverage distortion for β_2 and β_3.

7 Empirical experiment

In this section, we analyze the structure and dynamics of daily S&P 500 log returns, r_t, from 02/01/1990 to 31/12/2019, (sample size $n = 7558$). We employ robust regression estimation to assess whether the returns r_t can be modelled using a time-varying regression model of the form

$$r_t = \mu_t + u_t, \quad u_t = h_t \varepsilon_t, \tag{58}$$

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 the time-varying mean μ_t, the scale factor



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

Figure 12: Robust 95% confidence bands for μ_t in model (58) and $\beta_{1t} = h_t E|\varepsilon_t|$ in model (59), $n = 7558$, $H = n^{0.6}$.

h_t, and to test for the absence of autocorrelation in the absolute residuals $|u_t| = h_t|\varepsilon_t|$, thereby assessing the fit of the model (58) to the data.

It returns r_t follows the model (58) with i.i.d. noise ε_t, then the absolute residuals $|u_t|$'s are uncorrelated then for $t \neq s$:

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

Conversely, if the noise ε_t exhibits ARCH effects (stationary conditional heteroskedasticity), the sequence $|u_t|$ becomes autocorrelated, and the null hypothesis of uncorrelated absolute residuals $|u_t|$ would be rejected.

We estimate the the time varying mean μ_t using the time-varying OLS estimator with bandwidths $H = n^{0.4}, n^{0.5}, ..., n^{0.7}$. Figure 12(a) shows the estimated path of $\widehat{\mu}_t$ and the associated 95% confidence intervals for bandwidth $H = n^{0.6}$ indicating that μ_t is very likely to change over time.

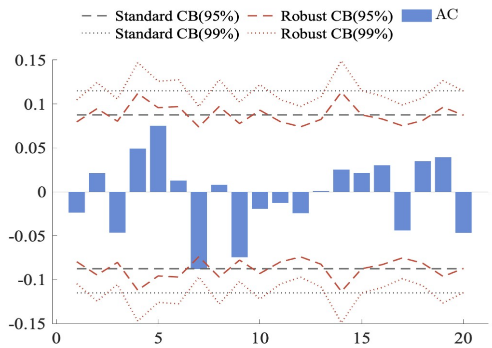
Assumption (58) implies that

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

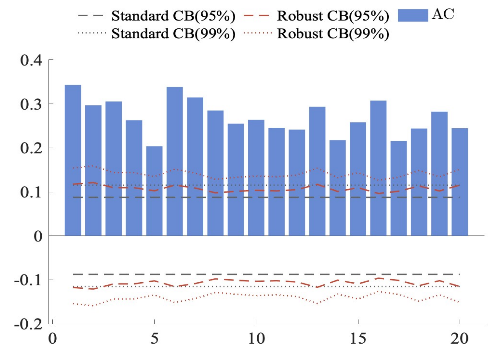
Therefore, $|\widehat{u}_t| = |r_t - \widehat{\mu}_t| \sim h_t E|\varepsilon_t| + h_t(|\varepsilon_t| - E|\varepsilon_t|)$ and thus $y_t = |\widehat{u}_t|$ follows a time-varying regression model of the form

$$y_t = \beta_{1t} + \widetilde{u}_t, \quad \widetilde{u}_t = g_t \eta_t, \tag{59}$$

where $\beta_{1t} = h_t E|\varepsilon_t|$ represents a time-varying intercept, $g_t = h_t$ denotes the scale factor, and $\eta_t = |\varepsilon_t| - E|\varepsilon_t|$ is an i.i.d. noise. Hence β_{1t} can be consistently estimated using the time-varying OLS estimator $\widehat{\beta}_{1t}$. Figure 12(b) displays the estimated path of $\widehat{\beta}_{1t}$ and the corresponding 95% confidence intervals for $\beta_{1t} = h_t E|\varepsilon_t|$ with bandwidth $H = n^{0.6}$, revealing pronounced time variation in the scale factor h_t.



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



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

Figure 13: 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%.

Figure 13(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 procedures developed in Giraitis et al. (2024). Given that the sample size is large ($n = 7558$) and β_{1t} is estimated non-parametrically with bandwidth $H = n^{0.6}$, we restrict the correlation analysis to the subsample $j \in [500, 1000]$. Both tests provide no evidence of significant correlation within this subsample, suggesting that the model (58) fits the returns r_t well during this time period.

The same is not likely to be true if $r_t^* = r_t - \widehat{\mu}_t$ follows a GARCH(1,1) process, as confirmed by the following experiment. We fit a GARCH(1,1) model to the demeaned returns $r_t^* = r_t - \widehat{\mu}_t$,

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

We generate a simulated GARCH(1,1) sample $r_{g1}^*, ..., r_{gn}^*$, apply the regression model (59) to the absolute values $y_t^* = |r_{gt}^*|$, and compute the residuals, $\widehat{u}_t^* = y_t^* - \widehat{\beta}_{1t}$. Figure 13(b) shows that both standard and robust tests detect significant correlation in residuals $\widehat{u}_t^*$, confirming the presence of conditional heteroskedasticity in the simulated GARCH data.

8 Conclusion

The robust OLS and time-varying OLS estimation and inference methods developed in this paper offer considerable flexibility for modelling economic and financial data. They allow for general heterogeneity in regression components and for structural change of regression coefficients over time. Moreover, the generalization of the structure of regressors and error terms further expands the range of empirical settings to which robust OLS regression framework can be applied. In particular, the paper develops asymptotic theory for general regression models with stochastic regressors possibly including a time varying mean, and provides data-based robust standard errors that enable the construction of confidence intervals for regression parameters. The Monte Carlo analysis demonstrates the strong performance of the robust estimation approach under complex settings, and confirms the asymptotic normality property and consistency of the proposed 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.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bardet, J.-M. and O. Wintenberger (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics 37*, 2730–2759.

Boldea, O., A. Cornea-Madeira, and A. Hall (2019). Bootstrapping structural change tests. *Journal of Econometrics 213*(2), 359–397.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. Springer.

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.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2024). Addressing covid-19 outliers in bvars with stochastic volatility. *The Review of Economics and Statistics 106*(5), 1403–1417.

Cattaneo, M., M. Jansson, and W. Newey (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association 113*(523), 1350–1361.

Chan, J. C. C., G. Koop, and Y. Xuewen (2024). Large order-invariant bayesian vars with stochastic volatility. *Journal of Business and Economic Statistics 42*(2), 825–837.

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.

Dahlhaus, R. and L. Giraitis (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis 19*(6), 629–655.

Dahlhaus, R. and S. Richter (2023). Adaptation for nonparametric estimators of locally stationary processes. *Econometric Theory 39*, 1123–1153.

Dahlhaus, R., S. Richter, and W. Wu (2019). Towards a general theory for locally stationary processes. *Bernoulli 25*(2), 1013–1044.

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

Doukhan, P. and O. Wintenberger (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications 118*(11), 1997–2013.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics 34*, 447–456.

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

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance 58*(3), 1269–1300.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *Annals of Statistics 27*(5), 1491–1518.

Fu, Z., Y. Hong, L. Su, and X. Wang (2023). Specification tests for time-varying coefficient models. *Journal of Econometrics 235*, 720–744.

Georgiev, I., D. Harvey, S. Leybourn, and A. Robert Taylor (2018). Testing for parameter instability in predictive regression models. *Journal of Econometrics 204*, 101–118.

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., G. Kapetanios, and T. Yates (2020). Asymptotic theory for time series with changing mean and variance. *Journal of Econometrics 219*, 281–313.

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.

Hall, A., S. Han, and O. Boldea (2012). Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics 270*(2), 281–302.

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

Hansen, B. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Hidalgo, J., Y. Lee, and M.-H. Seo (2019). Robust inference for threshold regression models. *Journal of Econometrics 210*, 291–309.

