Simultaneous Inference for Predictability with High Dimensional Mixed Roots*

Ziwei Mei
The Chinese University of Hong Kong

(JOB MARKET PAPER)

For the latest draft, click here.

December 6, 2024

Abstract

High dimensional nonstationary time series have been increasingly utilized for macroeconomic prediction, and the degrees of persistence usually vary among predictors. This paper devises a LASSO-based inference for high dimensional predictive regression with a wide class of persistent regressors, where the dimension of parameters of interest is allowed to exceed the sample size. We first establish the consistency of LASSO under mixed roots based on a new restricted eigenvalue condition of the Gram matrix. By virtue of bias-corrected LASSO estimators with the IVX instrumentation, we develop an IVX-desparsified-maximum (XDM) test that is robust to mixed degrees of persistence. Leveraging a Gaussian coupling result of martingales, we show that the XDM test asymptotically achieves the predetermined size and enjoys high power in detecting sparse alternatives. We apply the XDM test to validate the predictability of U.S. inflation with high dimensional macroeconomic data.

Key words: Local-to-Unity, Shrinkage, Gaussian Approximation, Maximum Test
JEL code: C22, C53, C55

*I am indebted to my advisors Qingliang Fan and Zhentao Shi for their constant help and guidance on my research. I am also grateful to Chunrong Ai, Jin Seo Cho, Yi Ding, Yue Fang, Xu Han, Merrick Li, Ming Li, Minghao Li, Xun Lu, Jun Yu, Xinghua Zheng, and Yang Zu for helpful comments on this paper. Ziwei Mei: zwmei@link.cuhk.edu.hk, 9/F Esther Lee Building, Department of Economics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China.

"... the ongoing interest in econometrics in uniform procedures of inference, the challenges presented by multiple predictors, and the pitfalls pointed out in the current contribution, there is substantial need for continuing econometric research on methods of inference that can cope with potential nonstationarities in many regressors, control size, and deliver good discriminatory power in detecting predictability. "

— Phillips, Peter C. B. "On confidence intervals for autoregressive roots and predictive regression." *Econometrica* 82.3 (2014): 1177-1195.

1 Introduction

The unprecedented burgeon of machine learning has reformed economic and financial studies in the modern big data era. Economists benefit from far-reaching toolkits to cope with high dimensional data, with LASSO (Tibshirani, 1996) being one of the most commonly used approaches. Machine learning extensively impacts multiple research topics of economic and financial prediction with high dimensional data, including macroeconomic forecasting (Smeekes and Wijler, 2018; Giannone et al., 2021; Medeiros et al., 2021; Babii et al., 2022; Goulet Coulombe et al., 2022), and empirical asset pricing (Feng et al., 2020; Gu et al., 2020).

Machine learning theory in regression models has been springing up for independently and identically distributed (i.i.d.) and weakly dependent data, where the number of regressors (p) is allowed to be larger than the sample size (n). Nevertheless, these theoretical results are inapplicable for highly persistent time series in macroeconomic and financial prediction. Persistent time series, though common in empirical applications, possess peculiar asymptotic properties that substantially complicate estimation and inference. The complexity is further entangled when the persistent time series are of high dimension.

This paper makes advancement in the linear predictive regression model

$$y_t = \sum_{j=1}^{p} x_{j,t-1}\beta_j^* + u_t, \ t = 1, 2, \ldots, n \tag{1}$$

under the $p \gg n$ regime. The regressors modeled as the autoregression of order one (AR(1)), given by $x_{j,t} = \rho_j^* x_{j,t-1} + e_{j,t}$, are allowed to be highly persistent. Our contributions are twofold: (i) taking *mixed root* regressors into account, and (ii) developing a procedure of simultaneous inference for *any* subset of coefficients where high dimensionality is allowed. To the best of our knowledge, this paper is the first to address these two issues when $p \gg n$.

Recent literature has witnessed theoretical progresses for the model (1) when persistent regressors are exact unit roots with the AR coefficient $\rho_j^* = 1$. In this framework, Mei and Shi (2024, MS hereafter) established the consistency of s*tandardized LASSO* (Slasso) where the regressors are normalized by their sample standard deviations (s.d.). Built upon the consistency of Slasso, Gao et al. (2024, GLMS hereafter) developed a tool of hypothesis testing called *IVX-Desparsified LASSO* (XDlasso) for a *scalar* coefficient of interest.

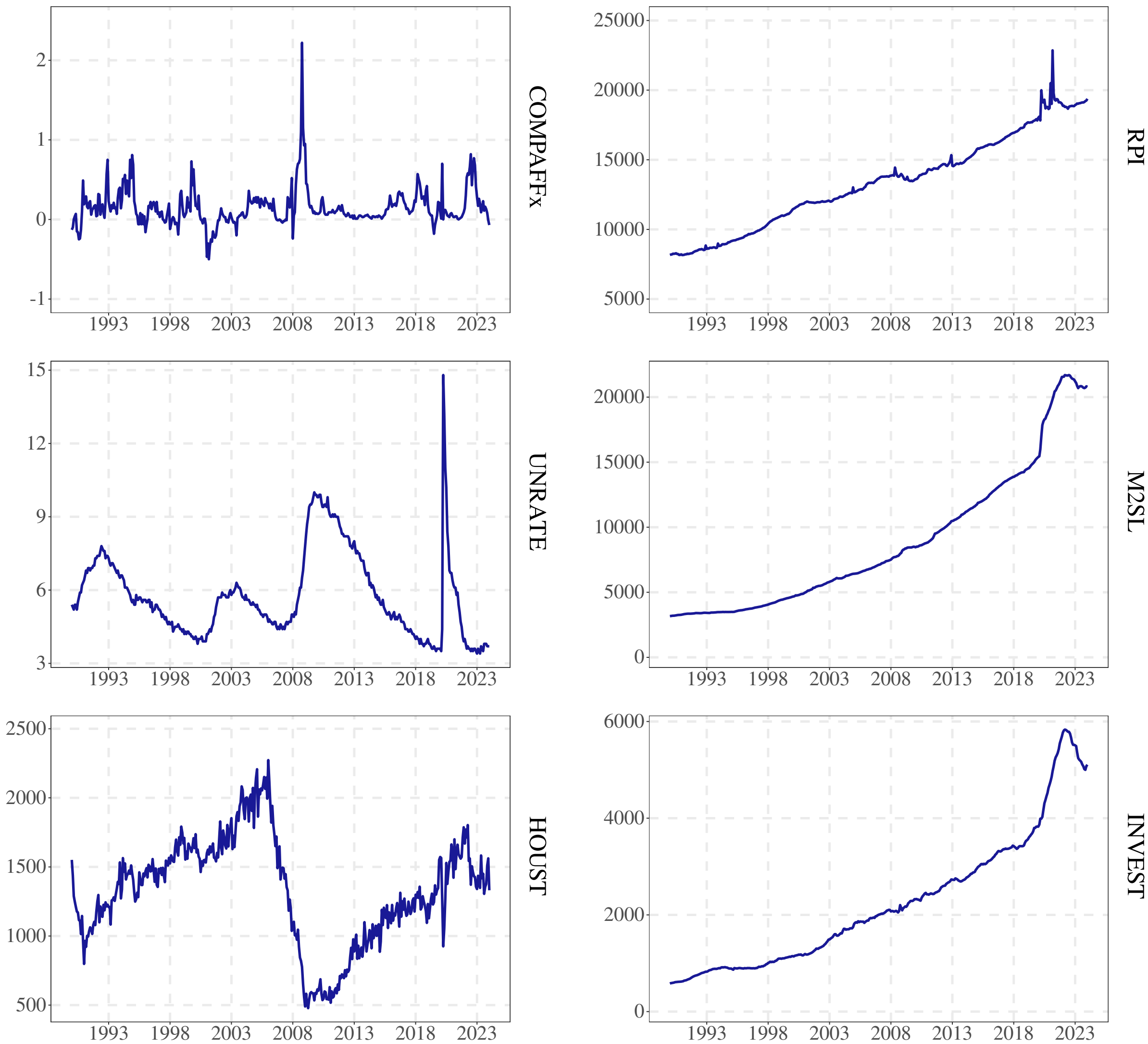
These important progresses are still insufficient to cover the practical applications. First, the unit root is not adequate to model nonstationary time series in the real world, since the persistence is usually heterogeneous among variables. The popular FRED-MD database (McCracken and Ng, 2016) of U.S. macroeconomic time series offers a typical example. Figure 1 plots representative variables from the FRED-MD database. The patterns of trends in these time series are visibly distinctive. We therefore generalize the previous unit roots by allowing ρ_j^* in the AR(1) process $x_{j,t} = \rho_j^* x_{j,t-1} + e_{j,t}$ to be close but not exactly equal to one, and heterogeneous among all j. This setup, called the *mixed root* in the literature (Phillips and Lee, 2016; Lee et al., 2022), consolidates the theory and broadens the applicability of LASSO and its variants.

Second, researchers are often interested in inference for multiple regressors with the joint null hypothesis

$$\mathbb{H}_0 : \beta_j^* = \beta_{0,j} \text{ for all } j \in \mathcal{I} \tag{2}$$

where $\mathcal{I}$ is an arbitrary subset of $\{1, 2, \ldots, p\}$ and allowed to be large. The global null hypothesis $\mathbb{H}_0 : \beta_j^* = 0$ for all $j = 1, 2, \ldots, p$ is commonly used to test the predictability of the outcome y_t by all observed predictors. Simultaneous inference for a subset of coefficients is also empirically oriented, allowing us to simultaneously test the predictability of a category of predictors. For instance, on account of the background of the Phillips Curve (Phillips, 1958), the predictability of inflation by unemployment rate is one of the central topics in macroeconomics (Stock and Watson, 1999; Groen et al., 2013). Nevertheless, the unemployment rate is just one measurement of the labor market, and FRED-MD includes other indicators. The simultaneous test provides a tool to infer if any of the high dimensional indicators of labor market has predictive power for inflation, which complements the classical wisdom. The simultaneous inference enables us to test the significance of predictive power for any subset of regressors, thereby enriching the empirical studies with high dimensional regressors from a multitude of perspectives. See Section 5 for illustrative examples.

To achieve the aforementioned two goals, we first establish the consistency of Slasso for mixed root regressors. The main challenges fall into the *restricted eigenvalue* (RE) condition for the Gram matrix of mixed roots, which is a pivotal property that commands LASSO's behavior (Bickel et al., 2009). When the AR(1) coefficient ρ_j^* converges to one as $n \to \infty$, the regressor $x_{j,t}$ is not strongly mixing and therefore the concentration inequalities for weakly dependent time series (Merlevède et al., 2011) do not apply. In particular, when $x_{j,t}$ is *local unit root* with $\rho_j^* = 1 + O(n^{-1})$, the sample s.d. of $x_{j,t}$ converges *in distribution* to



Notes: `COMPAPFFx:` 3-Month Commercial Paper Minus FEDFUNDS (Effective Federal Funds Rate); `UNRATE:` Unemployment Rate; `HOUST:` Housing Starts: Total New Privately Owned; `RPI:` Real Personal Income; `M2SL:` M2 Money Stock; `INVEST:` Securities in Bank Credit at All Commercial Banks. These time series exhibit stationary volatilities (COMPAFFx), persistent stochastic trends (UNRATE, HOUST), and explosive upward tendencies (RPI, M2SL, INVEST).

Figure 1: Representative Time Series

a peculiar random variable. Therefore, the Gram matrix does not concentrate around any deterministic matrix. The local unit root is more involved than the pure unit root with $\rho_j^* = 1$ considered in MS and GLMS, where $x_{j,t}$ has a partial sum representation $\sum_{s\leq t} e_{j,s}$ that simplifies theoretical derivations. We tackle the complexity of mixed roots and deduce the RE condition, yielding novel LASSO convergence rates distinguished from the previous works.

For simultaneous inference in high dimensional linear models, the off-the-shelf method is the maximum test by Zhang and Cheng (2017). It consists of the *desparsified LASSO* (Dlasso) that removes the shrinkage bias from LASSO penalty (Zhang and Zhang, 2014), and the multiplier bootstrap founded on Gaussian coupling theory (Chernozhukov et al., 2013; Chang et al., 2024). The theoretical foundation of Dlasso is ruined by persistent regressors, as the famous *Stambaugh bias* (Stambaugh, 1999) deviates the asymptotic distribution from normality. We thus propose a two-step testing procedure, named as the *IVX-desparsified-maximum* (XDM) test for the null hypothesis (2). First, we construct XDlasso estimators by GMLS for each coefficient of interest. The XDlasso by GLMS fuses the wisdom of the *desparsified LASSO* (Zhang and Zhang, 2014, Dlasso) that removes the shrinkage bias from LASSO penalty, and IVX (Phillips and Magdalinos, 2009) that removes the Stambaugh bias caused by high persistence. Second, we reject the null hypothesis if the maximum absolute value of the XDlasso t-statistics exceeds a critical value from the multiplier bootstrap.

Our XDM test is not a stack of ready-made components. Instead, it depends on a deep understanding of the asymptotic properties of persistent regressors, and the reconstruction of the theoretical foundation for XDlasso in the new scenario of high dimensional mixed roots. Persistent regressors invalidate Gaussian coupling theories designated for cross-sectional and weakly dependent data. Adopting the Gaussian approximation theory for martingales by Hall and Heyde (1980) and Belloni and Oliveira (2018), we exploit a high dimensional central limit theorem for XDlasso when the regressors violate the common assumption of weak dependence. We further show that the XDM test enjoys a correct asymptotic size and good discriminatory power in sparse alternatives. Similar to the classical IVX in low dimensions, the XDM test can detect local alternatives of order $1/\sqrt{n}$ for stationary regressors, and higher orders for persistent regressors.

We apply our XDM test to the FRED-MD database to infer the predictability of inflation using high dimensional macroeconomic time series. We discover that time series of output and income, labor market, money and credit, interest rate and exchange rate, prices, and stock market have significant predictive power for future inflation. The predictive power is more significant after the global financial crisis starting in late 2007. We further point out that the predictive power is strengthened by the inclusion of further lags of the predictors.

Literature review. Our work contributes to the literature of predictive regression. Non-standard asymptotic distribution caused by persistent regressors disables standard inference (Campbell and Yogo, 2006; Jansson and Moreira, 2006). Multiple solutions have been proposed for valid inference; see GLMS for a more comprehensive literature review. Phillips and Magdalinos (2009)'s IVX estimator enjoys the asymptotic normal distribution, enabling valid inference with a standard t-statistic for time series mean regressions (Kostakis et al., 2015, 2018; Phillips and Lee, 2013, 2016; Yang et al., 2020; Demetrescu et al., 2023), quantile regressions (Lee, 2016; Fan and Lee, 2019; Cai et al., 2023; Liu et al., 2023), and panel predictive regressions (Liao et al., 2024). The predictive regression also links to the recently intensive development of local projection in dynamic models (Jordà, 2005; Montiel Olea and Plagborg-Møller, 2021; Plagborg-Møller and Wolf, 2021; Mei et al., 2023; Li et al., 2024), with a main focus on weakly dependent data and low dimensions.

As Phillips (2014) pointed out in the Conclusion, econometric analysis of predictive regressions is challenging under multiple predictors, which is further complicated by the modern data-rich environment. A strand of recent works (Koo et al., 2020; Smeekes and Wijler, 2021; Phillips and Kheifets, 2024) has studied predictive regressions when the number of persistent regressors is allowed to grow with but far less than the sample size. Other papers discussed unit root test (Zhang et al., 2018), cointegration test (Onatski and Wang, 2018; Zhang et al., 2019; Bykhovskaya and Gorin, 2022, 2024), and factor analysis (Onatski and Wang, 2021), were also focused on the $p \ll n$ regime. After MS and GLMS, our paper is another stepping stone to understand predictive regressions with high dimensional nonstationary time series under $p \gg n$.

Inference by desparsified LASSO, since its inception a decade ago (Javanmard and Montanari, 2014; van de Geer et al., 2014; Zhang and Zhang, 2014), has been widely used in high dimensional models. Its popularity recently spreads to various econometric topics, including cross-sectional data (Gold et al., 2020; Fan et al., 2023, 2024) and time series regression (Chernozhukov et al., 2021; Adamek et al., 2023; Babii et al., 2024). Other machine learning techniques have been also used in various topics of time series, including cycle-trend decomposition (Phillips and Shi, 2021; Mei et al., 2024), cointegration (Lee et al., 2022), structural changes (Tu and Xie, 2023), and time-varying models (Yousuf and Ng, 2021). Recent advancements in machine learning have also witnessed applications in multiple topics of econometrics, such as instrumental variables (Belloni et al., 2012, 2014, 2022; Fan and Liao, 2014), moment equalities (Newey and Windmeijer, 2009; Cheng and Liao, 2015; Shi, 2016; Chang et al., 2018, 2021), to name a few.

Layout. The remainder of this paper is organized as follows. Section 2 instructs the high dimensional predictive regression with mixed root regressors, and establishes the consistency of Slasso. Section 3 proposes the XDM test and justifies its asymptotic size and power properties. Section 4 includes the Monte Carlo simulations to demonstrate the empirical size and power of the XDM test. Section 5 applies XDM to test predictability of the U.S. inflation. Section 6 concludes the paper. Technical proofs are neglected to Online Appendices.

Notations. We set up the notations before formal discussions. We define $\mathbf{1}\{\cdot\}$ as the indicator function, and Δ as the difference operator so that $\Delta x_t = x_t - x_{t-1}$. The set of natural numbers, integers, and real numbers are denoted as $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$, respectively. For some $n \in \mathbb{N}$, the integer set $\{1, 2, \cdots, n\}$ is denoted as $[n]$, and the space of n-dimensional real-number vectors is denoted as $\mathbb{R}^n$. For $x = (x_t)_{t\in[n]} \in \mathbb{R}^n$, the L_0-norm is $\|x\|_0 = \sum_{t=1}^n \mathbf{1}\{x_t \neq 0\}$, the L_2-norm is $\|x\|_2 = \sqrt{\sum_{t=1}^n x_t^2}$, the L_1-norm is $\|x\|_1 = \sum_{t=1}^n |x_t|$, and the sup-norm is $\|x\|_\infty = \sup_{t\in[n]} |x_t|$. Let 0_n be an $n \times 1$ zero vector, and 1_n be an $n \times 1$ vector of ones. For a generic matrix B, let B_{ij} be the (i,j)-th element, and $B^\top$ be its transpose. Let $\|B\|_\infty = \max_{i,j} |B_{ij}|$, and $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ be the minimum and maximum eigenvalues, respectively. Define $a \wedge b := \min\{a, b\}$, and $a \vee b := \max\{a, b\}$. An *absolute constant* is a positive, finite constant that is invariant with the sample size. The abbreviation "w.p.a.1" is short for "with probability approaching one". We use $\stackrel{\mathrm{p}}{\to}$ and $\stackrel{\mathrm{d}}{\to}$ to denote convergence in probability and in distribution, respectively. For any time series $\{a_t\}_{t=1}^n$, we use $\bar{a}$ to denote its sample mean $n^{-1}\sum_{t=1}^n a_t$. For any time series $\{a_t\}$ and $\{b_t\}$, we say they are *asymptotically uncorrelated* if their sample correlation coefficient $\frac{\sum_{t=1}^n (a_t-\bar{a})(b_t-\bar{b})}{\sqrt{\sum_{t=1}^n (a_t-\bar{a})^2 \sum_{t=1}^n (b_t-\bar{b})^2}} \stackrel{\mathrm{p}}{\to} 0$ as $n \to \infty$. Finally, we use "$A \otimes B$" to denote the Kronecker product of matrices A and B.