Hu, Z., I. Kasparis, and Q. Wang (2024). Time-varying parameter regressions with stationary presistent data. *Econometric Theory publisked online*, 1–17.

Jochmans, K. (2019). Heteroscedasticity-robust inference in linear regression models with many covariates. *Journal of the American Statistical Association 117*(538), 887–896.

Karmakar, S., S. Richter, and W. B. Wu (2022). Simultaneous inference for time-varying models. *Journal of Econometrics 227*(2), 408–428.

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.

Li, D., P. Phillips, and J. Gao (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics 215*, 607–632.

Linton, O. and Z. Xiao (2019). Efficient estimation of nonparametric regression in the presence of dynamic heteroskedasticity. *Journal of Econometrics 213*, 608–631.

MacKinnon, J. G. (2012). Thirty years of heteroscedasticity-robust inference. *in Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis, eds. X. Chen and N. R. Swanson, New York: Springer*, 437–461.

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

Nyblom, J. (1989b). 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.

Sun, Y., Y. Hong, T.-H. Lee, S. Wang, and X. Zhang (2021). Time-varying model averaging. *Journal of Econometrics 222*, 974–992.

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.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics 40*(5), 2601–2633.

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

White, H. (2014). *Asymptotic Theory for Econometricians*. Academic Press.

Wu, W. (2005). M-estimation of linear models with dependent errors. *Annals of Statistics 25*(2), 495–521.

Xu, K.-L. and P. Phillips (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics 142*, 265–280.

Zhang, T. and W. Wu (2012). Inference of time-varying regression models. *Annals of Statistics 40*(3), 1376–1402.

Zhang, T. and W. Wu (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Annals of Statistics 43*(2), 741–768.

Online Supplement to

"Unlocking the Regression Space"

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

[1]Queen Mary University of London

[2]King's College London

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, Corollaries 2.1 and 2.2, and Lemma 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

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

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. $\square$

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.

Recall the definition of the information set $\mathcal{F}_n^* = \sigma\big(\mu_t, g_t, h_t, t = 1, ..., n\big)$.

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),

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

Hence,

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

By (12.11) of Lemma 12.2,

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

This together with (9.3) implies:

$$\frac{a'D(\widehat{\beta} - \beta)}{\sqrt{a'D\Omega_nDa}} = v_n^{-1}d_nS_{zu} + o_p(1).$$

Write

$$s_n \;\; = \;\; v_n^{-1}d_nS_{zu} = \sum_{t=1}^{n}\xi_t, \quad \xi_t = v_n^{-1}d_nz_tu_t.$$

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

$\mathcal{F}_{n,t} = \sigma(\varepsilon_1, ..., \varepsilon_t;\ \mu_s, h_s, g_s, s = 1, ..., n)$:

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

The latter follows noting that the variables v_n^{-1}, d_n, h_t are $\mathcal{F}_{n,t-1}$-measurable since they are function of $\mu_s, h_s, g_s, s = 1, ..., n$. Similarly, since η_t's are $\mathcal{F}_{n,t-1}$ measurable (see Assumption 2.2), the variables $z_t = \mu_t + I_{gt}\eta$ are also $\mathcal{F}_{n,t-1}$-measurable. Finally, by assumption, $\{\mu_s, h_s, g_s, s = 1, ..., n\}$ and $\{\varepsilon_s, s = 1, ..., n\}$ are mutually independent, and therefore $E[\varepsilon_t|\mathcal{F}_{n,t-1}] = E[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ by Assumption 2.1. This shows that the conditional expectation property $E[\xi_t\,|\mathcal{F}_{n,t-1}] = 0$ is preserved for ξ_t and completes the argument showing that ξ_t is a martingale difference sequence with respect to the σ-field $\mathcal{F}_{n,t-1}$.

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

$$(a) \quad \sum_{t=1}^{n} E[\xi_t^2\,|\mathcal{F}_{n,t-1}] \to_p 1, \tag{9.7}$$

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

Observe that (a) holds with a non-random limit $\eta^2 = 1$. Thus, the verification of the condition (3.21) of Corollary 3.1, that the σ-fields are nested, $\mathcal{F}_{n,t} \subset \mathcal{F}_{n+1,t}$ for $t = 1, ..., n$ and $n \geq 1$, is unnecessary; see remark on page 59 in Hall and Heyde (1980). 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,

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

Recall that by (9.2), $D E[S_{zz}|\mathcal{F}_n^*]^{-1} D = O_p(1)$. 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

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

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,

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

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

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

Hence,

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

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 $j \ne k$. 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,

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

by (13). This completes the proof of the theorem. □

Proof of Corollary 2.1. We will show that

$$
\frac{\widehat{\omega}_{kk}}{\omega_{kk}} = 1 + o_p(1) \tag{9.9}
$$

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 (9.9), we will verify that

$$
D \widehat{\Omega}_n D = D \Omega_n D + o_p(1) \tag{9.10}
$$

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

$$a' D \Omega_n D a \geq b_n, \quad a' D \Omega_n D a \leq b_{n2} \tag{9.11}$$

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 $j \neq k$ and $a_k = 1$. Then $a' D \Omega_n D a = v_k^2 \omega_{kk}$, and by (9.11), $v_k^2 \omega_{kk} \geq b_n > 0$. This proves (9.9):

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

In addition, the bounds (9.11) imply that $\sqrt{\omega_{kk}} \asymp_p v_k^{-1}$:

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

Proof of (9.10). Set $V_n = D D_g^{-1}$. By (7) of Assumption 2.3, $V_n = O_p(1)$. We have

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

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

$$D^{-1} S_{zz\widehat{u}\widehat{u}} D^{-1} = D^{-1} S_{zzuu} D^{-1} + o_p(1). \tag{9.12}$$

This implies (9.10):

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

Proof of (9.12). By definition,

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

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 (9.12), it suffices to show that

$$i_n = o_p(1). \tag{9.13}$$

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

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

by (12.53) of Lemma 12.3. This implies (9.13) and completes the proof of the corollary. $\square$

Proof of Corollary 2.2. Let β_k be the true value of the k-th component of the parameter β, and suppose that $\beta_k \ne \beta_k^0$. Write

$$t_n = \frac{\widehat{\beta}_k - \beta_k^0}{\sqrt{\widehat{\omega}_{kk}}} = \frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} + \frac{\beta_k - \beta_k^0}{\sqrt{\widehat{\omega}_{kk}}} =: t_{n,1} + t_{n,2}.$$

By (16) of Corollary 2.1, $t_{n,1} \to_d \mathcal{N}(0,1)$ and $\sqrt{\omega_{kk}} \asymp_p v_k^{-1}$. Hence,

$$t_{n,1} = O_p(1), \quad t_{n,2} \asymp_p v_k \to_p \infty.$$

Then, $t_n = t_{n,1} + t_{n,2} = O_p(1) + t_{n,2} \asymp_p v_k \to_p \infty$, which proves the claim of Corollary 2.2. $\square$

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

$$i_n = v_{gk}^{-2} \max_{1\le t\le n} (g_{kt}^2 + \mu_{kt}^2) = o_p(1). \tag{9.14}$$

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

$$E[z_{kt}^2\,|\mathcal{F}_n^*] = \mu_{kt}^2 + 2\mu_{kt}g_{kt}E[\eta_{kt}\,|\mathcal{F}_n^*] + g_{kt}^2E[\eta_{kt}^2\,|\mathcal{F}_n^*] = \mu_{kt}^2 + g_{kt}^2. \tag{9.15}$$

In addition, by assumption (19) of lemma, $v_{gk}^{-2} = (\sum_{t=1}^n g_{kt}^2)^{-1} = O_p(n^{-1})$. Thus,

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

We will show that $Ei_{n,1} = o(1)$ which implies (9.14). Observe that for any $L \ge 1$,

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

By assumption (18), $E[z_{kt}^4] \le c < \infty$ where c does not depend on t, n. Hence,

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

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

Proof of (11). It suffices to verify that

$$i_n = v_k^{-2} \max_{1\le t\le n} (g_{kt}^2 + \mu_{kt}^2)h_t^2 = o_p(1). \tag{9.16}$$

By assumption (19) of lemma, $v_k^{-2} = O_p(n^{-1})$. This together with (9.15) implies that

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

We will show that $Ei_{n,2} = o(1)$ which implies $i_{n,2} = o_p(1)$ and proves (9.16).