2 LASSO for High Dimensional Mixed Roots

2.1 Predictive Regression

The linear predictive regression (1) can be written as the following form:

$$y_t = x_{t-1}^\top \beta^* + u_t,\ t = 1, 2, \ldots, n, \tag{3}$$

where $x_t = (x_{1,t}, \ldots, x_{p,t})^\top$ is a p-dimensional vector of regressors, and u_t is a stationary error term.[1] We formally set up the AR(1) model for the regressors as

$$x_{j,t} = \rho_j^* x_{j,t-1} + e_{j,t},\ \rho_j^* = 1 + \frac{c_j^*}{n^{\gamma_j}} \tag{4}$$

[1]We omit the intercept for simplicity of exposition. As MS pointed out, in LASSO the intercept can be handled by the well-known Frisch-Waugh-Lovell theorem.

for $j = 1, 2, \ldots, p$, where $e_{j,t}$ is a stationary innovation. For simplicity, let the initial value $\|x_{t=0}\|_\infty = O_p(1)$. We allow for serial correlation in the weakly dependent $e_{j,t}$, thereby accommodating misspecification of the AR process (4).

The AR(1) coefficient ρ_j^* measures the persistence of $x_{j,t}$, determined by the sample size n, the real number c_j^*, and the indicator $\gamma_j = 0$ or 1. In particular, the absolute constant γ_j determines the stochastic order and asymptotic distribution of $x_{j,t}$. We call γ_j the *degree of persistence* of $x_{j,t}$. Our general framework allows for a wide range of persistence, capturing the heterogeneous dynamic patterns of high dimensional macroeconomic data. The following categories of regressors are considered:

(a) Stationary (ST): $\gamma_j = 0$ and $-2 + \epsilon < c_j^* < -\epsilon$ for some fixed small constant $\epsilon \in (0, 1)$.

(b) Locally integrated (LI): $\gamma_j = 1$ and $c_j^* < 0$.

(c) Unit root (UR): $\gamma_j = 1$ and $c_j^* = 0$.

(d) Locally explosive (LE): $\gamma_j = 1$ and $c_j^* > 0$.

The latter three cases share similarities in asymptotic properties and thus can be unified into the same category, called *local unit root* (LUR). The LUR setup subsumes persistent regressors in reality with ρ_j^* close but not exactly equal to one.

Let $\mathcal{G}^{(1)} = \{j \in [p] : \gamma_j = 1\}$ collect the integers that index the locations of LUR regressors, and $\mathcal{G}^{(0)} = \{j \in [p] : \gamma_j = 0\}$ include the indexes for stationary regressors. Define the size of each group as $p_1 = |\mathcal{G}^{(1)}|$ and $p_0 = |\mathcal{G}^{(0)}|$. Without loss of generality, suppose that $\mathcal{G}^{(1)} = \{1, 2, \ldots, p_1\}$ and $\mathcal{G}^{(0)} = \{p_1 + 1, p_1 + 2, \ldots, p\}$. For $k = 0, 1$, let $\beta^{(k)*} = (\beta_j^*)_{j \in \mathcal{G}^{(k)}}$, $x_t^{(k)} = (x_{j,t})_{j \in \mathcal{G}^{(k)}}$, and $e_t^{(k)} = (e_{j,t})_{j \in \mathcal{G}^{(k)}}$ denote the subvectors of coefficients and time series indexed by $\mathcal{G}^{(k)}$. Further define the diagonal matrix $\mathbf{C}^{(k)} = \mathrm{diag}(\{c_j^*\}_{j \in \mathcal{G}^{(k)}})$. Then the data generating processes (3) and (4) can be written as

$$y_t = x_{t-1}^{(0)\top} \beta^{(0)*} + x_{t-1}^{(1)\top} \beta^{(1)*} + u_t, \tag{5}$$

$$x_t^{(k)} = \left(I_{p_k} + \frac{\mathbf{C}^{(k)}}{n^k} \right) x_{t-1}^{(k)} + e_{t-1}^{(k)}, \ \ k = 0, 1. \tag{6}$$

The index $s = \|\beta^*\|_0$ characterizes the sparsity of the model. Throughout this paper, we take the number of regressors $p = p\,(n)$ and the sparsity index $s = s\,(n)$ as deterministic functions of the sample size n. In formal asymptotic statements, we explicitly send the sample size n to infinity, and always assume $p(n) \to \infty$ with $n \to \infty$ while $s\,(n)$ is allowed to be either fixed or divergent.

2.2 Slasso

Define the sample standard deviations (s.d.) of the regressors as

$$\widehat{\sigma}_j = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_{j,t-1} - \bar{x}_j)^2}.$$

We consider the standardized LASSO (Slasso) estimator

$$\widehat{\beta}^{\mathrm{S}} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ n^{-1}\sum_{t=1}^{n}(y_t - x_{t-1}^{\top}\beta)^2 + \lambda\|D\beta\|_1 \right\}. \tag{7}$$

where the diagonal matrix $D = \operatorname{diag}(\widehat{\sigma}_1, \widehat{\sigma}_2, \ldots, \widehat{\sigma}_p)$ stores the sample s.d. of the variables. Slasso is scale-invariant in the sense that if the regressor $x_{j,t}$ is multiplied by a nonzero constant m, the estimator changes proportionally into $\widehat{\beta}_j^{\mathrm{S}}/m$. The magnitude of mixed root regressors vary among different degrees of persistence. The standardization renders the scales of all regressors into the same stochastic order of $O_p(1)$, for which the same tuning parameter λ is valid for all regressors with various degrees of persistence. In contrast, the plain LASSO (Plasso) with the matrix D in (7) replaced by the identity matrix is expected to be inconsistent. Plasso favors the LUR regressors of a larger order, and shrinks the coefficients of stationary regressors with a smaller order all the way to zero.

The consistency of Slasso is founded on two building blocks. The first one, which is essential and challenging, is the RE condition of the Gram matrix of the standardized regressors. For any $L > 1$, the *restricted eigenvalue* of any $p \times p$ matrix Σ is defined as

$$\kappa_H(\Sigma, L, s) := \inf_{\delta \in \mathcal{R}(L,s)} \frac{\delta^{\top} H^{-1}\Sigma H^{-1}\delta}{\delta^{\top}\delta}, \tag{8}$$

where $\mathcal{R}(L, s) = \{\delta \in \mathbb{R}^p \backslash \{0_p\} : \|\delta_{\mathcal{M}^c}\|_1 \leq L\|\delta_{\mathcal{M}}\|_1, \text{ for all } |\mathcal{M}| \leq s\}$. The generic matrix H is a placeholder and varies in different contexts. Let $\widehat{\Sigma} = \sum_{t=1}^{n} x_{t-1}x_{t-1}^{\top}/n$ be the sample Gram matrix of all regressors. In the context of Slasso, we consider $\Sigma = \widehat{\Sigma}$ and $H = D$ along with the scale standardization in (7). The choice of the constant L is related to the procedures of technical proofs and does not impact the rate of convergence. Following the common practice (Bickel et al., 2009; Mei and Shi, 2024), we set $L = 3$ as a convenient choice, and simplify the notation as $\widehat{\kappa}_D = \kappa_D(\widehat{\Sigma}, 3, s)$. The quantity $\widehat{\kappa}_D$ appears at the denominator of Slasso's convergence rates, according to Lemma 1 in MS. Therefore, a rate that bounds $\widehat{\kappa}_D$ away from zero is essential for the consistency of Slasso.

The second condition for Slasso's consistency is the deviation bound (DB) of the cross-product between the error term u_t in (3) and the standardized regressors. The theoretical order of the tuning parameter λ must be no smaller than that of $\|n^{-1}\sum_{t=1}^{n} D^{-1}x_{t-1}u_t\|_\infty$ to avoid overfitting. On the other hand, an excessively large λ causes over shrinkage and damages consistency. A tight upper bound of $\|n^{-1}\sum_{t=1}^{n} D^{-1}x_{t-1}u_t\|_\infty$ is therefore indispensable.

Next, we establish the RE and DB conditions for high dimensional mixed roots, and leverage them to derive the convergence rates of Slasso.

2.3 Consistency of Slasso

The assumptions in this paper, though sharing similarities with those in MS and GLMS, are tailored to accommodate mixed roots. We first state our assumptions, and then compare them to those in GLMS. Define $e_t = (e_{1,t}, e_{2,t}, \cdots, e_{p,t})^\top$. We assume that the stationary high dimensional vector e_t is generated by a linear transformation of independent innovations $\varepsilon_t = (\varepsilon_{j,t})_{j\in[p]}$:

$$e_t = \Phi\varepsilon_t \tag{9}$$

where Φ is a $p \times p$ deterministic matrix. Let $\mathcal{F}_t$ denote the σ-field generated by $\{u_s, \varepsilon_s\}_{s\leq t}$.

Assumption 1. *Suppose that u_t and ε_t are strictly stationary and ergodic. Moreover, u_t is a martingale difference sequence (m.d.s.) such that $\mathbb{E}\,(u_t|\mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2 > 0$. There exist absolute constants C_u, b_u, C_ε, and b_ε such that for all $t \in \mathbb{Z}$ and $a > 0$,*

$$\Pr\left(|u_t| > a\right) \leq C_u \exp(-a/b_u), \tag{10}$$

$$\Pr\left(|\varepsilon_{j,t}| > a\right) \leq C_\varepsilon \exp(-a/b_\varepsilon), \tag{11}$$

for all $j \in [p]$. Furthermore, $\{\varepsilon_{j,t}\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_{\ell,t}\}_{t\in\mathbb{Z}}$ are independent for all $j \neq \ell$.

For any two σ-fields $\mathcal{A}$ and $\mathcal{B}$, define $\alpha(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}} |\Pr\left(A\cap B\right) - \Pr(A)\Pr(B)|$ and $\alpha(d) = \sup_{s\in\mathbb{Z}} \alpha(\sigma(\{u_t,\varepsilon_t\}_{t\leq s}), \sigma(\{u_t,\varepsilon_t\}_{t\geq s+d}))$.

Assumption 2. *There exists absolute constants C_α, c_α, r, and c_ε such that*

$$\alpha(d) \leq C_\alpha \exp\left(-c_\alpha d^r\right), \quad \forall d \in \mathbb{Z}, \tag{12}$$

and $\mathbb{E}\left[\sum_{d=-\infty}^{\infty} \varepsilon_{k,t}\varepsilon_{k,t-d}\right] \geq c_\varepsilon$ for any $k \in [p]$.

Define $\Omega = \Phi\Phi^\top$ where Φ has appeared in (9).

Assumption 3. *There are absolute constants c_Φ, C_Φ and C_L such that: (a) $c_\Phi \leq \lambda_{\min}(\Phi\Phi^\top) \leq \lambda_{\max}(\Phi\Phi^\top) \leq C_\Phi$; (b) $\max_{j\in[p]}\sum_{\ell=1}^{p} |\Phi_{j\ell}| \leq C_L$.*

Assumption 4. *There exist some absolute constants $\nu > 0$ and $\xi \in (0, 1/8)$ such that: (a) $p = O(n^{\nu})$; (b) $s = O(n^{1/8-\xi} \wedge p^{1-\xi})$.*

Assumption 5. *There exist absolute constants $\epsilon \in (0, 1)$ and $\bar{c} > 0$, such that the parameters $\{c_j\}_{j\in[p]}$ in (4) satisfy: (a) $c_j^* \in (-2 + \epsilon, -\epsilon)$ when $\gamma_j = 0$, (b) $c_j^* \in [-\bar{c}, \bar{c}]$ when $\gamma_j = 1$.*

Assumptions 1-3 are the same as those in GMLS. Assumption 1 imposes the exponential tails of the innovations u_t and $\varepsilon_{j,t}$. Assumption 2 characterizes the weak dependence property by the α-mixing coefficient of the stationary innovations, and imposes a lower bound on the long-run variance. Assumption 3 bounds the eigenvalues and column L_1 norm of the covariance matrix of e_t. Under the transformation (9), these bounds come into effect on both the short-run and long-run covariance matrices of e_t. These assumptions are common in the literature of high dimensional models and time series regression.

Assumption 4(a) follows GLMS to admit the growth of variable dimension p at an arbitrary fast polynomial rate of n; an exponential growth rate is allowed with additional complexity. Assumption 4(b) restricts the sparsity index by $s = o(n^{1/8})$. The sparsity condition can be relaxed to $s = o(n^{1/4})$ if we follow MS to assume that the innovations $(e_{j,t})_{j\in[p]}$ are linear processes with i.i.d. shocks and exponentially decaying coefficients, but $s = o(n^{1/8})$ is indispensable for hypothesis testing; see Assumption $4'$ in Section 3.3.

For mixed roots, we additionally impose Assumption 5 on the parameter c_j^* to demarcate different degrees of persistence. Part (a) bounds the AR coefficients of stationary regressors away from -1 and 1; Part (b) bounds $|c_j^*|$ from above for LURs, accommodating $c_j^* = 0$ for unit roots.

Remark 1. In Assumption 1, the m.d.s. assumption of u_t and the conditional homoskedasticity condition for the error term u_t of the main regression (3) are not necessary for LASSO consistency, but required for inference. See Remarks 2–4 of GLMS for justifications and possible generalizations.

2.3.1 Restricted Eigenvalue

The coexistence of stationary regressors and LURs complicates the derivations of restricted eigenvalues. Fortunately, mixed root regressors have a convenient property: LUR and stationary regressors are asymptotically uncorrelated. Therefore, after standardization by the sample s.d. of the regressors, the Gram matrix $\widehat{\Sigma}$ is approximated by the following block-diagonal matrix

$$\widehat{\Delta} = \operatorname{diag}\left(\widehat{\Sigma}^{(0)}, \widehat{\Sigma}^{(1)}\right),$$

where for $k = 0, 1$, $\widehat{\Sigma}^{(k)} = n^{-1} \sum_{t=1}^{n} x_{t-1}^{(k)} x_{t-1}^{(k)\top}$ is the $p_k \times p_k$ Gram matrix of the regressors $x_t^{(k)}$. This result is formulated as the following lemma.

Lemma 1. *Under Assumptions 1-5,*

$$\|D^{-1}\left(\widehat{\Sigma}-\widehat{\Delta}\right)D^{-1}\|_{\infty}=O_p\left(\frac{(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}}\right) \tag{13}$$

as $n\to\infty$, where r is specified in Assumption 2.

With Lemma 1, we can streamline the derivation of the RE condition for the Gram matrix $\widehat{\Sigma}$ by building RE for each diagonal block in $\widehat{\Delta}$. The RE condition for the sample Gram matrix $\widehat{\Sigma}^{(0)}$ of stationary regressors is standard. It depends on the commonly used error bound (Fan et al., 2012) of the difference between the sample Gram matrix $\widehat{\Sigma}^{(0)}$ and the population Gram matrix $\Sigma^{(0)}=\mathbb{E}(x_t^{(0)}x_t^{(0)\top})$, given as $\|\widehat{\Sigma}^{(0)}-\Sigma^{(0)}\|_{\infty}=O_p(\sqrt{\log p/n})$. Therefore, the RE of $\widehat{\Sigma}^{(0)}$ is bounded away from below if the eigenvalues of $\Sigma^{(0)}$ are bounded away from above and below.

The RE condition for LUR is in essence distinct from that of stationary regressors and emerges as the most challenging hurdle for consistency of Slasso. In low dimensions, the Gram matrix $\widehat{\Sigma}^{(1)}$, after scaled by n^{-1}, converges *in distribution* to a non-degenerate stochastic integral, given as

$$n^{-1}\widehat{\Sigma}^{(1)}\xrightarrow{\mathrm{d}}\int_0^1\mathcal{J}_{\mathbf{C}^{(1)}}(r)\mathcal{J}_{\mathbf{C}^{(1)}}(r)^{\top}\mathrm{d}r \tag{14}$$

where for a generic matrix $\mathbf{C}$ the stochastic integral $\mathcal{J}_{\mathbf{C}}(t)=\int_0^t e^{\mathbf{C}(t-r)}\mathrm{d}\mathcal{B}(r)$ is a vector of Ornstein–Uhlenbeck processes, with $\mathcal{B}(r)$ being a multivariate Brownian motion. The diagonal entries of the stochastic integral on the right-hand side of (14) are nonnegative and continuously distributed, with a non-trivial probability in a neighborhood of zero. Consequently, in striking contrast to stationary and MI regressors, the minimum diagonal entry of $n^{-1}\widehat{\Sigma}^{(1)}$ diminishes to zero as the dimension of LUR regressors passes to infinity. In the following Lemma 2, we deduce a lower bound of RE for standardized LUR.

Lemma 2. *Under Assumptions 1-5, there exists an absolute constant $c_{\kappa}^{(1)}$ such that*

$$\widehat{\kappa}_D^{(1)}\geq\frac{c_{\kappa}^{(1)}}{s(\log p)^6} \tag{15}$$

w.p.a.1 as $n\to\infty$.

The lower bound in Lemma 2 converges to zero at a sufficiently slow speed, and therefore the consistency of Slasso is still ensured. To the best of our knowledge, this is the first result of eigenanalysis for the Gram matrix of LUR under the $p>n$ regime, where the constants $\{c_j^*\}_{j\in\mathcal{G}^{(1)}}$ are allowed to be heterogeneous among all regressors. This lower bound is slightly smaller than the rate $1/s(\log p)^4$ in Proposition 3(c) of MS for exact unit roots. This slight discrepancy, on one hand, indicates that we generalize the consistency of Slasso to LUR regressors at virtually no cost in the rate of convergence. On the other hand, it sheds light on the additional complexity of LUR in contrast to exact unit roots considered in MS. The LUR regressors are not partial sums of stationary time series, which brings about additional difficulties in theoretical justifications. The proof of Lemma 2 is sketched in the following remark.

Remark 2. In this remark, we impose zero initial values to simplify the discussion. In the proof, we first assume the innovations $e_t^{(1)}$ are i.i.d. normal. The LUR process $x_t^{(1)}$ can be viewed as a partial sum of the first order difference $\Delta x_t^{(1)} = e_t^{(1)} + n^{-1}\mathbf{C}^{(1)}x_{t-1}^{(1)}$. For unit roots $\mathbf{C}^{(1)}$ is a null matrix and thus $x_t^{(1)} = \sum_{s=1}^{t} e_t^{(1)}$ is a partial sum of a normally distributed sequence. MS deduces a lower bound of RE for unit roots governed by an intermediary number $\ell = C_\ell \cdot s \log p$ for some absolute constant C_ℓ, by virtue of non-asymptotic concentration inequalities for Wishart matrices. For local unit roots with a nonzero $\mathbf{C}^{(1)}$, the term $n^{-1}\mathbf{C}^{(1)}x_{t-1}^{(1)}$ induces deviation in the aforementioned lower bound of RE, and we show that this deviation is controlled by a function of ℓ. Under a sufficiently large choice of the absolute constant C_ℓ, the lower bound from the concentration inequalities of Wishart matrices dominate the minor deviation caused by $n^{-1}\mathbf{C}^{(1)}x_{t-1}^{(1)}$, thereby maintaining the convergence rate of RE. In the last step, we leverage the Gaussian approximation result in GLMS to generalize the RE condition to non-Gaussian innovations $e_t^{(1)}$.

By the block diagonal approximation in Lemma 1, we establish the following proposition of the RE for the whole Gram matrix $\widehat{\Sigma}$.

Proposition 1. *Under Assumptions 1-5, there exists an absolute constant c_κ such that*

$$\widehat{\kappa}_D \geq \frac{c_\kappa}{s(\log p)^6}. \tag{16}$$

w.p.a.1 as $n \to \infty$.

2.3.2 Deviation Bound

Compared to RE, the DB condition for mixed roots is more straightforward and not substantially distinguished from that in MS.

Proposition 2. *Under Assumptions 1-5, there exists some absolute constant C_{DB} such that*

$$4\left\|\frac{1}{n}\sum_{t=1}^{n} D^{-1}x_{t-1}u_t\right\|_\infty \leq \frac{C_{\mathrm{DB}}(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}} \tag{17}$$

w.p.a.1 as $n \to \infty$, where r is defined in Assumption 2.