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

$$\begin{aligned} z_{kt}^2h_t^2 &\le L + z_{kt}^2h_t^2I(z_{kt}^2h_t^2 \ge L) \\ &\le L + L_0^{-1}z_{kt}^4I(h_t^2 \le L_0^{-1}z_{kt}^2) + L_0h_t^4I(h_t^2 > L_0^{-1}z_{kt}^2)I(h_t^4L_0 \ge L) \\ &\le L + L_0^{-1}z_{kt}^4 + h_t^4L_0\big(\frac{h_t^4}{LL_0^{-1}}\big)^\delta \\ &\le L + L_0^{-1}z_{kt}^4 + h_t^{4+4\delta}A_L, \quad A_L = L^{-\delta}L_0^{1+\delta}. \end{aligned}$$

By assumption (18), $E[z_{kt}^4] \le c$ 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, $E[h_t^{4+4\delta}] = E[(E[u_t^2\,|\mathcal{F}_n^*])^{2+2\delta}] \leq E[|u_t|^{4+4\delta}] \leq c$. Notice that $A_L \to 0$ as $L \to \infty$. Therefore, as $n, L \to \infty$,

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

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

Proof of (7). By assumption (18) of Lemma 2.1, $E[z_{kt}^4] \leq c$ and $E[u_t^4] \leq c$ where $c < \infty$ does not depend on t, k, n. By (9.15),

$$\begin{aligned}
\mu_{kt}^2 &\leq E[z_{kt}^2\,|\mathcal{F}_n^*], \quad E[\mu_{kt}^2] \leq E[z_{kt}^2] \leq c, \\
g_{kt}^2 &\leq E[z_{kt}^2\,|\mathcal{F}_n^*], \quad E[g_{kt}^2] \leq E[z_{kt}^2] \leq c, \\
E[\mu_{kt}^4] &\leq E[(E[z_{kt}^2\,|\mathcal{F}_n^*])^2] \leq E[(E[z_{kt}^4\,|\mathcal{F}_n^*])] \leq E[z_{kt}^4] \leq c, \\
E[g_{kt}^4] &\leq E[(E[z_{kt}^2\,|\mathcal{F}_n^*])^2] \leq \ c, \\
E[h_t^4] &= E[(E[u_t^2\,|\mathcal{F}_n^*])^2] \leq E[(E[u_t^4\,|\mathcal{F}_n^*])] \leq E[u_t^4] \leq c, \\
E[\mu_{kt}^2 h_t^2] &\leq (E[\mu_{kt}^4]E[h_t^4])^{1/2} \leq c, \\
E[g_{kt}^2 h_t^2] &\leq (E[g_{kt}^4]E[h_t^4])^{1/2} \leq c, \qquad (9.17)
\end{aligned}$$

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

$$\begin{aligned}
&n^{-1}E[\textstyle\sum_{t=1}^n \mu_{kt}^2] \leq c, \quad \textstyle\sum_{t=1}^n \mu_{kt}^2 = O_p(n), \\
&n^{-1}E[\textstyle\sum_{t=1}^n \mu_{kt}^2 h_t^2] \leq c, \quad \textstyle\sum_{t=1}^n \mu_{kt}^2 h_t^2 = O_p(n), \\
&n^{-1}E[\textstyle\sum_{t=1}^n g_{kt}^2 h_t^2] \leq c, \quad \textstyle\sum_{t=1}^n g_{kt}^2 h_t^2 = O_p(n) \\
&n^{-1}E[\textstyle\sum_{t=1}^n g_{kt}^2] \leq c, \quad \textstyle\sum_{t=1}^n g_{kt}^2 = O_p(n).
\end{aligned}$$

By assumption (19), $n/v_k^2 = O_p(1)$ and $n/v_{gk}^2 = O_p(1)$. Thus,

$$\begin{aligned}
&v_{gk}^{-1}\textstyle\sum_{t=1}^n \mu_{kt}^2 = O_p(n/v_{gk}^2) = O_p(1), \\
&v_k^{-1}\textstyle\sum_{t=1}^n \mu_{kt}^2 h_t^2 = O_p(n/v_k^2) = O_p(1), \\
&v_{gk}^{-1}v_k = v_{gk}^{-1}\textstyle\sum_{t=1}^n g_{kt}^2 h_t^2 = O_p(n/v_{gk}^2) = O_p(1), \\
&v_k^{-1}v_{gk} = v_k^{-1}\textstyle\sum_{t=1}^n g_{kt}^2 = O_p(n/v_k^2) = O_p(1),
\end{aligned}$$

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

10 Proofs of Theorem 3.1 and Corollaries 3.1 and 3.2

Proof of Theorem 3.1. Recall the notation introduced in Section 3. Set

$$\widetilde{y}_j = b_{n,tj}^{1/2} y_j, \quad \widetilde{z}_j = b_{n,tj}^{1/2} z_j, \quad \widetilde{u}_j = b_{n,tj}^{1/2} u_j.$$

Then we can write

$$\widetilde{y}_j = \widetilde{z}_j' \beta_t + \widetilde{u}_j + r_j, \quad r_j = (\beta_j - \beta_t)' z_j.$$

Recall the estimator $\widehat{\beta}_t$ given in (22). In Section 3 we introduced an auxiliary regression model with a fixed parameter $\beta = \beta_t$:

$$y_j^* \;=\; \beta' \widetilde{z}_j + \widetilde{u}_j, \;\; \widetilde{u}_j = b_{n,tj}^{1/2} u_j, \quad j = 1, ..., n. \tag{10.1}$$

Recall the OLS estimator $\widehat{\beta}$ of the fixed parameter β in this model, given in (27):

$$\widehat{\beta} = \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{z}_j'\big)^{-1} \big(\sum_{j=1}^n \widetilde{z}_j y_j^*\big) = \beta + \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{z}_j'\big)^{-1} \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{u}_j\big).$$

In (28) we showed that following relation:

$$\widehat{\beta}_t - \beta_t \;=\; \widehat{\beta} - \beta + R_t, \quad R_t = \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{z}_j'\big)^{-1} \big(\sum_{j=1}^n \widetilde{z}_j \widetilde{z}_j' (\beta_j - \beta_t)\big) \tag{10.2}$$

The remainder $R_t = (R_{1t}, ..., R_{pt})'$ arises due to time variation in the parameter β_j and is negligible. We will obtain an upper bound for this term. The term $\widetilde{\beta} - \beta$ is the main component We will analyse it using the results of Section 2. Overall, equation (10.2) shows that properties of $\widehat{\beta}_t - \beta_t$ are determined by the properties of $\widehat{\beta} - \beta$, with an additional negligible term R_t.

First we will show that the components of $\widehat{\beta} - \beta = (\widetilde{\beta}_1 - \beta_1, ..., \widetilde{\beta}_p - \beta_p)'$ and R_t satisfy the following properties. For $k = 1, ..., p$,

$$\widetilde{\beta}_k - \beta_k \;=\; O_p(H^{-1/2}), \quad \frac{\widetilde{\beta}_k - \beta_k}{\sqrt{\omega_{kk,t}}} \to_d \mathcal{N}(0,1), \quad \sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}, \tag{10.3}$$

$$R_{kt} \;=\; O_p\big((H/n)^\gamma\big). \tag{10.4}$$

Proof of (10.3). Recall that $\widetilde{z}_j = (\widetilde{z}_{1j}, ..., \widetilde{z}_{pj})'$, and

$$\widetilde{z}_{kj} \;=\; \widetilde{\mu}_{kj} + \widetilde{g}_{kj} \eta_{kj}, \quad \widetilde{u}_j = \widetilde{h}_j \varepsilon_j, \tag{10.5}$$

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

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,

$$\widetilde{\beta}_k - \beta_k = O_p(v_k^{-1}) = O_p(H^{-1/2}),$$

where $v_k^2 \equiv v_{kt}^2 = \sum_{j=1}^n \widetilde{g}_{kj}^2 \widetilde{h}_j^2 = \sum_{j=1}^n b_{n,tj} g_{kj}^2 h_j^2$ and

$$v_k^2 \asymp_p H. \tag{10.6}$$

Indeed, $v_{kt}^{-2} = O_p(H^{-1})$ by (30) of Assumption 3.2. On the other hand, (9.17) implies that $Ev_{kt}^2 \le \sum_{j=1}^n b_{n,tj} E[g_{kj}^2 h_j^2] \le c \sum_{j=1}^n b_{n,tj} = O(H)$, where the last relation easily follows using definition of $b_{n,tj}$ and (24). Hence $v_{kt}^2 = O_p(H)$, which proves (10.6).

This complete the proof of the first claim in (10.3), while the second claim holds by (14) of Theorem 2.2. The third claim holds since by (16) of Corollary 2.1 and (10.6),

$$\sqrt{\omega_{kk,t}} \asymp_p v_k^{-1} = (\sum_{j=1}^n \widetilde{g}_{kj}^2 \widetilde{h}_j^2)^{-1/2} \asymp_p H^{-1/2}. \tag{10.7}$$

Proof of (10.4). Write

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

We will show that

$$||S_{\tilde{z}\tilde{z},t}^{-1}|| = O_p(H^{-1}), \quad ||S_{\tilde{z}\tilde{z}\beta,t}|| = O_p\big(H(H/n)^\gamma\big), \tag{10.8}$$

which implies $||R_t|| \le ||S_{\tilde{z}\tilde{z},t}^{-1}||\,||S_{\tilde{z}\tilde{z}\beta,t}|| = O_p\big((H/n)^\gamma\big)$. Then, $|R_{kt}| \le ||R_t|| = O_p\big((H/n)^\gamma\big)$ which proves (10.4).

To verify (10.8), recall notation of the $p \times p$ diagonal matrix

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

Notice that

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

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 (10.8). Next, bound

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

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 (24),

$$\sum_{j=1}^{n} b_{n,tj}(|t-j|/H)^{\gamma} = O(H).$$

This implies

$$E||S_{\tilde{z}\tilde{z}\beta,t}|| \ \le \ CH(H/n)^{\gamma}\Big(H^{-1}\sum_{j=1}^{n} b_{n,tj}(|t-j|/H)^{\gamma}\Big) \le CH(H/n)^{\gamma} \tag{10.9}$$

which proves the second claim in (10.8).

We now are ready to prove the claims (31) and (32) of the theorem. First, together with (10.2), the properties (10.3) and (10.4) establish the consistency result (31):

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

To prove the asymptotic normality property (32), recall assumption $H = o(n^{2\gamma/(2\gamma+1)})$. Then

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

because by (10.7) and (10.4),

$$\omega_{kk,t}^{-1/2}B_t = O_p\big(H^{1/2}\big)O_p\big((H/n)^{\gamma}\big) = O_p\big(H^{1/2}(H/n)^{\gamma}\big) = o_p(1)$$

under assumption $H = o(n^{2\gamma/(2\gamma+1)})$. 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 (10.3) which proves the asymptotic normality property (32) of the theorem. Noting that $\sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}$, as shown in (10.7), this completes the proof of the theorem. $\square$

Proof of Corollary 3.1. In the proof of Theorem 3.1 we wrote the time-varying regression model as a regression model

$$\widetilde{y}_j = \widetilde{z}_j'\beta + \widetilde{u}_j + r_j, \quad r_j = (\beta_j - \beta_t)'\widetilde{z}_j \tag{10.10}$$

with a fixed parameter $\beta = \beta_t$. We showed that the regressors $\widetilde{z}_j$ and the noise $\widetilde{u}_j$ satisfy assumptions of Theorem 2.2 and that the contribution of the term r_j is asymptotically negligible. That allowed us to establish the asymptotic normality property (32) 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 (34),

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

Proof of the corresponding result in the case of fixed parameter in Corollary 2.1 shows that we need to verify the validity of (9.12) for our regression model (10.10), i.e. to show that

$$j_n = D^{-1} S_{\widetilde{z}\widetilde{z}\widehat{u}\widehat{u}} D^{-1} = D^{-1} S_{\widetilde{z}\widetilde{z}\widetilde{u}\widetilde{u}} D^{-1} + o_p(1), \tag{10.11}$$

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

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 (9.12), $j_{n1} = D^{-1} S_{\widetilde{z}\widetilde{z}\widetilde{u}\widetilde{u}} D^{-1} + o_p(1)$. Hence, to prove (10.11), we need to show that

$$j_{n2} = o_p(1). \tag{10.12}$$

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 (10.12). Notice that

$$\begin{aligned}
\widehat{u}_j &= \widetilde{y}_j - \widehat{\beta}_t' \widetilde{z}_j = (\beta_t - \widehat{\beta}_t)' \widetilde{z}_j + \widetilde{u}_j + r_j = \widehat{u}_j^* + r_j, \\
\widehat{u}_j^2 - \widehat{u}_j^{*\,2} &= (\widehat{u}_j - \widehat{u}_j^*)^2 + 2(\widehat{u}_j - \widehat{u}_j^*)\widehat{u}_j^* \\
&= r_j^2 + 2r_j \widehat{u}_j^* = r_j^2 + 2r_j(\beta_t - \widehat{\beta}_t)' \widetilde{z}_j + 2r_j \widetilde{u}_j.
\end{aligned} \tag{10.13}$$

Using the inequality $2|ab| \le a^2 + b^2$, we can bound in (10.13),

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

Next we evaluate $|r_j \widetilde{u}_j|$ in (10.13). Let $L > 1$ be large number. Then,

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

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

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

$$+2L||\widetilde{z}_j||^3\,|\widetilde{u}_j|\,||\beta_j-\beta_t||^2.$$

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,

$$\begin{aligned}
|j_{n3}| &= H^{-1}\big|\textstyle\sum_{j=1}^n \widetilde{z}_j\widetilde{z}_j'(\widehat{u}_j^2-\widehat{u}_j^{*\,2})\big| \le H^{-1}\sum_{j=1}^n ||\widetilde{z}_j||^2|\widehat{u}_j^2-\widehat{u}_j^{*\,2}| \\
&\le L\{H^{-1}\textstyle\sum_{j=1}^n b_{n,tj}^2||\beta_j-\beta_t||^2\theta_j\} + (||\beta_t-\widehat{\beta}_t||^2+L^{-1})\{H^{-1}\sum_{j=1}^n b_{n,tj}^2\theta_j\} \\
&\le L\{\textstyle\sum_{j=1}^n b_{n,tj}||\beta_j-\beta_t||^2\}\{H^{-1}\sum_{j=1}^n b_{n,tj}\theta_j\} \\
&\qquad + (||\beta_t-\widehat{\beta}_t||^2+L^{-1})\{H^{-1}\textstyle\sum_{j=1}^n b_{n,tj}^2\theta_j\} \\
&= Lq_{n1}q_{n2} + (||\beta_t-\widehat{\beta}_t||^2+L^{-1})q_{n3}.
\end{aligned} \tag{10.14}$$

By (31) 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

$$Eq_{n1} = o(1), \quad Eq_{n2} = O(1), \quad Eq_{n3} = O(1). \tag{10.15}$$

Combining this with (10.14), 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 (10.12).