For low dimensional regressors under our assumptions, the classical law of large number (LLN) and functional central limit theorem (FCLT) for predictive regression (Magdalinos and Phillips, 2009) give

$$\frac{1}{n}\sum_{t=1}^{n}\frac{x_{j,t-1}u_t}{\widehat{\sigma}_j}=O_p(n^{-1/2}).$$

Specifically, the order of $n^{-1}\sum_{t=1}^{n}x_{j,t-1}u_t$ relates to the degree of persistence γ_j, while the standardization by $1/\widehat{\sigma}_j$ renders the convergence rate as the common order $n^{-1/2}$. The convergence rate in Proposition 2 is almost the same as the low dimensional analogy up to a logarithmic term.

2.3.3 Error Bounds

We partition the Slasso estimator into $\widehat{\beta}^{\mathrm{S}}=\left(\widehat{\beta}^{(0)\mathrm{S}\top},\widehat{\beta}^{(1)\mathrm{S}\top}\right)^{\top}$. Theorem 1 below gives the convergence rates of the Slasso estimator.

Theorem 1. *Under Assumptions 1-5, there exists a sufficiently large absolute constant C_{m}, such that when $\lambda=C_{\mathrm{m}}(\log p)^{\frac{3}{2}+\frac{1}{2r}}/\sqrt{n}$,*

$$\left\|\frac{1}{n}\sum_{t=1}^{n}x_{t-1}^{\top}(\widehat{\beta}^{\mathrm{S}}-\beta^{*})\right\|_2^2=O_p\left(\frac{s^2(\log p)^{8+\frac{1}{r}}}{n}\right), \tag{18}$$

$$\|\widehat{\beta}^{(1)\mathrm{S}}-\beta^{(1)*}\|_2+\sqrt{\frac{\log p}{n}}\|\widehat{\beta}^{(0)\mathrm{S}}-\beta^{(0)*}\|_2=O_p\left(\frac{s^{3/2}(\log p)^{6+\frac{1}{2r}}}{n}\right), \tag{19}$$

and

$$\|\widehat{\beta}^{(1)\mathrm{S}}-\beta^{(1)*}\|_1+\sqrt{\frac{\log p}{n}}\|\widehat{\beta}^{(0)\mathrm{S}}-\beta^{(0)*}\|_1=O_p\left(\frac{s^{2}(\log p)^{6+\frac{1}{2r}}}{n}\right) \tag{20}$$

as $n\to\infty$, where r is defined in Assumption 2.

Theorem 1 enriches the results in MS to characterize the LASSO convergence rates for mixed roots. The FCLT in low dimensional regression (Magdalinos and Phillips, 2009) suggests the following convergence rates of the ordinary least square (OLS) estimator

$$\widehat{\beta}^{(1)\mathrm{OLS}}-\beta^{(1)*}=O_p\left(\frac{1}{n}\right),\ \widehat{\beta}^{(0)\mathrm{OLS}}-\beta^{(0)*}=O_p\left(\frac{1}{\sqrt{n}}\right).$$

The convergence rates of LASSO are analogous to those in OLS, multiplied by an additional factor governed by the sparsity index s and a logarithmic factor. The $1/n$ in the denominators on the right-hand side of (19) and (20) yields the *super-consistency* for LUR regressors.

The Slasso estimator for the coefficients of stationary regressors follows the common $\sqrt{n}$-consistency. Lastly, the convergence rates are the same as those established in Theorem 2 of MS for pure unit roots up to a logarithmic term.

An essential difference from the results for i.i.d. data is the degree of the polynomial function of s. The well-known L_2 and L_1 convergence rates of LASSO for i.i.d. data are respectively $\sqrt{s \log p/n}$ and $s\sqrt{\log p/n}$. The additional factor s in (19) and (20) stems from the RE of LUR in Proposition B.1.

Remark 3. All theoretical orders of LASSO tuning parameters are only for technical proofs. Similar assumptions are commonly imposed in the literature of high dimensional regression (Bickel et al., 2009; Adamek et al., 2023). All numerical exercise in the current paper utilizes data-driven tuning parameter selection; see Section 4 for details.

Remark 4. Our theory does not take cointegrated regressors into account. Theorem 4 in MS shows that Slasso over-shrinks the coefficients of cointegrated regressors all the way to zero, even if their true values are nonzero. We are unaware of any regularization method that achieves estimation consistency when unit roots, stationary regressors, and cointegrated variables are mixed in the $p \gg n$ scheme. Cointegrated variables induce even more complications in the presence of mixed roots, thereby exceeding the scope of the current paper.

3 Simultaneous Inference

The maximum test, as the off-the-shelf device of simultaneous inference for high dimensional regression, depends on a bias-free estimator with an asymptotically normal distribution for each coefficient of interest. In the context of mixed roots, as introduced in Section 3.1, we utilize the XDlasso by GLMS that simultaneously removes the shrinkage bias and Stambaugh bias as a workhorse. Resorting to XDlasso, we propose the XDM test in Section 3.2, and formulate its asymptotic guarantee in Section 3.3.

3.1 IVX-Desparsified LASSO

When the regressor in a simple regression is highly persistent, OLS is known to follow a peculiar asymptotic distribution rather than normality. Therefore, OLS suffers the well-known Stambaugh bias that ruins standard inferential procedures comparing the t-statistic to the critical value from the standard normal distribution. Phillips and Magdalinos (2009)'s IVX is a powerful tool to remove the Stambaugh bias, depending on a self-generated instrument

$$z_{j,t} = \sum_{s=1}^{t} \rho_z^{t-s} \Delta x_{j,s}, \text{ and } \rho_z = 1 + \frac{c_z}{n^\theta}, \tag{21}$$

where $c_z < 0$ and $\theta \in (0,1)$ are user determined hyperparameters. In words, $z_{j,t}$ is an AR(1) process with ρ_z being the AR coefficient and the differenced regressor $\Delta x_{j,t}$ being the error term. When $x_{j,t}$ is an exact unit root with $\rho_j^* = 1$, the self-generated IV $z_{j,t}$ follows an AR(1) process $z_{j,t} = \rho_z z_{j,t-1} + e_{j,t}$ with a stationary innovation $e_{j,t}$. This is called a *mildly integrated* (MI) process with the degree of persistence $\theta \in (0,1)$, which is less persistent than an LUR. In the simple regression model, the IVX estimator of β_j^* is a two-stage least square estimator using $z_{j,t}$ as an instrument. This IVX instrumentation reduces the persistence of the regressor and recovers the asymptotically normal distribution.

In the context of high dimensional predictive regression, an auxiliary regression is in need to decorrelate the instrument $z_{j,t}$ and all other regressors $x_{-j,t} = (x_{k,t})_{k\neq j}$, for which we utilize the following XDlasso procedure by GLMS. The order of $z_{j,t}$'s sample s.d. is determined by $(\theta \wedge \gamma_j)$, which varies among the regressors. To unify the magnitude of the instrumental variable $z_{j,t}$ for each j, we standardize the instrument as

$$\widetilde{z}_{j,t} = \frac{z_{j,t}}{\widehat{\tau}_j}, \text{ where } \widehat{\tau}_j = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(z_{j,t} - \bar{z}_j)^2}. \tag{22}$$

Let $\widehat{r}_{j,t}$ be the residual from the following auxiliary Slasso regression:

$$\begin{aligned} \widehat{r}_{j,t} &= \widetilde{z}_{j,t} - x_{-j,t}\widehat{\varphi}^{[j]}, \\ \widehat{\varphi}^{[j]} &= \arg\min_{\varphi\in\mathbb{R}^{p-1}} \frac{1}{n}\sum_{t=1}^{n}(\widetilde{z}_{j,t} - x_{-j,t}\varphi)^2 + \mu_j\|D_{-j}\varphi\|_1. \end{aligned} \tag{23}$$

where $D_{-j} = \mathrm{diag}(\{\widehat{\sigma}_k\}_{k\neq j})$, and μ_j is the LASSO tuning parameter. The residual $\widehat{r}_{j,t}$ shares the same degree of persistence with the instrument $\widetilde{z}_{j,t}$, and is asymptotically uncorrelated to $x_{-j,t}$. The XDlasso estimator is given as

$$\widehat{\beta}_j^{\mathrm{XD}} = \widehat{\beta}_j^{\mathrm{S}} + \frac{\sum_{t=1}^{n}\widehat{r}_{j,t-1}(y_t - x_{t-1}^\top\widehat{\beta}^{\mathrm{S}})}{\sum_{t=1}^{n}\widehat{r}_{j,t-1}x_{j,t-1}}, \tag{24}$$

whose estimation error is decomposed as

$$\widehat{\beta}_j^{\mathrm{XD}} - \beta_j^* = \frac{\sum_{t=1}^{n}\widehat{r}_{j,t-1}u_t}{\sum_{t=1}^{n}\widehat{r}_{j,t-1}x_{j,t-1}} - \frac{\sum_{t=1}^{n}\widehat{r}_{j,t-1}x_{-j,t-1}^\top(\widehat{\beta}_{-j}^{\mathrm{S}} - \beta_{-j}^*)}{\sum_{t=1}^{n}\widehat{r}_{j,t-1}x_{j,t-1}}. \tag{25}$$

As u_t is an m.d.s., the first term is asymptotically normal since the instrument $z_{j,t}$, used to generate $\widehat{r}_{j,t}$, is less persistent than LUR. Moreover, the second term is bounded by $\|\sum_{t=1}^{n}x_{-j,t-1}\widehat{r}_{j,t-1}/\sum_{t=1}^{n}x_{j,t-1}\widehat{r}_{j,t-1}\|_\infty\|\widehat{\beta}_{-j}^{\mathrm{S}} - \beta_{-j}^*\|_1$. The Slasso estimation error $\|\widehat{\beta}_{-j}^{\mathrm{S}} - \beta_{-j}^*\|_1$

is small, based on the convergence rates we establish in Theorem 1. Furthermore, the factor $\| \sum_{t=1}^{n} x_{-j,t-1}\widehat{r}_{j,t-1} / \sum_{t=1}^{n} x_{j,t-1}\widehat{r}_{j,t-1}\|_\infty$ also has a small order, since the regressors $x_{-j,t}$ and the residual $\widehat{r}_{j,t}$ in the auxiliary LASSO regression (23) are asymptotically uncorrelated. The XDlasso estimator $\widehat{\beta}_j^{\mathrm{XD}}$ is therefore asymptotically unbiased and normally distributed.

Why is the IVX transformation necessary? When $x_{j,t}$ is LUR with high persistence, neither the Stambaugh bias nor the shrinkage bias is removed if we follow Dlasso (Zhang and Zhang, 2014) to replace $\widetilde{z}_{j,t}$ in (23) by the (standardized) regressor $x_{j,t}$. First, the regression (23) is *spurious* and thus the residual $\widehat{r}_{j,t}$ is still highly persistent. The first term on the right-hand side of (25) therefore follows a peculiar asymptotic distribution instead of normality. Second, the sample correlation coefficient of the residual $\widehat{r}_{j,t}$ from the spurious regression and the regressors $x_{j,t-1}$ does not diminish to zero, since both components follow highly persistent stochastic trends. Therefore, the second term of (25) is still not negligible. More details are available in Section 2.4 of GMLS. In the simulation studies, we compare our results to the maximum test based on Dlasso to highlight the necessity of IVX transformation.

3.2 The Maximum Test: Multiplier Bootstrap

The asymptotic covariance between $\widehat{\beta}_j^{\mathrm{XD}}$ and $\widehat{\beta}_m^{\mathrm{XD}}$ is estimated by

$$\widehat{\omega}_{j,m}^{\mathrm{XD}} = \frac{\widehat{\sigma}_u^2 \sum_{t=1}^{n} \widehat{r}_{j,t-1}\widehat{r}_{m,t-1}}{\sum_{t=1}^{n} \widehat{r}_{j,t-1}x_{j,t-1} \cdot \sum_{t=1}^{n} \widehat{r}_{m,t-1}x_{m,t-1}}, \tag{26}$$

where $\widehat{\sigma}_u^2 = n^{-1}\sum_{t=1}^{n} \widehat{u}_t^2$ with the Slasso residual $\widehat{u}_t = y_t - x_{t-1}^\top \widehat{\beta}^{\mathrm{S}}$. Under the null hypothesis (2), the t-statistic

$$\mathrm{t}_j^{\mathrm{XD}} = (\widehat{\beta}_j^{\mathrm{XD}} - \beta_{0,j})/\sqrt{\widehat{\omega}_{j,j}^{\mathrm{XD}}} \tag{27}$$

is asymptotically normal for any fixed $j \in \mathcal{I}$. The asymptotic correlation coefficient matrix of the vector of t-statistics $\mathrm{t}_{\mathcal{I}}^{\mathrm{XD}} = (\mathrm{t}_j^{\mathrm{XD}})_{j\in\mathcal{I}}$ is estimated by the positive semi-definite matrix

$$\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}} = (\widehat{\omega}_{j,m}^{\mathrm{XD}}/\sqrt{\widehat{\omega}_{j,j}^{\mathrm{XD}}\widehat{\omega}_{m,m}^{\mathrm{XD}}})_{j,m\in\mathcal{I}}. \tag{28}$$

It is worth noting that the diagonal entries of $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ are all one by construction. Under the null hypothesis (2), we expect all t-statistics in $\mathrm{t}_{\mathcal{I}}^{\mathrm{XD}}$ to be centered around zero. We therefore construct the XDM test statistic as the maximum absolute value of the t-statistics, defined as

$$\mathrm{XDM}_{\mathcal{I}} = \|\mathrm{t}_{\mathcal{I}}^{\mathrm{XD}}\|_\infty. \tag{29}$$

A hypothesis testing under significance level α is feasible by virtue of the $100(1-\alpha)$th percentile of $\mathrm{XDM}_{\mathcal{I}}$. We expect that $\mathrm{XDM}_{\mathcal{I}}$, the maximum norm of the vector $\mathrm{t}_{\mathcal{I}}^{\mathrm{XD}}$ with each entry being asymptotically normal, is well approximated by the maximum norm of the conditionally normal vector

$$\eta_{\mathcal{I}}|\mathcal{F}_n \sim \mathcal{N}\left(0_{|\mathcal{I}|}, \widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}\right), \tag{30}$$

where $\mathcal{F}_n$ is the σ-field defined above Assumption 1, and $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ defined in (28) estimates the asymptotic covariance matrix of $\mathrm{t}_{\mathcal{I}}^{\mathrm{XD}}$. We therefore utilize the $100(1-\alpha)$th percentile of $\|\eta_{\mathcal{I}}\|_\infty$

$$\mathrm{cv}_{\mathcal{I}}(\alpha) = \inf\left\{x \in \mathbb{R} : \Pr\left(\|\eta_{\mathcal{I}}\|_\infty \leq x|\mathcal{F}_n\right) \geq 1-\alpha\right\} \tag{31}$$

as the critical value. The XDM test rejects the null hypothesis (2) when

$$\mathrm{XDM}_{\mathcal{I}} > \mathrm{cv}_{\mathcal{I}}(\alpha). \tag{32}$$

Equivalently we can use the P-value

$$\mathrm{Pval}_{\mathcal{I}} = \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > \mathrm{XDM}_{\mathcal{I}}\middle|\mathcal{F}_n\right),$$

and reject the null hypothesis when $\mathrm{Pval}_{\mathcal{I}} < \alpha$.

Remark 5. The critical value (31) is infeasible. In practice we can estimate it by the multiplier bootstrap, simulating $\eta_{\mathcal{I}}^{[b]}|\mathcal{F}_n \sim \mathcal{N}\left(0_{|\mathcal{I}|}, \widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}\right)$ for $b = 1, 2, \ldots, B$, and store

$$\widehat{\mathrm{cv}}_{\mathcal{I}}(\alpha) = \inf\left\{x \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} 1(\|\eta_{\mathcal{I}}^{[b]}\|_\infty \leq x) \geq 1-\alpha\right\}. \tag{33}$$

The approximation error between the feasible $\widehat{\mathrm{cv}}_{\mathcal{I}}(\alpha)$ and the infeasible $\mathrm{cv}_{\mathcal{I}}(\alpha)$ is negligible if B is sufficiently large. We choose $B = 10000$ in all numerical exercises of the paper. Following the literature (Chernozhukov et al., 2013; Zhang and Cheng, 2017; Fan et al., 2024), we focus on the infeasible $\mathrm{cv}_{\mathcal{I}}(\alpha)$ in theoretical analysis to avoid complications. Similarly, we can also use the following bootstrapping P-value

$$\widehat{\mathrm{Pval}}_{\mathcal{I}} = \frac{1}{B}\sum_{b=1}^{B} 1\left(\|\eta_{\mathcal{I}}^{[b]}\|_\infty > \mathrm{XDM}_{\mathcal{I}}\right) \tag{34}$$

to conclude the XDM test.

We summarize the procedures of the XDM test in Algorithm 1.

Algorithm 1 (XDM Test for $\mathbb{H}_0 : \beta_j^* = \beta_{0,j}$ for all $j \in \mathcal{I}$).

Step1 Obtain $\widehat{\beta}^{\mathrm{S}}$ from the Slasso regression (7). Save the residual $\widehat{u}_t = y_t - x_{t-1}^\top \widehat{\beta}^{\mathrm{S}}$ and $\widehat{\sigma}_u^2 = n^{-1} \sum_{t=1}^n \widehat{u}_t^2$.

Step2 Obtain the IV $z_{j,t}$ by the transformation (21), and standardize it by (22).

Step3 Run the auxiliary LASSO regression (23), and save the residual $\widehat{r}_{j,t}$.

Step4 Compute the XDlasso estimator (24) for all $j \in \mathcal{I}$, and the covariance estimator $\widehat{\omega}_{j,m}^{\mathrm{XD}}$ by (26) for all $j, m \in \mathcal{I}$.

Step5 Obtain the t-statistic by (27) and generate the $\mathrm{XDM}_{\mathcal{I}}$ test statistic by (29).

Step6 For $b = 1, 2, \ldots, B$, simulate $\eta_{\mathcal{I}}^{[b]} \sim \mathcal{N}\left(0_{|\mathcal{I}|}, \widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}\right)$ with the covariance matrix $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ defined in (28). Calculate the critical value $\widehat{\mathrm{cv}}_{\mathcal{I}}(\alpha)$ by (33), or the P-value $\widehat{\mathrm{Pval}}_{\mathcal{I}}$ by (34).

Reject $\mathbb{H}_0$ under the significance level α if $\mathrm{XDM}_{\mathcal{I}} > \widehat{\mathrm{cv}}_{\mathcal{I}}(\alpha)$, or equivalently $\widehat{\mathrm{Pval}}_{\mathcal{I}} < \alpha$.

3.3 Theoretical Justifications

The asymptotic properties of the XDM test rely on the consistency of the auxiliary LASSO regression (23). We first heuristically discuss the consistency of the auxiliary regression. Note that two time series with different degrees of persistence are asymptotically uncorrelated. When $x_{j,t}$ is LUR with $\gamma_j = 1$, the instrument $z_{j,t}$ is as persistent as an MI process with a degree of persistence θ, and thus asymptotically uncorrelated to all LUR and stationary regressors. Therefore, the auxiliary LASSO estimator satisfies $\|\widehat{\varphi}^{[j]}\|_1 = o_p(1)$. When $x_{j,t}$ is stationary with $\gamma_j = 0$, the instrument $z_{j,t}$ is also stationary. Hence, the population truth of the Slasso estimator $\widehat{\varphi}_\ell^{[j]}$ is the linear projection of the (standardized) instrument $\widetilde{z}_{j,t}$ on $x_{-j,t}$.

Let $\varphi^{[j]*}$ denote the pseudo-true coefficients of the regression model (23). We neglect the complex formula of $\varphi^{[j]*}$ to Section A of the Online Appendices. The expression of $\varphi^{[j]*}$ is case-by-case, depending on the relative magnitude of θ, γ_j, and the population Gram matrix of the stationary regressors $\mathbb{E}(x_t^{(0)} x_t^{(0)\top})$.

Assumption 6. $\max_{j\in\mathcal{I}} \|\varphi^{[j]*}\|_0 \le s$ *where* $\mathcal{I}$ *indexes the regressors of interest in the null hypothesis (2), and* s *is specified in Assumption 4. Furthermore,* $\max_{j\in\mathcal{I}} \|\varphi^{[j]*}\|_1 \le C_1$ *for some absolute constant* C_1.