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 and recall (10.9). Then,

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

when $H = o(n^{2\gamma/(2\gamma+1)})$. This proves (10.15) 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 (24) it holds $H^{-1}\sum_{j=1}^n b_{n,tj} = O(1)$ and $b_{n,tj}^2 \le Cb_{n,tj}$. Hence,

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

This completes the proof of (10.15) and the corollary. $\square$

Proof of Corollary 3.2. Let β_{kt} be the true value of the k-th component of the time-varying parameter β_t. Suppose that $|\beta^0_{kt} - \beta_{kt}| \geq a > 0$ for $t = t_n \in [1, ..., n]$ as $n \to \infty$. Write

$$\tau_{n,t} = \frac{\widehat{\beta}_{kt} - \beta^0_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} = \frac{\widehat{\beta}_{kt} - \beta_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} + \frac{\beta_{kt} - \beta^0_{kt}}{\sqrt{\widehat{\omega}_{kk,t}}} =: \tau_{n1,t} + \tau_{n2,t}.$$

By (34) of Corollary 3.1, $\tau_{n1,t} \to_d \mathcal{N}(0,1)$ and $\sqrt{\omega_{kk,t}} \asymp_p H^{-1/2}$. Hence,

$$\tau_{n1,t} = O_p(1), \quad \tau_{n2,t} \asymp_p H^{1/2} \to_p \infty.$$

Then, $\tau_{n,t} = \tau_{n1,t} + \tau_{n2,t} = O_p(1) + \tau_{n2,t} \asymp_p H^{1/2} \to_p \infty$, which proves the claim of the Corollary 3.2. □

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 (10.5) satisfy Assumption 2.3 and Assumption 2.4(ii).*

Proof of Lemma 10.1. Notice that assumptions (24) 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 Theorems 4.1, 4.2 and Theorem 5.1

Proof of Theorem 4.1. Suppose that $y_t = \beta' z_t + u_t$ follows the regression model (1). In the presence of missing data, estimation of the parameter β is based on a regression model with the fixed parameter (39):

$$\widetilde{y}_t = \beta' \widetilde{z}_t + \widetilde{u}_t, \tag{11.1}$$

where the regressors $\widetilde{z}_t = (\widetilde{z}_{1t}, ..., \widetilde{z}_{pt})'$ and the noise $\widetilde{u}_t$ take the form

$$\begin{aligned} \widetilde{z}_{kt} &= \widetilde{\mu}_{kt} + \widetilde{g}_{kt}\eta_{kt}, \quad \widetilde{\mu}_{kt} = \tau_t \mu_{kt}, \quad \widetilde{g}_{kt} = \tau_t g_{kt}, \\ \widetilde{u}_t &= \widetilde{h}_t \varepsilon_t, \quad \widetilde{h}_t = \tau_t h_t, \end{aligned} \tag{11.2}$$

and τ_t is the missing data indicator. Under Assumptions 4.1 and 4.2 of the theoren, $\{\widetilde{\mu}_t, \widetilde{g}_t, \widetilde{h}_t\}$ are independent of $\{\varepsilon_t, \eta_t\}$. Therefore, (z_t, u_t) belongs to the regression space described in (2) and (3) of Section 2.

We estimate the fixed parameter β using the estimator defined in (41):

$$\widehat{\beta} \;=\; \Big(\sum_{t=1}^n \widetilde{z}_t \widetilde{z}_t'\Big)^{-1} \Big(\sum_{t=1}^n \widetilde{z}_t \widetilde{y}_t\Big). \tag{11.3}$$

We will show that $(\widetilde{z}_t, \widetilde{u}_t)$ satisfy Assumptions 2.1, 2.2, 2.3 and 2.4 of Theorem 2.2 of Section 2. Then, the required result (42) for $\widehat{\beta}$ of this theorem follows directly from the claims (16) of Corollary 2.1.

We split Assumptions 2.1, 2.2, 2.3 and 2.4 into two groups:

(a) Assumptions 2.1, 2.2, and 2.4(i), and

(b) Assumptions 2.3 and 2.4(ii).

Assumptions (a) imposed on the stationary processes η_t, ε_t are part of Assumption 4.2 of Theorem 4.1.

It remains to show the validity of the assumptions in group (b), i.e. that the means $\widetilde{\mu}_t$ and the scales $\widetilde{h}_t, \widetilde{g}_t$ satisfy Assumptions 2.3 and 2.4(ii). By Assumption 4.2, we have $Ez_{kt}^4 \le c$ and $Eu_t^4 \le c$. Moreover, $g_{kt} \ge c_1 > 0$ and $h_t \ge c_1 > 0$ where $c, c_1 > 0$ do not depend on k, t and n. For $k = 1, ..., p$, define:

$$\widetilde{v}_k^2 = \sum_{t=1}^n \widetilde{g}_{kt}^2 \widetilde{h}_t^2, \quad \widetilde{v}_{gk}^2 = \sum_{t=1}^n \widetilde{g}_{kt}^2. \tag{11.4}$$

Notice that

$$\widetilde{v}_k^2 \ge c_1^4 \sum_{t=1}^n \tau_t = c_1^4 N, \quad \widetilde{v}_{gk}^2 \ge c_1^2 \sum_{t=1}^n \tau_t = c_1^2 N,$$

where N is the size of the subsample (36). By assumption of the theorem, $n/N = O_p(1)$. Thus,

$$E\widetilde{z}_{kt}^4 \le Ez_{kt}^4 \le c, \quad E|\widetilde{u}_t|^{4+\delta} \le E|u_t|^{4+\delta} \le c, \tag{11.5}$$

$$n/\tilde{v}_k^2 = O(n/N) = O_p(1), \quad n/\tilde{v}_{gk}^2 = O(n/N) = O_p(1), \tag{11.6}$$

which confirms the validity of Assumptions 2.3 and 2.4(ii); see Lemma 2.1. This completes the proof of the theorem. □

Proof of Theorem 4.2. Now, suppose that $y_t = \beta_t' z_t + u_t$ follows the regression model (21) with a time-varying parameter β_t. In the presence of missing data, estimation of the time-varying parameter β_t is based on a model (43):

$$\widetilde{y}_t = \beta_t' \widetilde{z}_t + \widetilde{u}_t. \tag{11.7}$$

Here, the regressors $\widetilde{z}_t = (\widetilde{z}_{1t}, ..., \widetilde{z}_{pt})'$ and the noise $\widetilde{u}_t$ are the same as in (11.2). We showed in the proof of the Theorem 4.1 that (z_t, u_t) belongs to the regression space described in (2) and (3) of Section 2.

The estimator of the time-varying parameter β_t is given in (45):

$$\widehat{\beta}_t = \Big(\sum_{j=1}^n b_{n,tj} \widetilde{z}_j \widetilde{z}_j'\Big)^{-1} \Big(\sum_{j=1}^n b_{n,tj} \widetilde{z}_j \widetilde{y}_j\Big). \tag{11.8}$$

We will show that $(\widetilde{z}_t, \widetilde{u}_t)$ satisfy Assumptions 2.1, 2.2, 2.4(i), 3.1 and 3.2 of Theorem 3.1. Then, the results (46), (47) and (48) for $\widehat{\beta}_t$ of Theorem 4.2 follow from the results (31) of

Theorem 3.1 and (34) of Corollary 3.1.

Observe that Assumptions 2.1, 2.2, 2.4(i) on η_t, ε_t are part of Assumption 4.1 of this theorem, which also includes Assumption 3.1 for β_t.

It remains to show that $\widetilde{z}_t, \widetilde{u}_t$ satisfy Assumption 3.2. This requires to prove the validity of (11.5) and (11.6) under Assumption 4.2 of this theorem, which we showed in the proof of Theorem 4.1. □

Proof of Theorem 5.1. We consider a stationary AR(p) model (49),

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

where ε_t is a stationary $m.d.$ sequence with respect to the information set $\mathcal{F}_t = \sigma(\varepsilon_s, s \le t)$. Write it as a regression model (1),

$$y_t = \beta' z_t + u_t, \quad u_t = \varepsilon_t \tag{11.9}$$

with fixed parameter $\beta = (\beta_1, ..., \beta_{p+1})' = (\phi_0, ..., \phi_p)'$ and regressors $z_t = (z_{1t}, z_{2t}, ..., z_{p+1,t})' = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})'$. Under assumption (51) of theorem, AR(p) model has a stationary solution:

$$y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \text{where } \sum_{j=0}^{\infty} |a_j| < \infty, \ \mu = Ey_t, \tag{11.10}$$

and regressors

$$z_{kt} = \mu_{kt} + g_{kt}\eta_{kt}, \quad \mu_{kt} = E[y_{t-k}] = Ey_1, \quad g_{kt} = 1, \quad \eta_{kt} = y_{t-k} - E[y_{t-k}],$$

for $k = 2, ..., p+1$, satisfy regression assumption (3). From (11.10) it follows that the regressors $\eta_t = (\eta_{1t}, ..., \eta_{pt})' = (y_{t-1}, y_{t-2}, ..., y_{t-p})'$ are $\mathcal{F}_{t-1} = \sigma(\varepsilon_s, s \le t-1)$ measurable. Moreover, under the assumptions of the theorem, (ε_t, η_t) satisfy Assumptions 2.1, 2.2, 2.3 and 2.4 of Theorem 2.2 in Section 2. Finally, we show that $Ey_t^8 \le C < \infty$. Recall that by the assumption of theorem, ε_t is a stationary $m.d.$ sequence such that $E\varepsilon_t^8 < \infty$. It is known that if $E|\varepsilon_t|^p < \infty$, 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). Hence $E(y_t - \mu)^8 < \infty$ and $E\eta_{kt}^8 < \infty$ from $k = 1, ..., p$.

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*

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

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

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

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,

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

This implies

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

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

Recall notation

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

Recall definition $\mathcal{F}_n^* = \sigma(\mu_t, g_t, t = 1, \dots, n)$ 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 fol-*

lowing holds.

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

$$a'W_{zz}a \geq b_n, \quad ||W_{zz}^{-1}||_{sp} \leq b_n^{-1}, \tag{12.3}$$
$$||W_{zz}|| \leq b_{2n} = O_p(1). \tag{12.4}$$

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* $a = (a_1, ..., a_p)'$, $||a|| = 1$,

$$a'W_{zzuu}a \geq b_n, \quad ||W_{zzuu}^{-1}||_{sp} \leq b_n^{-1}, \tag{12.9}$$
$$||W_{zzuu}|| \leq b_{2n} = O_p(1), \tag{12.10}$$
$$a'D\Omega_n Da \geq b_n, \quad a'D\Omega_n Da \leq b_{2n} = O_p(1). \tag{12.11}$$

Moreover,

$$D^{-1} S_{zzuu} D^{-1} = W_{zzuu} + o_p(1), \tag{12.12}$$
$$D S_{zzuu}^{-1} D = W_{zzuu}^{-1} + o_p(1), \tag{12.13}$$
$$D^{-1} S_{zzuu}^{(c)} D^{-1} = W_{zzuu} + o_p(1), \qquad S_{zzuu}^{(c)} = \sum_{t=1}^{n} z_t z_t' E[u_t^2 \,|\mathcal{F}_{n,t-1}]. \tag{12.14}$$

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:*

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

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

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,

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

Then

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

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

$$\begin{aligned} &a'W_{zz}a = \sum_{t=1}^n a'D_g^{-1}E[z_t z_t'|\mathcal{F}_n^*]D_g^{-1}a \\ &= \sum_{t=1}^n a'D_g^{-1}E[\widetilde{z}_t\widetilde{z}_t'|\mathcal{F}_n^*]D_g^{-1}a + \sum_{t=1}^n (a'D_g^{-1}\mu_t)^2 + 2\sum_{t=1}^n (a'D_g^{-1}\mu_t)(e_t'D_g^{-1}a) \quad (12.21) \\ &= \sum_{t=1}^n a'D_g^{-1}E[\widetilde{z}_t\widetilde{z}_t'|\mathcal{F}_n^*]D_g^{-1}a + \sum_{t=1}^n (a'D_g^{-1}(\mu_t + e_t))^2 - \sum_{t=1}^n (a'D_g^{-1}e_t)^2. \quad (12.22) \end{aligned}$$

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

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

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):

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

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,

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

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

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

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):

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

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

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

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):

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

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

$$\begin{aligned} i_n &= \sum_{t=1}^n a' D_g^{-1} E[\widetilde{z}_t \widetilde{z}_t' | \mathcal{F}_n^*] D_g^{-1} a \\ &= a_1^2 \{ v_{g1}^{-2} \sum_{t=1}^n g_{1t}^2 \} + \sum_{t=1}^n \widetilde{a}' \widetilde{D}_g^{-1} \widetilde{I}_{gt} \widetilde{\Sigma} \widetilde{I}_{gt} \widetilde{D}_g^{-1} \widetilde{a} \\ &= i_{n,1} + i_{n,2}. \end{aligned}$$

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

It suffices to show that

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

We have

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

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,

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

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, $D_g^{-1}(\sum_{t=1}^{n}z_tz_t')D_g^{-1}=$

$O_p(1)$. 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,

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

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 $n \geq 1$,

$$q_{1,n} \geq b_n, \quad (12.38)$$

$$q_{2,n} \leq b_n/2. \quad (12.39)$$

Using these bounds in (12.37), we obtain

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

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

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

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),

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

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

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

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),

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

We will show that

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

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), (12.13) and (12.14). 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

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

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

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

By (12.20),

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

Then,

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

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,

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

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,

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

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. $\square$

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

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

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

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

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,

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

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

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),

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

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

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

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 $\varepsilon_{L,t}^{2+} = \varepsilon_t^2 I(\varepsilon_t^2 > L)$, we obtain

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

since, as $K \to \infty$, $L^2/K = K^{-1/2} \to 0$, $E[(\varepsilon_{L,1}^{2+})^2] \to 0$ 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), $||D^{-1} z_t u_t||^2 \leq 2b_{2t}\theta_{2t}$. Let $K > 0$ be a large number. Then,

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

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

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

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

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

$$\leq \quad \epsilon_n^{-1} K^2 O_p(1) o_p(1) + O_p(1)\Delta_K = o_p(1), \quad n, K \to \infty.$$

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

13 Additional Monte Carlo simulation results

In this section, we further evaluate the finite-sample performance of our robust OLS estimation method using two examples of regression models with fixed parameters, where the regressors z_t and regression noise u_t exhibit complex, non-standard structures.