Assumption 6 restricts the L_0 and L_1 norm of the pseudo true coefficient $\varphi^{[j]*}$. We need sparsity assumptions on both the main regression (3) and the auxiliary regressions (23) to bound the LASSO estimation errors. The sparsity of $\varphi^{[j]*}$ can be deduced by the sparsity of the precision matrix of the stationary regressors; see Zhang and Cheng (2017) for details. We abuse the same notation s to denote the sparsity level of all regression models under consideration.

For inference, we need a sparsity condition different from that in Assumption 4(b). We modify Assumption 4 into the following.

Assumption 4′. *There exist some absolute constants $\nu > 0$ and $\xi \in (0, \frac{\theta\wedge(1-\theta)}{4})$ such that: (a) $p = O(n^{\nu})$; (b) $s = O(n^{\frac{\theta\wedge(1-\theta)}{4}-\xi} \wedge p^{1-\xi})$.*

The only modification compared to Assumption 4 is to change $s = o(n^{1/8})$ into $s = o(n^{\frac{\theta\wedge(1-\theta)}{4}})$. The sparsity index s is unknown in practice. For practical implementation, we carry on the choice of $\theta = 1/2$ in GMLS, under which the quantity $\frac{\theta\wedge(1-\theta)}{4}$ takes the maximum and therefore admits the weakest restriction on the sparsity index s. This choice is in sharp contrast to the conventional wisdom of IVX (Kostakis et al., 2015; Phillips and Lee, 2016), where θ is chosen as large as 0.95 to enhance the power of the inference. In the context of high dimensional predictive regression, an excessively large θ leads to severe size distortion unless the true coefficients satisfy the restrictive sparsity condition. As documented by Theorem 3, though the power under our recommended choice $\theta = 0.5$ is lower than that under $\theta = 0.95$, the XDM test is still consistent under a wide class of local alternatives.

Recall that λ and μ_j are the LASSO tuning parameters in the regressions (3) and (23). The following theorem formally establish the asymptotic size of the XDM test.

Theorem 2. *Suppose Assumptions 1-3, 4′, 5, and 6 hold. There exist sufficiently large constants C_{m} and $\{C_j\}_{j\in\mathcal{I}}$, such that when $\lambda = C_{\mathrm{m}}(\log p)^{\frac{5}{2}+\frac{1}{2r}}/\sqrt{n}$ and $\mu_j = C_j(\log p)^{\frac{7}{2}+\frac{1}{2r}}/\sqrt{n^{\theta\wedge(1-\theta)}}$, we have under the null hypothesis (2)*

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\mathrm{XDM}_{\mathcal{I}} < x\right) - \Pr\left(\|\eta_{\mathcal{I}}\|_{\infty} \leq x \middle| \mathcal{F}_n \right) \right| = o_p(1) \tag{35}$$

as $n \to \infty$. Therefore, $\lim_{n\to\infty} \Pr\left(\mathrm{XDM}_{\mathcal{I}} > \mathrm{cv}_{\mathcal{I}}(\alpha)\right) = \alpha$.

The asymptotic size of XDM depends on the Gaussian coupling result in (35), deduced using the Gaussian approximation theory for martingales by Hall and Heyde (1980) and Belloni and Oliveira (2018). The convergence depends on high order moments of the error term u_t in the main regression (3), the convergence rates of LASSO estimators (7) and (23), and the sup-norm estimation error of the correlation matrix estimator $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ in (28). To the best of our knowledge, this is the first Gaussian coupling result for mixed root regressors that are not weakly dependent.

Remark 6. The challenges in the deduction of the Gaussian coupling result in Theorem 2 are twofold. First, in contrast to the case of exact unit root considered in GLMS, the instrumental variable $z_{j,t}$ in (21) is not completely an MI process when $x_{j,t}$ is LUR. In particular, the instrument can be decomposed as $z_{j,t} = \zeta_{j,t} + (1-\rho_j^*)\psi_{j,t}$, where $\zeta_{j,t}$ is an MI process and $\psi_{j,t}$ is a linear combination of $(x_{j,s})_{s\le t-1}$. The additional errors caused by the remainder $\psi_{j,t}$ need careful analysis. Second, the sup-norm convergence of the correlation coefficient matrix $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ under mixed roots is more complicated than that under i.i.d. or stationary data, for which the approximation error between the sample Gram matrix of MI processes and a peculiar deterministic matrix of an integral form is involved.

For power analysis, we consider the following alternative set

$$\mathcal{U}_{\mathcal{I}}(c) = \left\{ \beta = (\beta_1, \beta_2, \dots, \beta_p)^\top : \max_{j\in\mathcal{I}} \left| \sqrt{n^{1+(\theta\wedge\gamma_j)}}(\beta_j - \beta_{0,j}) \right| > c \log|\mathcal{I}| \right\} \tag{36}$$

for a generic constant $c > 0$. Distinguished from the alternative set used for cross-sectional data in Zhang and Cheng (2017), an additional factor $\sqrt{n^{(\theta\wedge\gamma_j)}}$ appears in (36) due to faster convergence rates for persistent regressors. Moreover, the lower bound in (36) is $\log|\mathcal{I}|$ rather than $\sqrt{\log|\mathcal{I}|}$ as in Zhang and Cheng (2017). When $x_{j,t}$ is LUR, the standard error of XDlasso $\widehat{\omega}_{j,j}^{\mathrm{XD}}$ defined in (26), multiplied by $\sqrt{n^{1+\theta}}$, converges *in distribution* to a peculiar random variable. The additional $\sqrt{\log|\mathcal{I}|}$ factor sources from the randomness introduced by the standard error $\widehat{\omega}_{j,j}^{\mathrm{XD}}$ of LUR regressors.

The following theorem characterizes the power property of the XDM test.

Theorem 3. *Suppose that $|\mathcal{I}| \to \infty$ as $n \to \infty$, and the conditions in Theorem 2 hold. For any absolute constant c_0, whenenver $\beta^* \in \mathcal{U}_{\mathcal{I}}(c_0)$ we have*

$$\lim_{n\to\infty} \Pr\left(\mathrm{XDM}_{\mathcal{I}} > \mathrm{cv}_{\mathcal{I}}(\alpha)\right) \to 1.$$

The condition $|\mathcal{I}| \to \infty$ is merely for simplicity. We allow for a fixed $|\mathcal{I}|$, under which the $\log|\mathcal{I}|$ in the alternative set (36) is replaced by any sequence that passes to infinity as $n \to \infty$. Theorem 3 indicates that the XDM test rejects the null hypothesis w.p.a.1 even when only one true coefficient locally deviates from the null hypothesis, thereby sensitively detecting sparse alternatives.

Consider a simple case where the scalar β_j^* deviates from the null hypothesis $\mathbb{H}_0 : \beta_j^* = 0$. By (36), for LURs with $\gamma_j = 1$ the XDM test is consistent under the alternative $\beta_j^* = O(1/n^{\delta_j})$ for any $\delta_j \in (0, \frac{1+\theta}{2})$. The range always includes the important rate $1/\sqrt{n}$, under which $\mathrm{var}(x_{j,t}\beta_j^*) = O(1)$, balancing the larger order for a persistent $x_{j,t}$ with $\gamma_j > 0$ and a standard order $O_p(1)$ for y_t. Like all other standard inference methods, XDM achieves the

$\sqrt{n}$-consistency for a stationary $x_{j,t}$ with $\gamma_j = 0$.

Our theoretical statements have again witnessed the sharp distinctions of mixed roots from the traditional i.i.d. and stationary setup. Some mild differences between the current results and those in MS, including the slightly faster convergence rate of RE in Lemma 2, highlight the additional complexity from the vicinity of unity. With mild costs in rates of convergence, we establish positive results of Slasso and XDlasso under mixed roots, broadening their usabilities for high dimensional macroeconomic and financial time series. The XDM test is also original in that the Gaussian coupling theory for persistent regressors has not been studied before.

4 Simulations

4.1 Setup

We consider the linear predictive regression model (3), where the p-dimensional predictors x_t follow the AR(1) processes (4). We separate the regressors into 4 groups, where the AR coefficient ρ_j^* is heterogeneous among different groups. Specifically, each group has $p_0 = p/4$ regressors, and the AR(1) coefficients $\rho^* = (\rho_1^*, \rho_2^*, \ldots, \rho_p^*)^\top$ satisfies

$$\rho^* = \varrho_0^* \otimes 1_{p_0}, \;\; \varrho_0^* = \left(0.5, \;\; 1 - \frac{1}{n}, \;\; 0, \;\; 1 + \frac{1}{n}\right)^\top. \tag{37}$$

Specifically, $x_{j,t}$ is stationary for $1 \leq j \leq p_0$, and LUR for $p_0 + 1 \leq j \leq 4p_0$ including LI, UR, and LE regressors. The coefficients of the predictive regression (3) are configured as

$$\beta^* = 1_4 \otimes (b_0, 0_{p_0-1}^\top)^\top. \tag{38}$$

In other words, the coefficient accompanied with the first regressor in each group is b_0, with all other coefficients being zero. In the simulation studies, we vary the value of b_0 to investigate the size and power of the tests. The innovations $(u_t, e_t^\top)^\top$ are i.i.d. samples from a multivariate normal distribution with mean zero and a covariance matrix $(0.5^{|i-j|})_{i,j\in[p+1]}$. We consider the sample sizes $n \in \{180, 240, 360\}$ to mimic 15-year, 20-year, and 30-year monthly data, and the variable dimensions $p_0 = \{30, 50, 70\}$, under which $p = 4p_0 = \{120, 200, 280\}$. We conduct 1000 replications for each setting.

Tuning parameters. The LASSO tuning parameter for the main regression (3) is chosen

Table 1: Size of the XDM and DM tests

p	n	(a) XDM			(b) DM		
		All	ST	LUR	All	ST	LUR
140	180	0.040	0.053	0.043	0.206	0.053	0.219
	240	0.062	0.051	0.067	0.250	0.046	0.269
	360	0.054	0.052	0.055	0.342	0.045	0.379
210	180	0.042	0.049	0.040	0.168	0.041	0.194
	240	0.037	0.049	0.031	0.257	0.044	0.279
	360	0.055	0.052	0.047	0.362	0.055	0.396
280	180	0.039	0.046	0.039	0.196	0.054	0.208
	240	0.032	0.038	0.035	0.215	0.038	0.252
	360	0.050	0.052	0.052	0.370	0.046	0.403

by minimizing the following BIC criterion (Ahrens et al., 2020)

$$\lambda = \arg\min_{\lambda} \; \log\left[\sum_{t=1}^{n}(y_t - x_{t-1}^{\top}\widehat{\beta}(\lambda))^2\right] + \|\widehat{\beta}(\lambda)\|_0 \cdot \log n,$$

where $\widehat{\beta}(\lambda)$ denotes the LASSO estimator under the tuning parameter λ. The tuning parameter for the auxiliary LASSO regression (23) is also chosen in a parallel way. The hyperparameters for the self-generated instrument carry on the choices in GLMS $c_z = -5$ and $\theta = 0.5$. As illustrated below Assumption 4′, the choice $\theta = 0.5$ allows for the weakest condition of sparsity.

4.2 Size

We consider the following null hypotheses:

(a) $\mathbb{H}_0^{\mathrm{All}} : \beta_1^* = \beta_2^* = \cdots = \beta_p^* = 0,$

(b) $\mathbb{H}_0^{\mathrm{ST}} : \beta_1^* = \beta_2^* = \cdots = \beta_{p_0}^* = 0,$

(c) $\mathbb{H}_0^{\mathrm{LUR}} : \beta_{p_0+1}^* = \beta_{p_0+2}^* = \cdots = \beta_p^* = 0.$

The first item is the global null hypothesis. The last two hypotheses are for different categories of regressors according to their degrees of persistence. Though the maximum test by Zhang and Cheng (2017) based on Dlasso (Zhang and Zhang, 2014) is not designed for persistent regressors, we compare XDM test with it to highlight the necessity of IVX transformation under mixed roots. We abbreviate it as the Dlasso maximum (DM) test for convenience. All tests are carried out under 5% significance level.

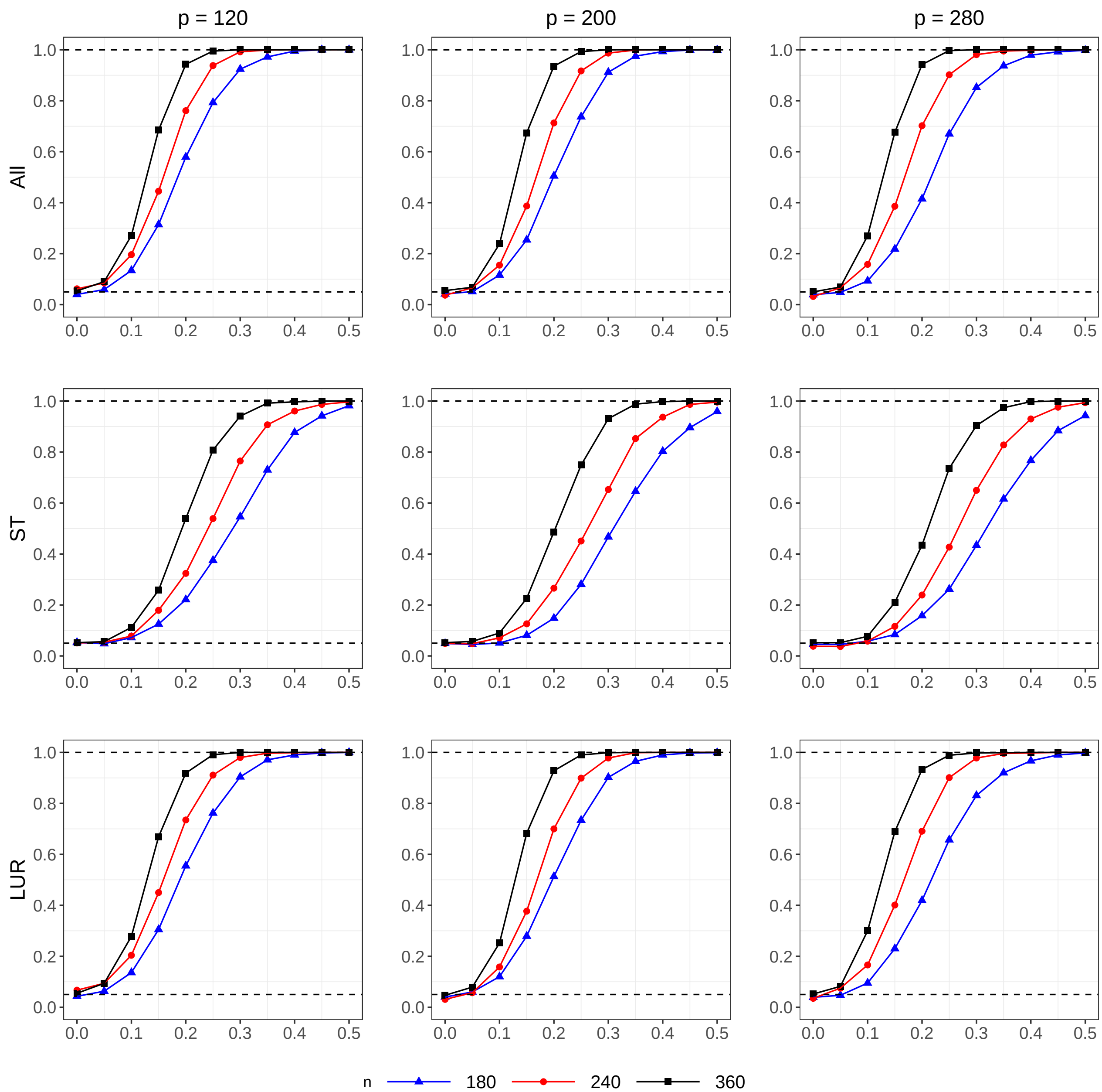
Table 1(a) summarizes the empirical size of the XDM test for the null hypotheses in the itemized list. The primary finding is that XDM effectively controls the empirical size around the nominal level. This conclusion is congruent among different sample sizes, variable dimensions, and regressors of interest. In sharp contrast, as displayed in Table 1(b), the empirical size of the global test by DM severely distorts from the nominal size. Scrutinizing the DM test for each category of regressors, we observe that the size distortion is caused by the LUR regressors, highlighting the necessity IVX transformation to remove the Stambaugh bias.

4.3 Power

To examine the power of the XDM test, we vary b_0 from 0 to 0.5 and report the rejection rate for the null hypotheses $\mathbb{H}_0^{\mathrm{All}}$, $\mathbb{H}_0^{\mathrm{ST}}$, and $\mathbb{H}_0^{\mathrm{LUR}}$. Each graph in Figure 2 includes power curves under various n's for one variable dimension p and one null hypothesis. The XDM test exhibits increasingly high power as the sample size n increases and the true coefficients shift away from the null hypothesis.

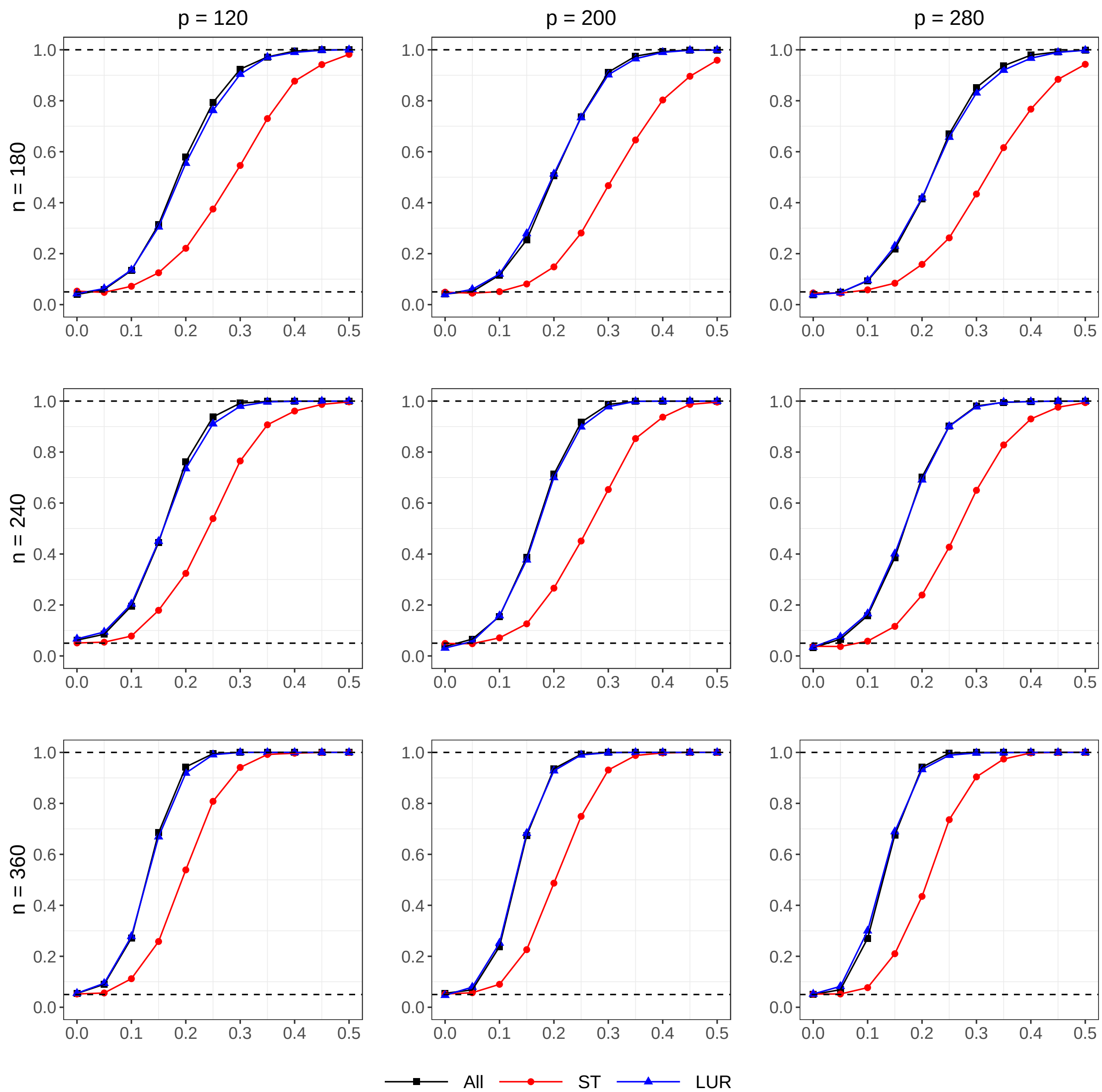
To corroborate with Theorem 3, we reorganize the power curves by the category of regressors of interest in different null hypotheses. In Figure 3, each graph sketches the power curves for hypotheses $\mathbb{H}_0^{\mathrm{All}}$, $\mathbb{H}_0^{\mathrm{ST}}$, and $\mathbb{H}_0^{\mathrm{LUR}}$ under each pair of sample size and variable dimension (n, p). The test for $\mathbb{H}_0^{\mathrm{ST}}$ is evidently less powerful than that for $\mathbb{H}_0^{\mathrm{LUR}}$. This finding echoes the classical theory of predictive regression that estimation and inference for more persistent regressors enjoy faster rates of convergence. Nevertheless, there is almost no distinctions among the tests for $\mathbb{H}_0^{\mathrm{All}}$ and $\mathbb{H}_0^{\mathrm{LUR}}$. This phenomenon concerts with Theorem 3 that the order of XDM test's power is determined by the indicator $\max_{j\in\mathcal{I}}(\theta \wedge \gamma_j)$. In our practice, this indicator equals 0.5 across these two configurations of null hypotheses. Therefore, the power of XDM tests for these two hypotheses share the same theoretical order.

This simulation exercise demonstrates the desirable size and power properties of the XDM test under finite sample. The results not only support the theoretical statements in Section 3.3, but also evidence the credibility of XDM's practical application.



Notes: The x-axis represents the value of b_0 in (38). The horizontal dashed lines mark the significance level 0.05 and power 1.

Figure 2: Power Curves by Sample Size



Notes: The x-axis represents the value of b_0 in (38). The horizontal dashed lines mark the significance level 0.05 and power 1.

Figure 3: Power Curves by Null Hypothesis

5 Predictability of Inflation

Predictability of inflation is one of the central topics in macroeconomic prediction, and has provoked a long debate in academia (Stock and Watson, 2003). The sharp decline and spike of inflation caused by the recent Covid-19 shock in developed economies, including the United States, pose challenges to inflation modeling and forecasting (Ball et al., 2022; Bobeica and Hartwig, 2023). A surging strand of literature (Medeiros et al., 2021; Aliaj et al., 2023; Hong et al., 2024) has surfaced the improvement of inflation forecasting by machine learning under the data-rich environment. The benefits of machine learning to real-time tracking of inflation has also caught attention from policymakers; see Knotek and Zaman (2024) for a review. This section applies the XDM test to investigate the predictability of inflation using high dimensional macroeconomic predictors. Our novel testing method for high dimensional coefficients unpacks new insights and more thorough understanding of this topic, which were unexplored due to the lack of feasible toolkits. Throughout this section, we focus on $\beta_{0,j} = 0$ in the null hypothesis (2).

5.1 Data

We utilize the monthly time series from the FRED-MD database (McCracken and Ng, 2016), with the sample period from January 1990 to January 2024. The inflation rate of our central interest is calculated by $\pi_t = \Delta \log(P_t) \times 100$, where P_t denotes the *Consumer Price Index for All Urban Consumers: All Items (CPI)* from the FRED-MD dataset. We focus on the linear predictive regression $\pi_t = \beta_0^* + x_{t-1}^\top \beta^* + u_t$, and apply XDM to test the joint significance of the coefficients β^* and its subvectors. Predictors in x_t include other 121 time series after removing those with missing values during the sample period. Figure 4 plots the inflation rate. Graphically the time series of inflation rate is stationary, which is also supported by statistical evidence of the AR(1) coefficient 0.467 and the Augmented Dickey-Fuller (ADF) test P-value 0.010.

In macroeconomics, the global financial crisis (GFC) around 2008 has been an important time node. The pre-GFC period within our sample (from January 1990 to November 2007) falls in the Great Moderation of the United States coined by Stock and Watson (2002), characterizing the mitigation of the volatility in business cycle fluctuations. In contrast, the post-GFC period is featured by the Great Recessions during the financial crises in economies around the world from late-2007 to mid-2009, and the Covid-19 recession starting in 2020. These recessions and the recoveries that follow result in greater volatilities in the economies compared to the pre-GFC period. We thus highlight the pre-GFC and the post-GFC periods using different colors in Figure 4, and carry out empirical studies using these two subsamples respectively additional to the full-sample analysis. Visually, the inflation rate is more volatile after the GFC, especially during the two aforementioned Great Recessions and their recoveries, hinting the heterogeneity in its predictability between these two subsamples.

Figure 5 portrays the persistence of the 121 predictors. Panel (a) is a histogram of the AR(1) coefficients. An overwhelming majority of the AR(1) coefficients concentrate around 1, indicating high degrees of persistence. Meanwhile, the AR coefficients range from 0.38 to 1.01, highlighting the necessity to consider mixed roots. Panel (b) depicts the distribution of P-values of the Augmented Dickey–Fuller (ADF) test for each predictor. Only a small portion of ADF tests reject the null hypothesis under the 10% level. These summary statistics exhibit the nature of nonstationarity of the predictors, thereby motivating our XDM test for persistent regressors.

5.2 Empirical Findings

The first problem of interest is the global significance of all coefficients in the predictive model. Beyond that, the FRED-MD database categorizes all macroeconomic time series into 8 groups by economic implications, including (1) output and income, (2) labor market, (3) consumption and housing, (4) orders and inventories, (5) money and credit, (6) interest rate and exchange rate, (7) prices, and (8) stock market. The CPI indicator that generates our outcome variable π_t belongs to the "price" group. We thus respond to a more profound question: which groups of predictors have significant predictive power for future inflation?

The P-values reported by Table 2(a) show the answer. We first focus on the results by the XDM test until further clarifications. The global null hypothesis is always unambiguously rejected by XDM no matter for the full sample or the two subsamples, indicating that statistical testing strongly evidences the predictability of inflation. In terms of the groups by economic implications, "output and income" and "prices" also have strongly significant predictive power for inflation across all considered sample periods. The former echos the dynamic correlation between output and inflation extensively documented in macroeconomic literature (Lucas, 1973; Galı and Gertler, 1999; Burstein, 2006). The significance of price indicators' predictive power is unsurprising; these indicators, in the same group as CPI, have strong economic relevance to inflation.

Macro-finance indicators, including "money and credit", "interest rate and exchange rate", and "stock market", show significant predictive power to future inflation under 10% level in the full sample and the post-GFC period, but not the pre-GFC period. This discrepancy among different sample periods results from the sharp contrast between the Great Moderation in the pre-GFC period and the Great Recessions in the post-GFC period, since during the latter period the inflation is more volatile and responsive to financial and monetary shocks. As these monetary and financial indicators instantly respond to monetary policies, we conjecture that this discrepancy is also a consequence of the increasing attention of the Federal Reserve on inflation targeting after the GFC. [2]

The predictive power of "labor market" variables is statistically significant under 10% level for the full sample, exhibiting a silver lining of the Phillips Curve. Nevertheless, the predictive power is not significant for the subsamples. There is no significant evidence from the data for the predictive power of other economic indicators like "consumption and housing" and "orders and inventories".

Whilst it is a common practice to transform the regressors into stationary time series, recent empirical applications of machine learning (Smeekes and Wijler, 2018, 2021; Mei and Shi, 2024) have illuminated the benefits of nonstationary data in prediction. In FRED-MD, each variable is accompanied with a transformation code (TCODE) denoting the elementary transformation to stationarize the time series. Table 2(b) reports the results using transformed variables according to the TCODEs. Compared to the results using raw data in Table 2(a), stationarizing the regressors undermines the statistical significance of predictive power for the full set of regressors, and the essential macroeconomic indicators like outputs and prices. The impairment of predictive power in the pre-GFC period is the most apparent. These results evidence that transforming the predictors into stationary time series prior to prediction is not a silver bullet.

Recent empirical studies (Medeiros et al., 2021; Mei and Shi, 2024) have substantiated the improvement of prediction with the lagged outcome and more lags of predictors included in the model. To further evidence this argument, we consider $\pi_t = \beta_0^* + \sum_{h=1}^4 \pi_{t-h}\beta_{\pi,h}^* + \sum_{h=1}^4 x_{t-h}^\top \beta_{x,h}^* + u_t$ and test the significance the coefficients of $\beta_h^* = (\beta_{\pi,h}^*, \beta_{x,h}^{*\top})^\top$ for $h = 1, 2, 3, 4$. Table 3 displays the testing results. The global null hypothesis is again indisputably rejected by the XDM test. Also, the predictive power of the first order lag is strongly evidenced, echoing the results from the one-period predictive regression. The predictive power of the second order lag is also statistically significant using the full sample and the pre-GFC period, but not the post-GFC period. In terms of the third and the fourth order lags, the predictive power is not supported by XDM. These results exhibit that lags of predictors can improve predictive power for future inflation, but the improvement mitigates as the lag order increases.

We finally turn to the DM test without IVX transformation in Tables 2(a) and 3 with highly persistent regressors. Across all configurations, the DM test exhibits stronger statis-

[2] As remarked by Cúrdia (2022) in the Conclusion, "In 2020 the Federal Reserve incorporated average inflation targeting into its policy framework ... an AIT strategy that responds to both current and expected future inflation could have achieved substantially better economic outcomes and reduced uncertainty in the aftermath of the Global Financial Crisis. "

tical significance than the XDM test. The sharpest contrasts occur in the "consumption and housing" and "orders and inventories" categories in Table 2(a), and the lags of the second and third orders in Table 3. As we have documented, the DM test suffers severe size distortion when regressors of interest are highly persistent. Therefore, the discrepancy between DM and XDM tests is ascribed to the spurious significance caused by the bias in the DM test. This phenomenon again sheds light on the need for IVX transformation in our XDM test to remove the Stambaugh bias.

6 Conclusion

This paper proposes the XDM test for simultaneous inference of high dimensional predictive regression with mixed roots. As a preparatory theoretical foundation, we first reestablish the consistency of the workhorse Slasso estimator by the essential RE and DB conditions. We then leverage XDlasso to remove both the shrinkage bias from LASSO penalty and the Stambaugh bias from nonstationary regressors. The XDM test rejects the null hypothesis if the maximum absolute value of the XDlasso t-statistic is larger than a critical value from a high dimensional central limit theorem for persistent regressors. The validity of our testing method is further supported by simulations and empirical applications.

In addition to the maximum test, the literature has also developed sum-type tests (Chen et al., 2024) and L_p-norm-based tests (Kock and Preinerstorfer, 2023, 2024) for dense alternatives. It is unclear how to generalize these methods to nonstationary data. These extensions deserve explorations in the future.

While mixed root is a popular model of nonstationary time series due to its mature asymptotic theories, it is still insufficient to cover all peculiar characteristics of macroeconomic and financial time series in the real world. Topics with profound influence in low dimensional nonstationary time series include cointegration (Phillips, 1991), long memory (Shi and Yu, 2023), bubbles (Phillips et al., 2015b,a), and so forth. It is unclear how to extend these models to high dimensions. Encouraged by the preface in our paper quoted from Phillips (2014), we are still on the way toward the terminal of predictive regression with many nonstationary regressors.

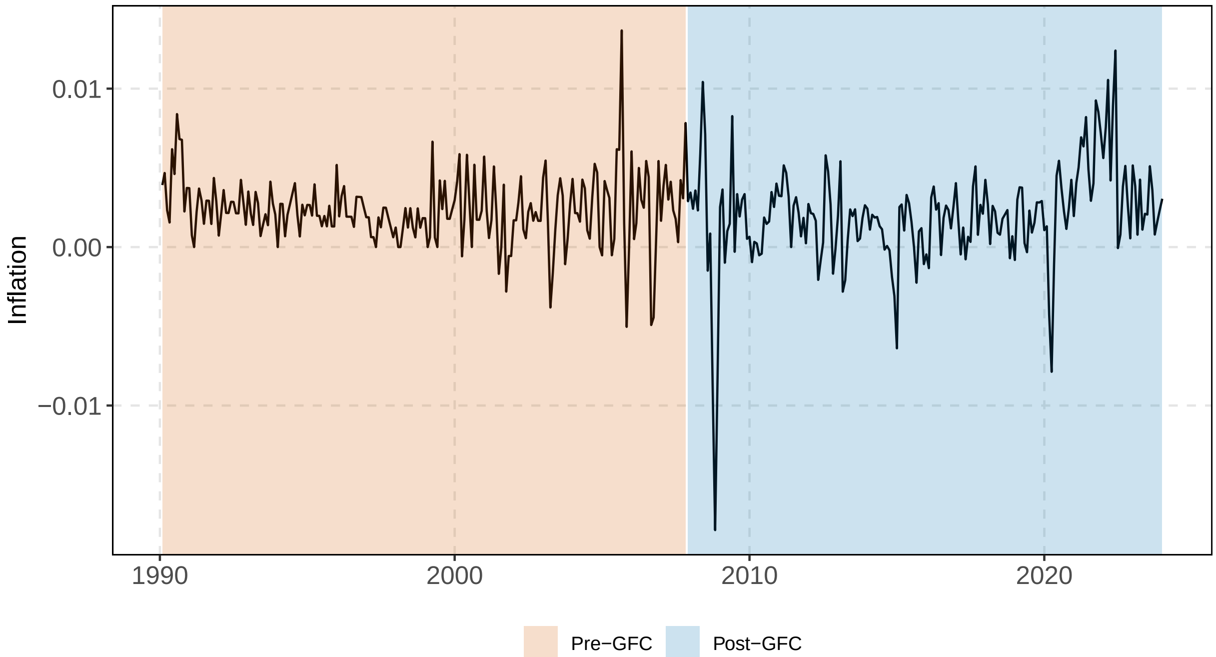
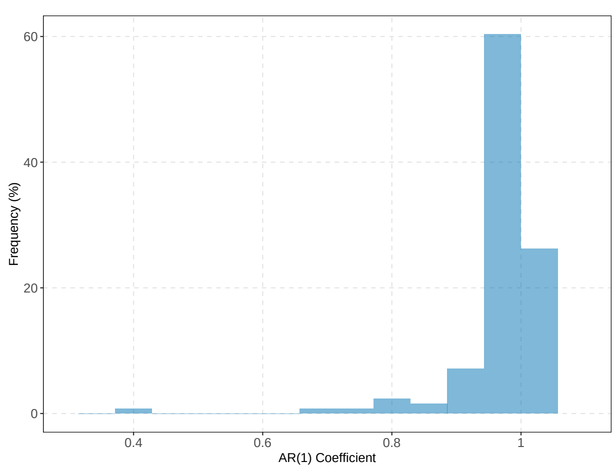
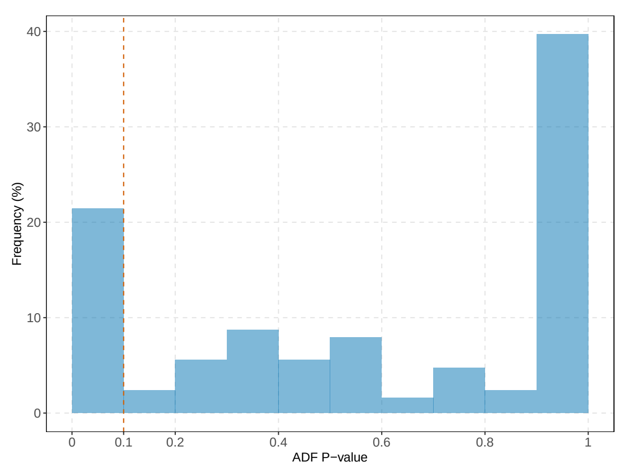


Figure 4: Inflation Rate



(a) AR(1) Coefficients



(b) ADF P-values

Notes: The vertical dashed line in Panel (b) mark the significance level 0.1.

Figure 5: Histograms of Summary Statistics of Predictors

Table 2: P-values of the XDM and DM Tests

(a) Raw Data

Groups	XDM			DM		
	Full	preGFC	postGFC	Full	preGFC	postGFC
All	0.000	0.000	0.000	0.000	0.000	0.000
Output and Income	0.041	0.000	0.017	0.008	0.001	0.000
Labor Market	0.058	0.877	0.311	0.013	0.817	0.000
Consumption and Housing	0.170	0.315	0.944	0.091	0.694	0.083
Orders and Inventories	0.437	0.158	0.723	0.559	0.008	0.026
Money and Credit	0.029	0.797	0.097	0.015	0.234	0.049
Interest Rate and Exchange Rate	0.095	0.644	0.072	0.031	0.675	0.000
Prices	0.000	0.000	0.000	0.000	0.000	0.000
Stock Market	0.000	0.789	0.000	0.000	0.165	0.000

(b) Stationarized Predictors

Groups	XDM			DM		
	Full	preGFC	postGFC	Full	preGFC	postGFC
All	0.001	0.244	0.032	0.000	0.312	0.000
Output and Income	0.056	0.162	0.464	0.014	0.176	0.195
Labor Market	0.028	0.469	0.801	0.111	0.338	0.176
Consumption and Housing	0.865	0.810	0.242	0.017	0.111	0.116
Orders and Inventories	0.000	0.222	0.007	0.000	0.264	0.000
Money and Credit	0.179	0.033	0.138	0.342	0.043	0.674
Interest Rate and Exchange Rate	0.216	0.747	0.006	0.022	0.516	0.022
Prices	0.043	0.226	0.050	0.053	0.078	0.084
Stock Market	0.008	0.663	0.001	0.146	0.482	0.787

Notes:The P-values are calculated by Equation (34).

Table 3: P-values of the XDM and DM Tests with More Lagged Predictors

Lags	XDM			DM		
	Full	preGFC	postGFC	Full	preGFC	postGFC
All	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.002	0.007	0.349	0.000	0.000	0.022
3	0.257	0.217	0.486	0.355	0.819	0.008
4	0.561	0.542	0.087	0.224	0.197	0.581

Notes: The P-values are calculated by Equation (34).

References

Adamek, R., S. Smeekes, and I. Wilms (2023). Lasso inference for high-dimensional time series. *Journal of Econometrics 235*(2), 1114–1143.

Ahrens, A., C. B. Hansen, and M. E. Schaffer (2020). lassopack: Model selection and prediction with regularized regression in stata. *The Stata Journal 20*(1), 176–235.

Aliaj, T., M. Ciganovic, and M. Tancioni (2023). Nowcasting inflation with lasso-regularized vector autoregressions and mixed frequency data. *Journal of Forecasting 42*(3), 464–480.

Babii, A., E. Ghysels, and J. Striaukas (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics 40*(3), 1094–1106.

Babii, A., E. Ghysels, and J. Striaukas (2024). High-dimensional granger causality tests with an application to vix and news. *Journal of Financial Econometrics 22*(3), 605–635.

Ball, L., D. Leigh, and P. Mishra (2022). Understanding us inflation during the covid-19 era. *Brookings Papers on Economic Activity 2022*(2), 1–80.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Belloni, A., C. Hansen, and W. Newey (2022). High-dimensional linear models with many endogenous variables. *Journal of Econometrics 228*(1), 4–26.

Belloni, A. and R. I. Oliveira (2018). A high dimensional central limit theorem for martingales, with applications to context tree models. *arXiv preprint arXiv:1809.02741*.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of statistics 37*(4), 1705–1732.

Bobeica, E. and B. Hartwig (2023). The covid-19 shock and challenges for inflation modelling. *International journal of forecasting 39*(1), 519–539.