Example 1. As in the Monte Carlo section of the main paper, we generate arrays of samples from a regression model with a fixed parameter and an intercept, using a sample size of $n = 1500$ and 1000 replications. We first consider the following model:

$$\begin{aligned} y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + u_t, \quad u_t = h_t \varepsilon_t, \\ \beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.4, 0.3)'. \end{aligned} \tag{13.1}$$

We specify the scale factor h_t in the regression noise $u_t = h_t\varepsilon_t$ as a deterministic trend $h_t = 0.4(t/n)$, and a stationary martingale difference noise ε_t is generated from a GARCH$(1,1)$ process

$$\varepsilon_t = \sigma_t e_t, \quad \sigma_t^2 = 1 + 0.7\sigma_{t-1}^2 + 0.2\varepsilon_{t-1}^2, \quad e_t \sim i.i.d.\mathcal{N}(0,1). \tag{13.2}$$

Define the regressors as $z_{1t} = 1$ and $z_{kt} = \mu_{kt} + g_{kt}\eta_{kt}$ for $k = 2, 3$, where

$$\begin{aligned} \mu_{2t} &= 0.5\sin(\pi t/n) + 1, \qquad g_{2t} = \Big|\frac{1}{2\sqrt{n}}\sum_{j=1}^{t}\nu_j\Big| + 0.25, \quad \nu_j \sim \text{i.i.d.}\mathcal{N}(0,1), \\ \mu_{3t} &= 0.5\sin(0.5\pi t/n) + 1, \qquad g_{3t} = 0.5\sin(3\pi t/n) + 1, \\ \eta_{kt} &= 0.5\eta_{k,t-1} + \xi_{kt}, \qquad \xi_{2t} = \varepsilon_{t-1}, \quad \xi_{3t} = \varepsilon_{t-2}. \end{aligned} \tag{13.3}$$

Figure 14 displays plots of a sample of variables y_t, z_t, and u_t for $t = 1, \ldots, 1500$ generated by Model (13.1)-(13.3), which exhibit clear patterns of non-stationary behavior. The Monte Carlo simulation results for sample size $n = 1500$ based on 1000 replications are reported in the Table 6. Since the regressors z_t and regression noise u_t in this model satisfy the assumptions of Corollary 2.1, as expected, the Monte Carlo simulation results confirm excellent performance of the robust OLS estimator. In particular, the empirical coverage of the 95% confidence intervals is close to the nominal 95%, whereas the standard OLS estimator exhibits significant coverage distortions.

Table 6: Robust OLS estimation in Model (13.3), $n = 1500$.

Parameters	Bias	RMSE	CP	CP$_{st}$	SD
β_1	-0.00036	0.02738	94.3	89.8	0.02738
β_2	0.00050	0.01681	93.8	79.5	0.01680
β_3	-0.00003	0.00682	95.6	85.5	0.00682

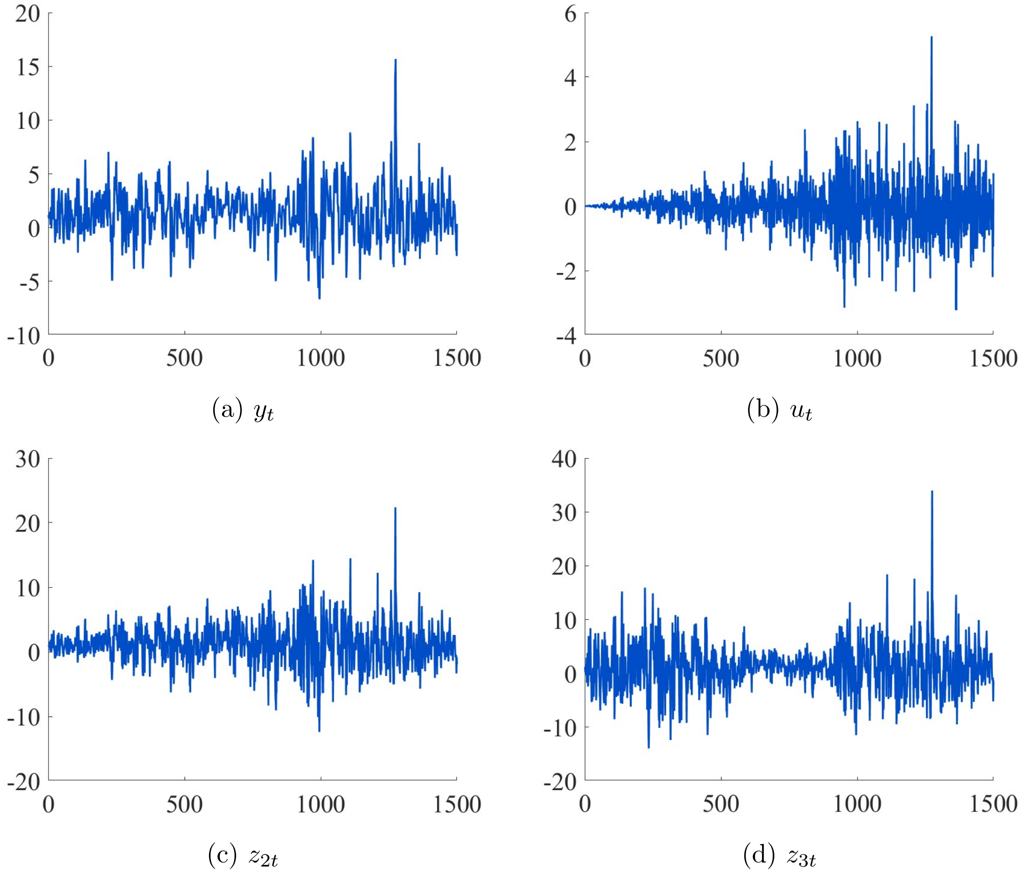


Figure 14: Plots of y_t, u_t, z_{2t} and z_{3t} in Model (13.3), $n = 1500$.

Example 2. Next, we provide an example of a regression model in which the components $\beta_1, \beta_2, \beta_3$ of the fixed regression parameter are estimated at different rates. Consider regression model (13.1) with ε_t, η_{2t}, η_{3t} defined as in Example 1. Set $h_t \equiv 1$, and let the means μ_{kt} and scale factors g_{kt}, $k = 2, 3$ be defined as follows:

$$\begin{aligned} \mu_{2t} &= [0.5\sin(10\pi t/n) + 1]\sqrt{g_{2t}}, \qquad g_{2t} = t, \\ \mu_{3t} &= [0.5\sin(5\pi t/n) + 1]\, g_{3t}, \qquad g_{3t} = t^{\gamma}, \quad \gamma = \frac{1}{2}, 0, -\frac{1}{4}, -\frac{1}{2}. \end{aligned} \tag{13.4}$$

This model satisfies the assumptions of Corollary 2.1 (see also Remark 2.1 in the main paper).

Therefore, the corresponding t-statistics for $k = 1, 2, 3$ have the following property:

$$\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{\omega}_{kk}}} \to_d \mathcal{N}(0,1), \quad \sqrt{\widehat{\omega}_{kk}} \asymp_p v_k^{-1}, \tag{13.5}$$

where, the robust standard errors $\sqrt{\widehat{\omega}_{kk}}$ are inversely proportional to the consistency rate

$$v_k = \big(\sum_{j=1}^{n} g_{jt}^2\big)^{1/2}.$$

In this model, the intercept β_1 associated with the regressor $z_{1t} = 1$ is estimated at the consistency rate $v_1 = \sqrt{n}$; the parameter β_2 linked with the regressor z_{2t} (with $g_{2t} = t$) at the rate $v_2 \sim n^{3/2}$, and the parameter β_3 linked with the regressor z_{3t} (with $g_{3t} = t^\gamma$) at the rate $v_3 \sim n^{\gamma+1/2}$. The rate v_3 is super-fast, n, when $\gamma = 1/2$; standard, $n^{1/2}$, when $\gamma = 0$; super-slow, $n^{1/4}$, when $\gamma = -1/4$; and logarithmic, $\log n$, when $\gamma = -1/2$. Monte Carlo results reported in Table 7 confirm the validity of the normal approximation (13.5) in finite samples ($n = 1500$, based on 1000 replications). In particular, the coverage of the robust 95% confidence intervals is close to the nominal level for all three parameters $\beta_1, \beta_2, \beta_t$ and for all values of γ considered in the construction of the regressor z_{3t}. In contrast, the coverage rates based on the standard OLS method exhibit noticeable distortions, especially for β_{2t} and β_{3t}.