Burstein, A. T. (2006). Inflation and output dynamics with state-dependent pricing decisions. *Journal of Monetary Economics 53*(7), 1235–1257.

Bykhovskaya, A. and V. Gorin (2022). Cointegration in large VARs. *The Annals of Statistics 50*(3), 1593–1617.

Bykhovskaya, A. and V. Gorin (2024). Asymptotics of cointegration tests for high-dimensional VAR(k). *Review of Economics and Statistics*, 1–38.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(2), 349–372.

Cai, Z., H. Chen, and X. Liao (2023). A new robust inference for predictive quantile regression. *Journal of Econometrics 234*(1), 227–250.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*(1), 27–60.

Chang, J., S. X. Chen, C. Y. Tang, and T. T. Wu (2021). High-dimensional empirical likelihood inference. *Biometrika 108*(1), 127–147.

Chang, J., X. Chen, and M. Wu (2024). Central limit theorems for high dimensional dependent data. *Bernoulli 30*(1), 712–742.

Chang, J., C. Y. Tang, and T. T. Wu (2018). A new scope of penalized empirical likelihood with high-dimensional estimating equations. *The Annals of Statistics 46*(6B), 3185–3216.

Chen, D., L. Feng, P. A. Mykland, and L. Zhang (2024). High dimensional regression coefficient test with high frequency data. *Journal of Econometrics*, 105812.

Cheng, X. and Z. Liao (2015). Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics 186*(2), 443–464.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.

Chernozhukov, V., W. Karl Härdle, C. Huang, and W. Wang (2021). Lasso-driven inference in time and space. *The Annals of Statistics 49*(3), 1702–1735.

Cúrdia, V. (2022). Average inflation targeting in the financial crisis recovery. *FRBSF Economic Letter 2022*(01), 1–05.

Demetrescu, M., I. Georgiev, P. M. Rodrigues, and A. R. Taylor (2023). Extensions to IVX methods of inference for return predictability. *Journal of Econometrics 237*(2), 105271.

Fan, J. and Y. Liao (2014). Endogeneity in high dimensions. *Annals of statistics 42*(3), 872.

Fan, J., J. Zhang, and K. Yu (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association 107*(498), 592–606.

Fan, Q., Z. Guo, and Z. Mei (2024). A heteroskedasticity-robust overidentifying restriction test with high-dimensional covariates. *Journal of Business & Economic Statistics* (just-accepted), 1–22.

Fan, Q., Z. Guo, Z. Mei, and C.-H. Zhang (2023). Uniform inference for nonlinear endogenous treatment effects with high-dimensional covariates. *arXiv preprint arXiv:2310.08063*.

Fan, R. and J. H. Lee (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics 213*(1), 261–280.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Galı, J. and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics 44*(2), 195–222.

Gao, Z., J. H. Lee, Z. Mei, and Z. Shi (2024). On LASSO inference for high dimensional predictive regression. *arXiv preprint arXiv:2409.10030*.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Gold, D., J. Lederer, and J. Tao (2020). Inference for high-dimensional instrumental variables regression. *Journal of Econometrics 217*(1), 79–111.

Goulet Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics 37*(5), 920–964.

Groen, J. J., R. Paap, and F. Ravazzolo (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics 31*(1), 29–44.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application.* Academic press.

Hong, Y., F. Jiang, L. Meng, and B. Xue (2024). Forecasting inflation using economic narratives. *Journal of Business & Economic Statistics* (just-accepted), 1–27.

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica 74*(3), 681–714.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research 15*(1), 2869–2909.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Knotek, E. S. and S. Zaman (2024). Nowcasting inflation.

Kock, A. B. and D. Preinerstorfer (2023). Consistency of p-norm based tests in high dimensions: Characterization, monotonicity, domination. *Bernoulli 29*(3), 2544–2573.

Kock, A. B. and D. Preinerstorfer (2024). Enhanced power enhancements for testing many moment equalities: Beyond the 2-and ∞-norm. *arXiv preprint arXiv:2407.17888*.

Koo, B., H. M. Anderson, M. H. Seo, and W. Yao (2020). High-dimensional predictive regression in the presence of cointegration. *Journal of Econometrics 219*(2), 456–477.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies 28*(5), 1506–1553.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2018). Taking stock of long-horizon predictability tests: Are factor returns predictable? *Available at SSRN 3284149*.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics 192*(1), 105–118.

Lee, J. H., Z. Shi, and Z. Gao (2022). On LASSO for predictive regression. *Journal of Econometrics 229*(2), 322–349.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024). Local projections vs. vars: Lessons from thousands of dgps. *Journal of Econometrics*, 105722.

Liao, C., Z. Mei, and Z. Shi (2024). Nickell meets stambaugh: A tale of two biases in panel predictive regressions. *arXiv preprint arXiv:2410.09825*.

Liu, X., W. Long, L. Peng, and B. Yang (2023). A unified inference for predictive quantile regression. *Journal of the American Statistical Association*, 1–15.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American economic review 63*(3), 326–334.

Magdalinos, T. and P. C. B. Phillips (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory 25*(2), 482–526.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics 39*(1), 98–119.

Mei, Z., P. C. B. Phillips, and Z. Shi (2024). The boosted hodrick-prescott filter is more general than you might think. *Journal of Applied Econometrics*.

Mei, Z., L. Sheng, and Z. Shi (2023). Nickell bias in panel local projection: Financial crises are worse than you think. *arXiv preprint arXiv:2302.13455*.

Mei, Z. and Z. Shi (2024). On LASSO for high dimensional predictive regression. *Journal of Econometrics 242*(2), 105809.

Merikoski, J. K. and R. Kumar (2004). Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes 4*(150-159), 9.

Merlevède, F., M. Peligrad, and E. Rio (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*(4), 1789–1823.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*(3), 687–719.

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica 86*(4), 1465–1478.

Onatski, A. and C. Wang (2021). Spurious factor analysis. *Econometrica 89*(2), 591–614.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *Economica 25*(100), 283–299.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, 283–306.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica 82*(3), 1177–1195.

Phillips, P. C. B. and I. L. Kheifets (2024). High-dimensional iv cointegration estimation and inference. *Journal of Econometrics 238*(2), 105622.

Phillips, P. C. B. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics 177*(2), 250–264.

Phillips, P. C. B. and J. H. Lee (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics 192*(2), 433–450.

Phillips, P. C. B. and T. Magdalinos (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper 7.*

Phillips, P. C. B., S. Shi, and J. Yu (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International Economic Review 56*(4), 1043–1078.

Phillips, P. C. B., S. Shi, and J. Yu (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review 56*(4), 1079–1134.

Phillips, P. C. B. and Z. Shi (2021). Boosting: Why you can use the HP filter. *International Economic Review 62*(2), 521–570.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and VARs estimate the same impulse responses. *Econometrica 89*(2), 955–980.

Shi, S. and J. Yu (2023). Volatility puzzle: Long memory or antipersistency. *Management Science 69*(7), 3861–3883.

Shi, Z. (2016). Econometric estimation with high-dimensional moment equalities. *Journal of Econometrics 195*(1), 104–119.

Smeekes, S. and E. Wijler (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting 34*(3), 408–430.

Smeekes, S. and E. Wijler (2021). An automated approach towards sparse single-equation cointegration modelling. *Journal of Econometrics 221*(1), 247–276.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Stock, J. H. and M. W. Watson (2002, August). Has the business cycle changed and why? Working Paper 9127, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

Strassex, V. (1967). Almost sure behavior of sums of independent random variables. In *Proceedings*, Volume 2, pp. 315. University of Caifornia Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

Tu, Y. and X. Xie (2023). Penetrating sporadic return predictability. *Journal of Econometrics 237*(1), 105509.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

Yang, B., W. Long, L. Peng, and Z. Cai (2020). Testing the predictability of US housing price index returns based on an IVX-AR model. *Journal of the American Statistical Association 115*(532), 1598–1619.

Yousuf, K. and S. Ng (2021). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics 224*(1), 60–87.

Zhang, B., G. Pan, and J. Gao (2018). Clt for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *The Annals of Statistics 46*(5), 2186–2215.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(1), 217–242.

Zhang, R., P. Robinson, and Q. Yao (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association 114*(526), 916–927.

Zhang, X. and G. Cheng (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association 112*(518), 757–768.

Online Appendices for "Simultaneous Inference for Predictability with High Dimensional Mixed Roots"

Ziwei Mei
The Chinese University of Hong Kong

A The Expression of $\varphi^{[j]*}$

Define $\Sigma^{(0)} = \mathbb{E}(x_t^{(0)} x_t^{(0)\top})$. For any $j \in \mathcal{G}^{(0)}$, define $e_{-j,t}^{(0)}$ as the subvector of $e_t^{(0)}$ removing $e_{j,t}$, $\Sigma_{-j,-j}^{(0)}$ as the submatrix of $\Sigma^{(0)}$ removing the row and column associated with $x_{j,t}$, $\Sigma_{\cdot,j}^{(0)}$ as the column of $\Sigma^{(0)}$ associated with $x_{j,t}$, and $\Sigma_{-j,j}^{(0)}$ as the subvector of $\Sigma_{\cdot,j}^{(0)}$ removing the entry associated with $x_{j,t}$. Write $\varphi^{[j]*} = (\varphi^{[j](0)*\top}, \varphi^{[j](1)*\top})^\top$, where $\varphi^{[j](k)*}$ collects the coefficients of regressors in the group $\mathcal{G}^{(k)}$, denoted as $\{x_{\ell,t}\}_{\ell \in \mathcal{G}^{(k)} \backslash \{j\}}$.

CASE I: $j \in \mathcal{G}^{(1)}$. Then

$$\varphi^{[j]*} = 0_{p-1}.$$

CASE II: $j \in \mathcal{G}^{(0)}$. Then

$$\varphi^{[j](k)*} = \begin{cases} (\Sigma_{-j,-j}^{(0)})^{-1} \Sigma_{-j,j}^{(0)}, & k = 0, \\ 0_{p_1}, & k = 1. \end{cases}$$

B Proofs

Throughout the proofs, we use c and C to denote generic positive constants that may vary from place to place. Let $\mathbf{I}_n$ be the $n \times n$ identity matrix. For any positive sequences $\{a_n\}$ and $\{b_n\}$, "$a_n \overset{\mathrm{p}}{\preccurlyeq} b_n$" means that there is an absolute constant, say c, such that the event $\{a_n \leq c b_n\}$ holds with probability approaching one (w.p.a.1.). Symmetrically, "$a_n \overset{\mathrm{p}}{\succcurlyeq} b_n$" means "$b_n \overset{\mathrm{p}}{\preccurlyeq} a_n$". The integer floor function is denoted as $\lfloor \cdot \rfloor$. For an n-dimensional vector $x = (x_t)_{t \in [n]}$, the L_2-norm is $\|x\|_2 = \sqrt{\sum_{t=1}^n x_t^2}$. For notational simplicity, in the proofs we assume $p \geq n^{\nu_1}$ for some absolute constant ν_1, which is reasonable as we focus on the high dimensional case with a larger p relative to n.[B.1] We always assume the initial value $x_{j,t=0} = 0$ without loss of generality.

[B.1] There is no technical difficulty in allowing p to grow either slowly at a logarithmic or fast at an exponential rate of n, but we have to compare $\log p$ and $\log n$ in many places, and in many conditions and rates the term "$\log p$" has to be changed into $\log(np)$.

B.1 Proofs for Section 3

B.1.1 Preparatory Propositions for DB and RE

Proposition B.1. *Under Assumptions 1-5, w.p.a.1*

$$\kappa_{\mathbf{I}}(\widehat{\Sigma}^{(1)}, 3, s) \geq \frac{c_{\kappa}^{(1)} n}{s \cdot \log p}. \tag{B.1}$$

Proof of Proposition B.1. This proposition is a direct corollary of Lemma B.5 by taking $L = 3$ and $c_{\kappa}^{(1)} = 9$, and the fact that $\kappa_{\mathbf{I}}(\widehat{\Sigma}^{(1)}, 3, s) \geq \kappa_{\mathbf{I}}(\ddot{\Sigma}^{(1)}, L, s)$. □

Proposition B.2. *Under Assumptions 1-5,*

(a) For stationary regressors, there exists some absolute constants $\sigma_{\min} < \sigma_{\max}$, such that

$$\sigma_{\min} \leq \min_{j \in \mathcal{G}^{(0)}} \widehat{\sigma}_j \leq \max_{j \in \mathcal{G}^{(0)}} \widehat{\sigma}_j \leq \sigma_{\max}. \tag{B.2}$$

(b) For nonstationary regressors,

$$\sqrt{\frac{n}{\log p}} \overset{\mathrm{p}}{\preccurlyeq} \min_{j \in \mathcal{G}^{(1)}} \widehat{\sigma}_j \leq \min_{j \in \mathcal{G}^{(1)}} \widehat{\sigma}_j \overset{\mathrm{p}}{\preccurlyeq} \sqrt{n \log p}. \tag{B.3}$$

Proof of Proposition B.2. **For Part (a)**, note that $\Sigma^{(0)}$ is a nonrandom positive definite matrix with diagonal entries bounded away from zero and above. Also, $\{n^{-1} \sum_{t=1}^{n} x_{j,t-1}^2\}_{j \in \mathcal{G}^{(0)}}$ are diagonal entries of $\widehat{\Sigma}^{(0)}$ Therefore, there exists some absolute there exists some absolute constants $\sigma_{\min}$, such that

$$\begin{aligned}
\min_{j \in \mathcal{G}^{(0)}} \frac{1}{n} \sum_{t=1}^{n} x_{j,t-1}^2 &\geq 2\sigma_{\min}^2 - \|\widehat{\Sigma}^{(0)} - \Sigma^{(0)}\|_{\infty} \\
&\geq 1.5\sigma_{\min}^2
\end{aligned}$$

when n is sufficiently large, where the second inequality applies (B.87). In addition,

$$\bar{x}_j = n^{-1} \sum_{t=1}^{n} \sum_{s=0}^{t-1} \rho_j^{*t-1-s} e_{j,s} = n^{-1} \sum_{t=1}^{n} (\sum_{s=t}^{n} \rho_j^{*(t-1-s)}) e_{j,t}$$

where

$$\max_{j \in \mathcal{G}^{(0)}} \sum_{s=t-1}^{n} \rho_j^{*(s-t+1)} = \max_{j \in \mathcal{G}^{(0)}} \frac{1 - \rho_j^{*(n-t+1)}}{1 - \rho_j^*} = O(1).$$

Therefore, $(\sum_{s=t}^{n} \rho_j^{*(t-1-s)})e_{j,t}$ is a sup-exponential and α-mixing sequence. Standard bounds of partial sum of centered sup-exponential weakly dependent time series, like (B.29), yield

$$\max_{j\in\mathcal{G}^{(0)}} \left| \frac{1}{n}\sum_{t=1}^{n}(\sum_{s=t}^{n}\rho_j^{*(t-1-s)})e_{j,t}\right| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{\frac{\log p}{n}}.$$

Therefore,

$$\max_{j\in\mathcal{G}^{(0)}} |\bar{x}_j| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{\frac{\log p}{n}}.$$

By $\widehat{\sigma}_j^2 = n^{-1}\sum_{t=1}^{n} x_{j,t-1}^2 - \bar{x}_j^2$, we have

$$\begin{aligned}\min_{j\in\mathcal{G}^{(0)}} \frac{1}{n}\widehat{\sigma}_j^2 &\geq \min_{j\in\mathcal{G}^{(0)}} \frac{1}{n}\sum_{t=1}^{n} x_{j,t-1}^2 - \max_{j\in\mathcal{G}^{(k)}} \bar{x}_j^2 \\ &\geq 1.5\sigma_{\min}^2 - \frac{\log p}{n} \geq \sigma_{\min}^2\end{aligned}$$

when n is sufficiently large. This result verifies the lower bound of (B.2). The upper bound can be deduced in a parallel way.

For Part (b), the lower bound follows by

$$\min_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j^2 \geq \kappa_{\mathbf{I}}(\ddot{\Sigma}^{(1)}, 3, 1) \geq \frac{c_\kappa^{(1)}}{\log p}$$

where the second inequality applies Lemma B.5. For the upper bound, by (B.50), the regressor $x_{j,t} = \sum_{s=0}^{t-1}\rho_j^{*t-1-s}e_{j,s}$ is a partial sum of a sub-exponential and α-mixing sequence. By Lemma B.2 of MS,

$$\max_{j\in\mathcal{G}^{(1)}, t\in[n]} |x_{j,t}| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{n\log p} \tag{B.4}$$

and therefore

$$\max_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j^2 \leq \max_{j\in\mathcal{G}^{(1)}} n^{-1}\sum_{t=1}^{n} x_{j,t-1}^2 \leq \max_{j\in\mathcal{G}^{(1)}, t\in[n]} |x_{j,t}|^2 \overset{\mathrm{p}}{\preccurlyeq} n\log p.$$

We complete the proof of Proposition B.2. $\square$

Proposition B.3. *Under Assumptions 1-5, for $k \in \{0, 1\}$ we have*

$$\|n^{-1}\sum_{t=1}^{n} x_{t-1}^{(k)}u_t\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-k}}}. \tag{B.5}$$

as $n \to \infty$.

Proof of Proposition B.3. Equation (B.5) is a direct corollary of (B.45) in Lemma B.2, since u_t is stationary, strong mixing, and m.d.s. such that $\mathbb{E}(x_{t-1}^{(k)} u_t) = 0$. $\square$

B.1.2 Proof of Main Results

Proof of Lemma 1. By Proposition B.2,

$$\frac{1}{\min_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j \min_{\ell\in\mathcal{G}^{(0)}} \widehat{\sigma}_\ell} \overset{\mathrm{p}}{\preccurlyeq} \frac{1}{\sqrt{\log p}/\sqrt{n}} = \sqrt{\frac{n}{\log p}}. \tag{B.6}$$

Therefore,

$$\begin{aligned}
\|D^{-1}\left(\widehat{\Sigma} - \widehat{\Delta}\right) D^{-1}\|_\infty &\leq \frac{\|n^{-1}\sum_{t=1}^{n} x_{t-1}^{(0)} x_{t-1}^{(1)\top}\|_\infty}{\min_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j \min_{\ell\in\mathcal{G}^{(0)}} \widehat{\sigma}_\ell} \\
&\overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{1+\frac{1}{2r}}}{\sqrt{n/\log p}} \\
&\overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}} = O_p\left(\frac{(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}}\right),
\end{aligned}$$

where the second row applies Lemma B.4 and (B.6). $\square$

Proof of Lemma 2. Define $\widehat{\sigma}_{\max}^{(1)} = \max_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j$, $\widehat{\sigma}_{\min}^{(1)} = \min_{j\in\mathcal{G}^{(1)}} \widehat{\sigma}_j$, and $\widehat{\varsigma}^{(1)} = \widehat{\sigma}_{\max}^{(1)}/\widehat{\sigma}_{\min}^{(1)}$. Further define $\widetilde{\delta}^{(1)} := (D^{(1)})^{-1}\delta = (\widehat{\sigma}_j^{-1}\delta_j)_{j\in\mathcal{G}^{(1)}}$. Note that $\widehat{\sigma}_{\min}\|\widetilde{\delta}_{\mathcal{M}^c}\|_1 \leq \|\delta_{\mathcal{M}^c}\|_1$ and $\|\delta_{\mathcal{M}}\|_1 \leq \widehat{\sigma}_{\max}\|\widetilde{\delta}_{\mathcal{M}}\|_1$. Therefore, whenever $\delta \in \mathcal{R}(3,s)$ such that $\|\delta_{\mathcal{M}^c}\|_1 \leq 3\|\delta_{\mathcal{M}}\|_1$ for any $|\mathcal{M}| \leq s$, we have $\|\widetilde{\delta}_{\mathcal{M}^c}\|_1 \leq \widehat{\varsigma}^{(1)}\|\widetilde{\delta}_{\mathcal{M}}\|_1$ and $\widetilde{\delta} \in \mathcal{R}\left(3\widehat{\varsigma}^{(1)}, s\right)$. Then

$$\begin{aligned}
\widehat{\kappa}_D^{(1)} &= \inf_{\delta\in\mathcal{R}(3,s)} \frac{\delta^\top (D^{(1)})^{-1}\widehat{\Sigma}^{(1)}(D^{(1)})^{-1}\delta}{\delta^\top\delta} = \inf_{\delta\in\mathcal{R}(3,s)} \frac{\widetilde{\delta}^\top\widehat{\Sigma}^{(1)}\widetilde{\delta}}{\widetilde{\delta}^\top(D^{(1)})^2\widetilde{\delta}} \\
&\geq \inf_{\widetilde{\delta}\in\mathcal{R}\left(3\widehat{\varsigma}^{(1)},s\right)} \frac{\widetilde{\delta}^\top\widehat{\Sigma}^{(1)}\widetilde{\delta}}{\widetilde{\delta}^\top(D^{(1)})^2\widetilde{\delta}} \geq (\widehat{\sigma}_{\max}^{(1)})^{-2} \inf_{\widetilde{\delta}\in\mathcal{R}\left(3\widehat{\varsigma}^{(1)},s\right)} \frac{\widetilde{\delta}^\top\widehat{\Sigma}^{(1)}\widetilde{\delta}}{\widetilde{\delta}^\top\widetilde{\delta}} = \frac{\kappa(\widehat{\Sigma}^{(1)}, 3\widehat{\varsigma}^{(1)}, s)}{(\widehat{\sigma}_{\max}^{(1)})^2}.
\end{aligned}$$

Taking $L = 3\widehat{\varsigma}^{(1)}$. By Proposition B.5 we have $\kappa(\widehat{\Sigma}^{(1)}, 3\widehat{\varsigma}^{(1)}, s) \geq \dfrac{cn}{9s\log p(\widehat{\sigma}_{\max}^{(1)})^2 \cdot (\widehat{\varsigma}^{(1)})^2}$ w.p.a.1 for some absolute constant c. By (B.3), there exists some absolute constant c' such that

$$(\widehat{\varsigma}^{(1)})^2 \geq c'(\log p)^2 \text{ and } (\widehat{\sigma}_{\max}^{(1)})^2 \geq c' n \log p.$$

Therefore,

$$\kappa(\widehat{\Sigma}^{(1)}, 3\widehat{\varsigma}^{(1)}, s) \geq \frac{cn}{9s(\log p)^4(c')^2}$$

w.p.a.1. Then Lemma 2 follows with $c_\kappa^{(1)} = c/(3c')^2$. $\square$

Proof of Proposition 1. We have for any δ

$$\delta^\top D^{-1}\widehat{\Sigma}D^{-1}\delta \geq \delta^\top D^{-1}\widehat{\Delta}D^{-1}\delta - \|D^{-1}(\widehat{\Sigma}-\widehat{\Delta})D^{-1}\|_\infty\|\delta\|_1^2.$$

Lemmas 2 and (B.88) suggest that for any $\delta \in \mathcal{R}(3,s)$

$$\frac{\delta^\top D^{-1}\widehat{\Delta}D^{-1}\delta}{\|\delta\|_2^2} \geq \frac{c}{s(\log p)^4}$$

for some absolute constant c. By (B.79) and Lemma 1,

$$\begin{aligned}
\frac{\|D^{-1}(\widehat{\Sigma}-\widehat{\Delta})D^{-1}\|_\infty\|\delta\|_1^2}{\|\delta\|_2^2} &= O_p\left(\frac{(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}}\right)\cdot\frac{\|\delta\|_1^2}{\|\delta\|_2^2} \\
&= O_p\left(\frac{s(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}}\right).
\end{aligned}$$

Therefore,

$$\frac{\delta^\top D^{-1}\widehat{\Sigma}D^{-1}\delta}{\|\delta\|_2^2} \geq \frac{c}{s(\log p)^4} + O_p\left(\frac{s(\log p)^{\frac{3}{2}+\frac{1}{2r}}}{\sqrt{n}}\right) \geq \frac{0.5c}{s(\log p)^4}$$

when n is sufficiently large, where the second inequality applies Assumption 4. Proposition 1 thus follows with $c_\kappa = 0.5c$. $\square$

Proof of Proposition 2. The DB condition follows by

$$\begin{aligned}
n^{-1}\|\sum_{t=1}^n D^{-1}x_{t-1}u_t\|_\infty &= \max_{j\in[p]} n^{-1}\left|\sum_{t=1}^n \frac{x_{j,t-1}}{\widehat{\sigma}_j}u_t\right| \\
&\overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{2+\frac{1}{2r}}}{\min_{k\in\{0,1\}}\sqrt{n^{1+k}}\min_{j\in\mathcal{G}^{(k)}}\widehat{\sigma}_j} \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{\frac{5}{2}+\frac{1}{2r}}}{\sqrt{n}},
\end{aligned}$$

where the first inequality in the second row applies Proposition B.3, and the last inequality applies Proposition B.2. $\square$

Proof of Theorem 1. By Lemma 1 of MS, our Propositions 1 and 2 imply

$$\|n^{-1}\sum_{t=1}^n x_{t-1}^\top(\widehat{\beta}^{\mathrm{S}}-\beta^*)\|_2^2 = O_p\left(\frac{s(\log p)^{5+\frac{1}{r}}/n}{1/s(\log p)^4}\right) = O_p\left(\frac{s^2(\log p)^{9+\frac{1}{r}}}{n}\right), \tag{B.7}$$

$$\|D(\widehat{\beta}^{\mathrm{S}}-\beta^*)\|_2 = O_p\left(\frac{\sqrt{s}(\log p)^{\frac{5}{2}+\frac{1}{2r}}/\sqrt{n}}{1/s(\log p)^4}\right) = O_p\left(\frac{s^{3/2}(\log p)^{\frac{13}{2}+\frac{1}{2r}}}{\sqrt{n}}\right), \tag{B.8}$$

and

$$\|D(\widehat{\beta}^{\mathrm{S}}-\beta^{*})\|_{1}=O_{p}\left(\frac{s(\log p)^{\frac{5}{2}+\frac{1}{2r}}/\sqrt{n}}{1/s(\log p)^{4}}\right)=O_{p}\left(\frac{s^{2}(\log p)^{\frac{13}{2}+\frac{1}{2r}}}{\sqrt{n}}\right). \tag{B.9}$$

as $n\to\infty$. (18) follows (B.7). (19) and (20) follow by (B.8), (B.9), Proposition B.2, and the inequality

$$\|\widehat{\beta}^{(k)\mathrm{S}}-\beta^{(k)*}\|_{q}\leq\frac{1}{\min_{j\in\mathcal{G}^{(k)}}\widehat{\sigma}_{j}}\|D^{(k)}(\widehat{\beta}^{(k)\mathrm{S}}-\beta^{(k)*})\|_{q}\leq\frac{1}{\min_{j\in\mathcal{G}^{(k)}}\widehat{\sigma}_{j}}\|D(\widehat{\beta}^{\mathrm{S}}-\beta^{*})\|_{q}$$

for $q=1,2$. □

B.2 Proofs for Section 3

B.2.1 Auxiliary Regression and Bias Correction

Define

$$\widetilde{\varphi}^{[j]*}=\widehat{\tau}_{j}^{-1}\varphi^{[j]*}$$

where $\widehat{\tau}_{j}$ is the sample s.d. of the instrument $z_{j,t}$ defined in (22), and $\varphi^{[j]*}$ is the pseudo-true coefficient is defined in Section A. Define $\delta_{\min}:=\theta\wedge(1-\theta)$.

Proposition B.4. *Under the conditions in Theorem 2, when* $\mu_{j}=\frac{C_{j}(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}$

$$\max_{j\in[p]}\|D_{-j}(\widehat{\varphi}^{[j]}-\widetilde{\varphi}^{[j]*})\|_{1}\overset{\mathrm{p}}{\preccurlyeq}\frac{s^{2}(\log p)^{\frac{15}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}.$$

Proof of Proposition B.4. The lower bound of RE, $c_{\kappa}/s(\log p)^{4}$, has been established by Proposition 1. By Lemma 1 in MS, it suffices to show the following DB condition

$$\max_{j\in[p]}\|n^{-1}\sum_{t=1}^{n}D_{-j}^{-1}x_{-j,t-1}(\widetilde{z}_{j,t-1}-x_{-j,t-1}^{\top}\widetilde{\varphi}^{[j]*})\|_{\infty}\overset{\mathrm{p}}{\preccurlyeq}\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}.$$

We only need to show

$$\max_{j\in\mathcal{G}^{(k)}}\|n^{-1}\sum_{t=1}^{n}D_{-j}^{-1}x_{-j,t-1}(\widetilde{z}_{j,t-1}-x_{-j,t-1}^{\top}\widetilde{\varphi}^{[j]*})\|_{\infty}\overset{\mathrm{p}}{\preccurlyeq}\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}} \tag{B.10}$$

for $k=\{0,1\}$.

CASE I. When $k=1$, by the definition of $\varphi^{[j]*}$ in Section A we have $\widetilde{\varphi}^{[j]*}=0_{p-1}$ for any

$j \in \mathcal{G}^{(1)}$. In addition, by (B.115) we have for $m \in \{0,1\}$,

$$\max_{\ell \in \mathcal{G}^{(m)}, j \in \mathcal{G}^{(1)}} |\sum_{t=1}^{n} x_{\ell,t-1} z_{j,t-1}| \overset{\mathrm{p}}{\preccurlyeq} (n^{\theta \wedge m} + n^{\theta + \frac{m-1}{2}}) n (\log p)^{3+\frac{1}{2r}}.$$

Therefore,

$$\begin{aligned} &\max_{j \in \mathcal{G}^{(1)}} \|n^{-1} \sum_{t=1}^{n} D_{-j}^{-1} x_{-j,t-1} (\widetilde{z}_{j,t-1} - x_{-j,t-1}^{\top} \widetilde{\varphi}^{[j]*})\|_{\infty} \\ &\le \max_{m \in [K]} \frac{\max_{\ell \in \mathcal{G}^{(m)}, j \in \mathcal{G}^{(1)}} |\sum_{t=1}^{n} x_{\ell,t-1} z_{j,t-1}|}{n \cdot \min_{\ell \in \mathcal{G}^{(m)}} \widehat{\sigma}_{\ell} \min_{j \in \mathcal{G}^{(1)}} \widehat{\tau}_{j}} \overset{\mathrm{p}}{\preccurlyeq} \frac{(n^{\theta \wedge m} + n^{-\frac{1}{2} + \theta + \frac{m}{2}})(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^m}\sqrt{n^{\theta}}}, \end{aligned} \tag{B.11}$$

where the last inequality applies Proposition B.2 and Lemma B.8. In addition,

$$\frac{n^{\theta \wedge m}}{\sqrt{n^m}\sqrt{n^{\theta}}} = \frac{1}{\sqrt{n^{(m \vee \theta) - (\theta \wedge m)}}} \le \frac{1}{\sqrt{n^{\delta_{\min}}}},$$

and

$$\frac{n^{-\frac{1}{2} + \theta + \frac{m}{2}}}{\sqrt{n^m}\sqrt{n^{\theta}}} = \frac{1}{\sqrt{n^{(1-\theta)}}} \le \frac{1}{\sqrt{n^{\delta_{\min}}}}.$$

Therefore, by (B.11) we have

$$\max_{j \in \mathcal{G}^{(1)}} \|n^{-1} \sum_{t=1}^{n} D_{-j}^{-1} x_{-j,t-1} (\widetilde{z}_{j,t-1} - x_{-j,t-1}^{\top} \widetilde{\varphi}^{[j]*})\|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}. \tag{B.12}$$

CASE II. When $k = 0$, by (B.115)

$$\max_{\ell \in \mathcal{G}^{(m)}, j \in \mathcal{G}^{(0)}} |\sum_{t=1}^{n} x_{\ell,t-1} z_{j,t-1}| \overset{\mathrm{p}}{\preccurlyeq} (n^{k \wedge m} + n^{k + \frac{m-\theta}{2}}) \cdot n (\log p)^{3+\frac{1}{2r}}. \tag{B.13}$$

By the definition of $\varphi^{[j]*}$ in Section A, we have for any $j \in \mathcal{G}^{(0)}$, $\varphi_i^{[j]*} \neq 0$ if and only if $\gamma_i = k$. Therefore, when $m = 1$,

$$\begin{aligned} \max_{\ell \in \mathcal{G}^{(m)}, j \in \mathcal{G}^{(0)}} |\sum_{t=1}^{n} x_{\ell,t-1} x_{-j,t-1}^{\top} \varphi^{[j]*}| &\le \max_{\ell \in \mathcal{G}^{(m)}, i \in \mathcal{G}^{(0)}} |\sum_{t=1}^{n} x_{\ell,t-1} x_{i,t-1}| \max_{j \in \mathcal{G}^{(k)}} \|\varphi^{[j]*}\|_1 \\ &\overset{\mathrm{p}}{\preccurlyeq} n \cdot (\log p)^{1+\frac{1}{2r}}, \end{aligned} \tag{B.14}$$

where the last inequality applies (B.4) and Assumption 6. By (B.13) and (B.14), we have

$$\begin{aligned}
&\max_{\ell\in\mathcal{G}^{(1)},j\in\mathcal{G}^{(0)}}|n^{-1}\sum_{t=1}^{n}\frac{x_{\ell,t-1}}{\widehat{\sigma}_\ell}(\widetilde{z}_{j,t-1}-x_{-j,t-1}^\top\widetilde{\varphi}^{[j]*})|\\
\leq&\frac{\max_{\ell\in\mathcal{G}^{(1)},j\in\mathcal{G}^{(0)}}|\sum_{t=1}^{n}x_{\ell,t-1}(z_{j,t-1}-x_{-j,t-1}^\top\varphi^{[j]*})|}{n\cdot\min_{\ell\in\mathcal{G}^{(1)}}\widehat{\sigma}_\ell\min_{j\in\mathcal{G}^{(0)}}\widehat{\tau}_j}\\
\overset{\mathrm{p}}{\preccurlyeq}&\frac{n^{\frac{1-\theta}{2}}(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n}},
\end{aligned}\tag{B.15}$$

where the last inequality applies Proposition B.2 and Lemma B.8. Following the same arguments for (B.12), we have

$$\max_{\ell\in\mathcal{G}^{(1)},j\in\mathcal{G}^{(0)}}|n^{-1}\sum_{t=1}^{n}\frac{x_{\ell,t-1}}{\widehat{\sigma}_\ell}(z_{j,t-1}-x_{-j,t-1}^\top\varphi^{[j]*})|\overset{\mathrm{p}}{\preccurlyeq}\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}.\tag{B.16}$$

When $m=0$, recall that $z_{j,t}=x_{j,t}-(1-\rho_z)\psi_{j,t}$ by (B.102), and the definitions of $\Sigma^{(0)}_{-j,-j}$ and $\Sigma^{(k)}_{-j,j}$ in Section A. Again, recall by the definition in Section A that we have for any $j\in\mathcal{G}^{(0)}$, $\varphi_i^{[j]*}\neq 0$ if and only if $\gamma_i=\gamma^{(0)}$. Therefore,

$$\begin{aligned}
&\max_{\ell,j\in\mathcal{G}^{(0)},\ell\neq j}|n^{-1}\sum_{t=1}^{n}x_{\ell,t-1}(x_{j,t-1}-x_{-j,t-1}^\top\varphi^{[j]*})|\\
=&\max_{j\in\mathcal{G}^{(0)}}\|n^{-1}\sum_{t=1}^{n}x^{(k)}_{-j,t-1}(x_{j,t-1}-x_{-j,t-1}^\top\varphi^{[j]*})\|_\infty\\
\leq&\max_{j\in\mathcal{G}^{(0)}}\|n^{-1}\sum_{t=1}^{n}x^{(k)}_{-j,t-1}x_{j,t-1}-\Sigma^{(k)}_{\cdot,j}\|_\infty+\max_{j\in\mathcal{G}^{(0)}}\|n^{-1}\sum_{t=1}^{n}x^{(k)}_{-j,t-1}x^{(k)\top}_{-j,t-1}\varphi^{[j]*}-\Sigma^{(k)}_{\cdot,j}\|_\infty\\
\leq&\max_{j\in\mathcal{G}^{(0)}}\|n^{-1}\sum_{t=1}^{n}x^{(k)}_{-j,t-1}x_{j,t-1}-\Sigma^{(k)}_{\cdot,j}\|_\infty+\max_{j\in\mathcal{G}^{(0)}}\|n^{-1}\sum_{t=1}^{n}x^{(k)}_{-j,t-1}x^{(k)\top}_{-j,t-1}-\Sigma^{(k)}_{-j,-j}\|_\infty\cdot\|\varphi^{[j]*}\|_1\\
\overset{\mathrm{p}}{\preccurlyeq}&(\log p)^{3+\frac{2}{r}}+n^{-\frac{1}{2}}(\log p)^{3+\frac{1}{2r}},
\end{aligned}$$

where the last inequality applies (B.87) and the upper bound of $\varphi^{[j]*}$'s L_1-norm Assumption

6. Therefore,

$$
\begin{aligned}
&\max_{\ell,j\in\mathcal{G}^{(0)},\ell\neq j}|n^{-1}\sum_{t=1}^{n}\frac{x_{\ell,t-1}}{\widehat{\sigma}_\ell}(\widetilde{z}_{j,t-1}-x_{-j,t-1}^\top\widetilde{\varphi}^{[j]*})| \\
\leq&\frac{\max_{\ell,j\in\mathcal{G}^{(0)},\ell\neq j}|\sum_{t=1}^{n}x_{\ell,t-1}(z_{j,t-1}-x_{-j,t-1}^\top\varphi^{[j]*})|}{n\cdot\min_{\ell\in\mathcal{G}^{(k)}}\widehat{\sigma}_\ell\min_{j\in\mathcal{G}^{(k)}}\widehat{\tau}_j}\overset{\mathrm{p}}{\preccurlyeq}\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n}} \\
=&\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n}}\leq\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}.
\end{aligned}
\tag{B.17}
$$

By (B.16) and (B.16), when $k=1$,

$$
\max_{j\in\mathcal{G}^{(k)}}\|n^{-1}\sum_{t=1}^{n}D_{-j}^{-1}x_{-j,t-1}(\widetilde{z}_{j,t-1}-x_{-j,t-1}^\top\widetilde{\varphi}^{[j]*})\|_\infty\overset{\mathrm{p}}{\preccurlyeq}\frac{(\log p)^{\frac{7}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}. \tag{B.18}
$$

Equation (B.10) follows by (B.12) for $k=1$ and (B.18) for $k=0$. We complete the proof of Proposition B.4. □

Recall from (B.128) that $r_{j,t}^*=z_{j,t}-x_{-j,t}^\top\varphi^{[j]*}$ is the true error term in the auxiliary regression. Also, $\check{r}_{j,t}=\widehat{\tau}_j\widehat{r}_{j,t}$ is the residual of the auxiliary regression (23) standardized by the sample s.d. of the instrument. By some fundamental calculations, we have

$$
\frac{\widehat{\beta}_j^{\mathrm{XD}}-\beta_j^*}{\widehat{\omega}_j^{\mathrm{XD}}}=\mathrm{sgn}_j\left(\mathrm{N}_j+\mathrm{Err}_{j,1}+\mathrm{Err}_{j,2}\right), \tag{B.19}
$$

where $\mathrm{sgn}_j=|\sum_{t=1}^{n}\check{r}_{j,t-1}x_{j,t-1}|/\sum_{t=1}^{n}\check{r}_{j,t-1}x_{j,t-1}$ equals either 1 or -1, and

$$
\mathrm{N}_j=\frac{\sum_{t=1}^{n}r_{j,t-1}^*u_t}{\sqrt{n\Pi_{j,j}}}, \tag{B.20}
$$

with $\Pi_{j,j}=n^{\theta\wedge\gamma_j}\widetilde{\Pi}_{j,j}$ by (B.143), and $\widetilde{\Pi}_{j,j}$ defined in (B.142) is a positive constant with an $O(1)$ order. Therefore,

$$
\sqrt{n\Pi_{j,j}}=O(\sqrt{n^{1+\theta\wedge\gamma_j}}).
$$

As for the errors,

$$
\mathrm{Err}_{j,1}=\sum_{t=1}^{n}r_{j,t-1}^*u_t\left(\frac{1}{\sqrt{n\Pi_{j,j}}}-\frac{1}{\sqrt{\sum_{t=1}^{n}\widetilde{r}_{j,t-1}^2}}\right),
$$

and

$$\mathrm{Err}_{j,2} = \frac{\sum_{t=1}^{n}(\check{r}_{j,t-1} - r^*_{j,t-1})u_t - \sum_{t=1}^{n}\check{r}_{j,t-1}x_{t-1}^\top(\widehat{\beta}^{\mathrm{S}}_{-j} - \beta^*_{-j})}{\sqrt{\sum_{t=1}^{n}\check{r}^2_{j,t-1}}}.$$

Proposition B.5. *Under the conditions in Theorem 2, we have*

$$\sup_{j\in[p]}|\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}| \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{10+1/r}}{\sqrt{n^{\delta_{\min}}}}$$

as $n \to \infty$, where $\delta_{\min} = \theta \wedge (1-\theta)$.