As expected, smaller values of γ are associated with slower consistency rates v_3, wider confidence intervals, and larger standard deviations for the estimator of β_3.

Table 7: Robust OLS estimation in Model (13.4), $n = 1500$.

γ	Parameters	Bias	RMSE	CP	CP$_{st}$	SD
	β_1	-0.00331	0.08888	94.1	93.3	0.08882
1/2	β_2	2.4E-06	0.00004	95.3	86.4	0.00004
	β_3	-0.00008	0.00128	94.5	85.6	0.00128
	β_1	-0.00406	0.08976	94.5	93.4	0.08967
0	β_2	2.4E-06	0.00004	95.8	86.9	0.00004
	β_3	0.00272	0.03384	94.9	85.9	0.03373
	β_1	-0.00397	0.08884	94.6	93.9	0.08875
$-1/4$	β_2	2.6E-06	0.00004	95.6	85.9	0.00004
	β_3	0.01275	0.14219	95	87.2	0.14162
	β_1	-0.00319	0.08628	95	94.7	0.08622
$-1/2$	β_2	3.0E-06	0.00004	95.5	86.3	0.00004
	β_3	0.04468	0.43022	95.1	91.3	0.42790

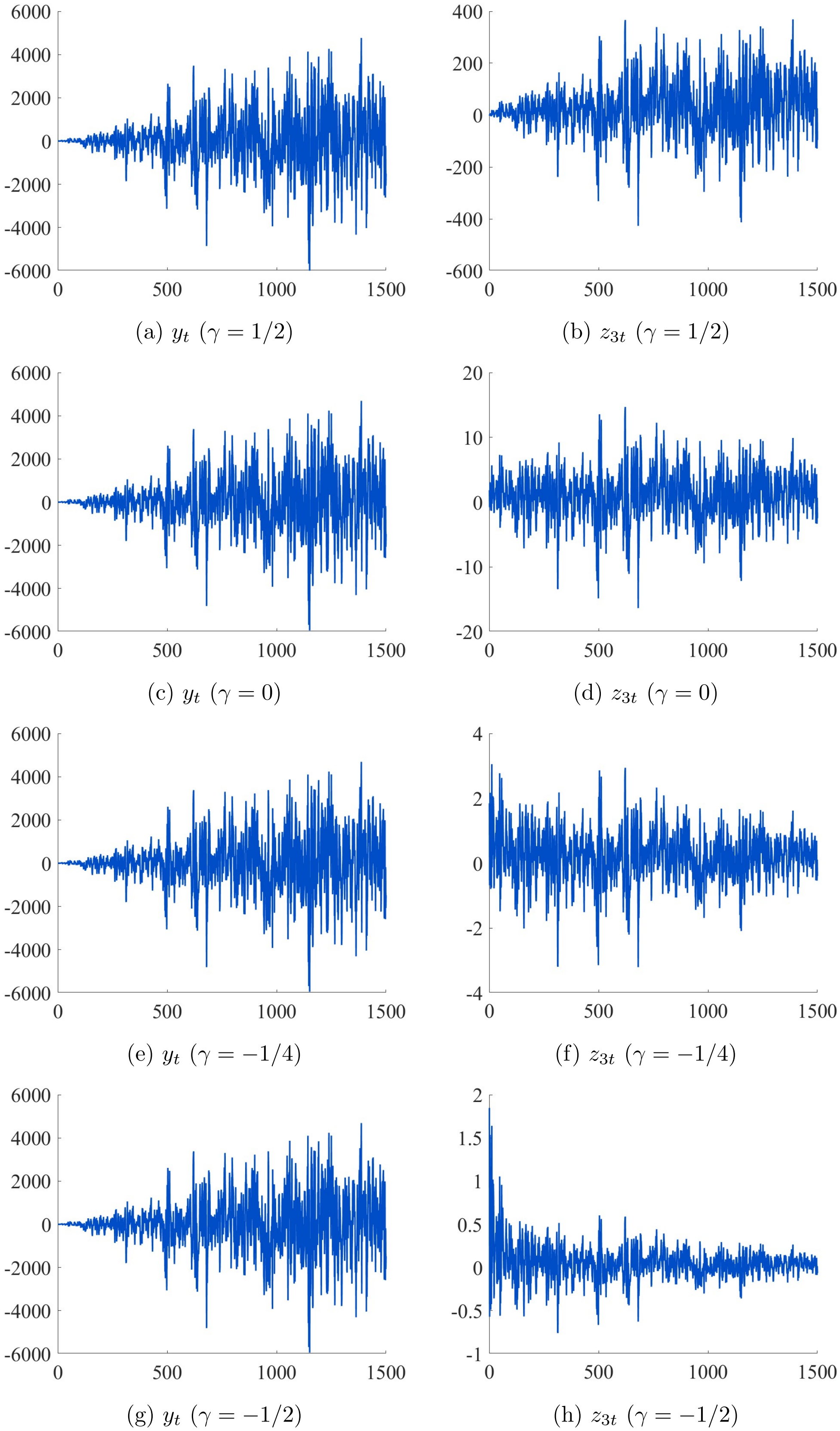
Table 8 reports the estimation results for the parameters $\beta_1, \beta_2, \beta_3$ for sample sizes $n = 200, 800, 1500, 3000$, when the regressor z_t is generated with $\gamma = -1/4$ and β_3 is estimated with the super-slow rate $v_3 = n^{1/4}$. The coverage rates for the robust OLS method are close to the nominal level in all cases. As expected, as n increases, the standard errors of all three parameter estimates decrease; however, for β_3, which is estimated with the super-slow rate

$n^{1/4}$, the reduction in the standard deviation is relatively slow.

Table 8: Robust OLS estimation in Model (13.4), $\gamma = -1/4$, $n = 1500$.

n	Parameters	Bias	RMSE	CP	CP$_{st}$	SD
	β_1	-0.01307	0.23584	94.9	95.1	0.23548
200	β_2	0.00008	0.00073	92.5	86.6	0.00073
	β_3	0.03138	0.22902	94.6	91.7	0.22686
	β_1	-0.00740	0.12187	95.1	94.6	0.12164
800	β_2	0.00001	0.00010	94.3	86.3	0.00009
	β_3	0.01302	0.16151	94.8	88.1	0.16098
	β_1	-0.00397	0.08884	94.6	93.9	0.08875
1500	β_2	2.6E-06	0.00004	95.6	85.9	0.00004
	β_3	0.01275	0.14219	95	87.2	0.14162
	β_1	-0.00319	0.06599	93.2	92.2	0.06592
3000	β_2	0.00000	0.00001	94.1	83.3	0.00001
	β_3	0.00913	0.12430	94.8	84.3	0.12396

Figure 15 displays plots of a single sample of the variables y_t and z_{3t} for $t = 1, \ldots, 1500$ generated by Model (13.4) for $\gamma = 1/2, 0, -1/4, -1/2$. These samples exhibit clear patterns of non-stationary behavior.



(a) y_t ($\gamma = 1/2$) (b) z_{3t} ($\gamma = 1/2$)

(c) y_t ($\gamma = 0$) (d) z_{3t} ($\gamma = 0$)

(e) y_t ($\gamma = -1/4$) (f) z_{3t} ($\gamma = -1/4$)

(g) y_t ($\gamma = -1/2$) (h) z_{3t} ($\gamma = -1/2$)

Figure 15: Plots of y_t, z_{3t} of a single sample of the model (13.4) for $\gamma = 1/2, 0, -1/4, -1/2$.

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.