Proof of Proposition B.5. Define $\phi_j = \theta \wedge \gamma_j$. We first bound $\mathrm{Err}_{j,1}$. By (B.146),

$$\sup_{j\in[p]}\sqrt{n^{1+\phi_j}}\left|\frac{1}{\sqrt{n\Pi_{j,j}}} - \frac{1}{\sqrt{\sum_{t=1}^{n}\tilde{r}^2_{j,t-1}}}\right| \le \|Q_n^{1/2}(D_\Pi^{-1/2} - \widehat{D}_{\Pi,n}^{-1/2})\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}. \quad \text{(B.21)}$$

In addition, by (B.135) and the definition of $v_{j,t}$ in (B.136),

$$\begin{aligned}
\sup_{j\in[p]}\left|\frac{1}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}r^*_{j,t-1}u_t\right| &\le \sup_{j\in[p]}\left|\frac{1}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}v_{j,t-1}u_t\right| + \sup_{j\in[p]}\left|\frac{1}{n^{\theta\vee\gamma_j}\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}\psi_{j,t-1}u_t\right| \\
&\overset{\mathrm{p}}{\preccurlyeq} \sup_{j\in[p]}\frac{\sqrt{n^{1+\phi_j}}(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1+\phi_j}}} + \sup_{j\in[p]}\frac{n^{\frac{1}{2}+\frac{\gamma_j\vee\theta}{2}+(\theta\wedge\gamma_j)}(\log p)^{2+\frac{1}{2r}}}{n^{\theta\vee\gamma_j}\sqrt{n^{1+(\theta\wedge\gamma_j)}}} \\
&\le 2(\log p)^{2+\frac{1}{2r}}, \qquad \text{(B.22)}
\end{aligned}$$

where in the second line, the upper bound of the first term Proposition B.3, and the upper bound of the second term applies (B.105). Then (B.21) and (B.22) yield

$$\sup_{j\in[p]}|\mathrm{Err}_{j,1}| \overset{\mathrm{p}}{\preccurlyeq} \sup_{j\in[p]}\left|\frac{1}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}r^*_{j,t-1}u_t\right| \cdot \sup_{j\in[p]}\sqrt{n^{1+\phi_j}}\left|\frac{1}{\sqrt{n\Pi_{j,j}}} - \frac{1}{\sqrt{\sum_{t=1}^{n}\check{r}^2_{j,t-1}}}\right| \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{13+\frac{1}{r}}}{\sqrt{n^{\delta_{\min}}}}.$$

We then bound $\mathrm{Err}_{j,2}$. By (B.134), there exists some absolute constant c such that

$$\min_{j\in[p]}\sqrt{\frac{1}{n^{1+\phi_j}}\sum_{t=1}^{n}\check{r}^2_{j,t-1}} = \min_{j\in[p]}\sqrt{\frac{1}{n^{1+\phi_j}}\sum_{t=1}^{n}\check{r}^2_{j,t-1}} > c. \qquad \text{(B.23)}$$

In addition,

$$\begin{aligned}\max_{j\in[p]}\left|\frac{1}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}(\check r_{j,t-1}-r^*_{j,t-1})u_t\right| &= \max_{j\in[p]}\left|\frac{\widehat\tau_j}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}u_t x_{-j,t}^\top D_{-j}^{-1}D_{-j}(\widehat\varphi^{[j]}-\widetilde\varphi^{[j]*})\right|\\ &\leq \max_{j\in[p]}\frac{\widehat\tau_j}{\sqrt{n^{1+\phi_j}}}\|\sum_{t=1}^{n}D_{-j}^{-1}x_{-j,t}u_t\|_\infty\cdot\|D_{-j}(\widehat\varphi^{[j]}-\widetilde\varphi^{[j]*})\|_1\\ &\overset{\text{p}}{\preccurlyeq}\frac{s^2(\log p)^{\frac{19}{2}+\frac{1}{r}}}{\sqrt{n^{\delta_{\min}}}},\end{aligned}\tag{B.24}$$

where the last inequality applies (B.122) that bounds $\widehat\tau_j$, the DB condition in Proposition 2, and the estimation error of the auxiliary regression by Proposition B.4. Finally,

$$\begin{aligned}&\max_{j\in[p]}\left|\frac{1}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}\check r_{j,t-1}x_{t-1}^\top(\widehat\beta^{\text{S}}_{-j}-\beta^*_{-j})\right|\\ &=\max_{j\in[p]}\left|\frac{\widehat\tau_j}{\sqrt{n^{1+\phi_j}}}\sum_{t=1}^{n}\widehat r_{j,t-1}x_{t-1}^\top D_{-j}^{-1}D_{-j}(\widehat\beta^{\text{S}}_{-j}-\beta^*_{-j})\right|\\ &\leq\max_{j\in[p]}\frac{\widehat\tau_j}{\sqrt{n^{1+\phi_j}}}\|\sum_{t=1}^{n}D_{-j}^{-1}x_{t-1}^\top\widehat r_{j,t-1}\|_\infty\cdot\|D_{-j}(\widehat\beta^{\text{S}}_{-j}-\beta^*_{-j})\|_1\\ &\overset{\text{p}}{\preccurlyeq}\max_{j\in[p]}\frac{\widehat\tau_j}{\sqrt{n^{1+\phi_j}}}\cdot n\mu_j\cdot\frac{s^2(\log p)^{\frac{13}{2}+\frac{1}{2r}}}{\sqrt n}\overset{\text{p}}{\preccurlyeq}\frac{s^2(\log p)^{10+\frac{1}{r}}}{\sqrt{n^{\delta_{\min}}}},\end{aligned}\tag{B.25}$$

where the third inequality applies the Karush–Kuhn–Tucker condition of the auxiliary regression (23) with μ_j being the LASSO tuning parameter, and the LASSO regression error (B.9); the last inequality applies (B.122) that bounds $\widehat\tau_j$, and the convergence rate of μ_j specified in Proposition B.4. Then $\text{Err}_{j,2}$ is bounded by (B.23), (B.24), and (B.25). We complete the proof of Proposition B.5. □

B.2.2 Coupling

Recall that N_j is defined in (B.20), and write $\text{N}=(\text{N}_j)_{j\in[p]}$.

Proposition B.6. *Under the conditions in Theorem 2, there exists a sequence of vectors $\text{Z}\in\mathbb{R}^p$ such that (a) $\text{Z}_{\mathcal{I}}$ is normally distributed conditioning on $\{r^*_t\}_{t\in[n-1]}$, and (b) there exists some absolute constant c such that*

$$\|\text{N}-\text{Z}\|_\infty\overset{\text{p}}{\preccurlyeq}n^{-c},\tag{B.26}$$

and

$$\|\mathbb{E}(\text{Z}\text{Z}^\top|\{r^*_{t-1}\}_{t\in[n]})-V_n\|_\infty=O_p\left(n^{-1/4}(\log p)^c\right).\tag{B.27}$$

as $n \to \infty$, where V_n is defined above Lemma B.12.

Proof. We first use a normal approximation for $\chi_{j,t} = r^*_{j,t-1} u_t$, which is a martingale. Also, $\mathrm{N}_j = \frac{\sum_{t=1}^n \chi_{j,t}}{\sqrt{n\Pi_{j,j}}}$. Without loss of generality, suppose $\sigma_u^2 = 1$ to simplify the notation. According to the Skorokhod representation theorem (Strassex, 1967, Theorem 4.3), there exists a richer probability space that supports a standard Brownian motion $\mathcal{B}_j(\cdot)$ and a sequence of stopping time $\{\tau_{jt}\}_{t\geq 0}$ such that for all $t \geq 0$,

$$
\begin{aligned}
\sum_{s=0}^{t-1} \chi_{j,s} = \mathcal{B}_j\left(\sum_{s=0}^{t-1} \tau_{j,s}\right), \quad \text{almost surely.}\\
\mathbb{E}\left[\tau_{j,t}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\chi_{j,t}^2|\mathcal{F}_{t-1}\right] = r^{*2}_{t-1},\\
\mathbb{E}(|\tau_{jt}|^{\varpi}|\mathcal{F}_{t-1}) \leq \check{C}_{\varpi}\mathbb{E}(|\chi_{j,t}|^{2\varpi}|\mathcal{F}_{t-1}); \quad \forall \varpi > 4(\nu+1)
\end{aligned}
\tag{B.28}
$$

for some positive constant $\check{C}_{\varpi}$. The Brown motion conditional on the sequence of stopping time follows

$$
\left[\mathcal{B}_j\left(\sum_{t=1}^{n} \tau_{j,t-1}\right) - \mathcal{B}_j(\sum_{t=1}^{n} r^{*2}_{j,t-1})\right] \Bigg| \{\tau_{j,t}, r^*_t\}_{t\in[n-1]} \sim \mathcal{N}\left(0, \left|\sum_{t=1}^{n} \widetilde{\tau}_{j,t-1}\right|\right)
$$

where $\widetilde{\tau}_{j,s} = \tau_{j,s} - \mathbb{E}\left[\chi^2_{j,t-1}|\mathcal{F}_{t-1}\right]$ is the demeaned version of $\tau_{j,s}$. The moment properties of the normal distribution gives

$$
\mathbb{E}\left[\left|\mathcal{B}_j\left(\sum_{t=1}^{n} \tau_{j,t-1}\right) - \mathcal{B}_j(\sum_{t=1}^{n} r^{*2}_{t-1})\right|^{\varpi} \Bigg| \{\widetilde{\tau}_{jt}\}_{0\leq t\leq n-1}\right] = \pi^{-1/2} 2^{\varpi/2}\Gamma\left(\frac{\varpi+1}{2}\right)\left|\sum_{t=1}^{n} \widetilde{\tau}_{j,t-1}\right|^{\varpi/2}, \quad \forall \varpi > 0,
\tag{B.29}
$$

where $\Gamma(\cdot)$ is the Gamma function. For simplicity, write $\mathbb{E}_r(\cdot) = \mathbb{E}(\cdot|\{r^*_{t-1}\}_{t\in[n]})$. Note that $\{\widetilde{\tau}_{js}\}_{s\geq 0}$ is a martingale difference sequence, and it thus satisfies for $\varpi \geq 4$ that

$$
\begin{aligned}
\mathbb{E}_r\left(\left|\sum_{s=0}^{n-1} \widetilde{\tau}_{js}\right|^{\varpi/2}\right) &\leq \frac{\varpi}{\varpi-2} C^{(1)}_{\varpi}\left[\mathbb{E}_r\left(\sum_{s=0}^{n-1} \mathbb{E}[\widetilde{\tau}^2_{js}]\right)^{\varpi/4} + \sum_{s=0}^{n-1} \mathbb{E}_r[|\widetilde{\tau}_{js}|^{\varpi/2}]\right]\\
&\leq \frac{\varpi}{\varpi-2} C^{(1)}_{\varpi}\left[\mathbb{E}_r\left(\check{C}_2\sum_{s=0}^{n-1} \mathbb{E}_r[\chi^4_{j,s}]\right)^{\varpi/4} + \check{C}_{\varpi/2}\sum_{s=0}^{n-1} \mathbb{E}_r|\chi_{j,s}|^{\varpi}\right]\\
&= \frac{\varpi}{\varpi-2} C^{(1)}_{\varpi}\left(\check{C}_2^{\varpi/4} \vee \check{C}_{\varpi/2}\right)\left[\mathbb{E}_r\left(\sum_{s=0}^{n-1} \mathbb{E}[\chi^4_{j,s}]\right)^{\varpi/4} + \sum_{s=0}^{n-1} \mathbb{E}_r|\chi_{j,s}|^{\varpi}\right]
\end{aligned}
\tag{B.30}
$$

where the first inequality by the Rosenthal's inequality (Hall and Heyde, 1980, Theorem 2.12, p.23) with an absolute constant $C_{\varpi}^{(1)}$, and the second inequality by (B.28). Notice that Lemma B.9 implies

$$\sup_{j\in[p]} \mathbb{E}_r \left(\sum_{s=0}^{n-1} \mathbb{E}_r \left[\frac{\chi_{j,s}^4}{\Pi_{j,j}^4}\right]\right)^{\varpi/4} + \sup_{j\in[p]} \sum_{s=0}^{n-1} \mathbb{E}_r \left|\frac{\chi_{j,s}}{\Pi_{j,j}}\right|^{\varpi} = O(n^{\varpi/4}(\log p)^{5\varpi/2} + n(\log p)^{5\varpi/2}) \tag{B.31}$$
$$= O\left(n^{\varpi/4}(\log p)^{5\varpi/2}\right).$$

Define $\mathrm{Z}_j = \frac{1}{\sqrt{n}\Pi_{j,j}}\mathcal{B}_j(\sum_{t=1}^n r_{j,t-1}^{*2})$. Given $\varpi > 4$, by the law of iterated expectations and (B.29) we have the unconditional moment

$$\begin{aligned}
& \sup_{j\in[p]} \mathbb{E}\left[|(\mathrm{N}_j - \mathrm{Z}_j)|^{\varpi}\right] \\
= \;& \sup_{j\in[p]} \mathbb{E}\left[\left|\frac{1}{\sqrt{n}\Pi_{j,j}}\left(\sum_{t=1}^n \chi_{j,t-1} - \mathcal{B}_j(\sum_{t=1}^n r_{j,t-1}^{*2})\right)\right|^{\varpi}\right] \\
= \;& n^{-\varpi/2}\mathbb{E}\left[\mathbb{E}\left[\left|\mathcal{B}_j(\sum_{t=1}^n \tau_{j,t-1}) - \mathcal{B}_j(\sum_{t=1}^n r_{j,t-1}^{*2})\right|^{\varpi} \middle| \{\tau_{j,t}, r_t^*\}_{t\in[n-1]}\right]\right] \\
= \;& \pi^{-1/2}2^{\varpi/2}\Gamma\left(\frac{\varpi+1}{2}\right) n^{-\varpi/2}\mathbb{E}\left(\left|\sum_{t=1}^n \widetilde{\tau}_{j,t-1}\right|^{\varpi/2}\right) \\
= \;& O(n^{-\varpi/2}n^{\varpi/4}(\log p)^c) = O(n^{-\varpi/4}(\log p)^{5\varpi/2}),
\end{aligned} \tag{B.32}$$

uniformly for all $j \in [p]$, where the order follows by (B.30) and (B.31). The above bound for the ϖth moment allows us to use the Markov inequality to bound the probability

$$\begin{aligned}
\Pr\left(\sup_{j\in[p]} |\mathrm{N}_j - \mathrm{Z}_j| > \mu \middle| \{r_{t-1}^*\}_{t\in[n]}\right) & \leq p \sup_{j\in[p]} \Pr\left(|\mathrm{N}_j - \mathrm{Z}_j| > \mu \middle| \{r_{t-1}^*\}_{t\in[n]}\right) \\
& \leq \frac{p}{\mu^{\varpi}} \sup_{j\in[p+1]} \mathbb{E}_r\left[|\mathrm{N}_j - \mathrm{Z}_j|^{\varpi}\right] \\
& \overset{\mathrm{p}}{\preccurlyeq} p \cdot n^{-\delta_{\min}\varpi/2}\mu^{-\varpi} = O_p\left(n^{\nu-\delta_{\min}\varpi/2}\mu^{-\varpi}(\log p)^{5\varpi/2}\right)
\end{aligned}$$

for any $\mu > 0$, given the order $p = O(n^{\nu})$. Taking $\varpi = \max\{4, 6\nu/\delta_{\min}\}$ and $\mu = n^{-\nu/\varpi}$, we have

$$\Pr\left(\sup_{j\in[p]} |\mathrm{N}_j - \mathrm{Z}_j| > n^{-\nu/\varpi} \middle| \{r_{t-1}^*\}_{t\in[n]}\right) \overset{\mathrm{p}}{\preccurlyeq} n^{-\nu}(\log p)^c \to 0.$$

The conditional probability is a random variable uniformly bounded in $[0, 1]$. Applying the

Bounded Convergence Theorem to the conditional probability, we have

$$\Pr\left(\sup_{j\in[p]}|\mathrm{N}_j-\mathrm{Z}_j|>n^{-\nu/\varpi}\right)\to 0$$

as $n\to\infty$. Then (B.26) follows by taking $\mathrm{Z}=(\mathrm{Z}_j)_{j\in[p]}$.

We then prove (B.27). By (B.32) we have

$$\sup_{j\in[p]}\mathbb{E}_r\left[|(\mathrm{N}_j-\mathrm{Z}_j)|^2\right]=O_p(n^{-1/2}(\log p)^{5\varpi/2}).$$

By definition of V_n above Lemma B.12,

$$\begin{aligned}
\|\mathbb{E}_r(\mathrm{Z}\mathrm{Z}^\top)-V_n\|_\infty &= \sup_{j.\ell\in[p]}|\mathbb{E}_r\left(\mathrm{N}_j\mathrm{N}_\ell-\mathrm{Z}_j\mathrm{Z}_\ell\right)|\\
&\le \sup_{j.\ell\in[p]}|\mathbb{E}_r\left(\mathrm{N}_j(\mathrm{N}_\ell-\mathrm{Z}_\ell)\right)|+\sup_{j.\ell\in[p]}|\mathbb{E}_r\left(\mathrm{Z}_\ell(\mathrm{N}_j-\mathrm{Z}_j)\right)|\\
&\le \sqrt{\sup_{j.\ell\in[p]}\mathbb{E}_r(\mathrm{N}_j^2+\mathrm{Z}_j^2)\cdot\sup_{j\in[p]}\mathbb{E}_r\left((\mathrm{N}_j-\mathrm{Z}_j)^2\right)}=O_p(n^{-1/4}(\log p)^{5\varpi/2}),
\end{aligned}$$

where the order applies (B.32). We complete the proof of Proposition B.6. □

Recall that $\eta_{\mathcal{I}}$ defined in (30) is asymptotically normal with covariance matrix $\widetilde{\Omega}_{\mathcal{I}}^{\mathrm{XD}}$ conditionally on the observed data.

Proposition B.7. *Under the conditions in Theorem 2, we have*

$$\sup_{x\in\mathbb{R}}|\Pr\left(\|\mathrm{N}_{\mathcal{I}}\|_\infty>x\right)-\Pr\left(\|\eta_{\mathcal{I}}\|_\infty>x|\mathcal{F}_n\right)|=o_p(1).$$

as $n\to\infty$.

Proof of Proposition B.7. Define $\mathrm{Z}_{\mathcal{I}}=(\mathrm{Z}_j)_{j\in\mathcal{I}}$, where $\mathrm{Z}=(\mathrm{Z}_j)_{j\in[p]}$ is the conditional Gaussian vector specified in Proposition B.6. By the Gaussian Perturbation Lemma in Belloni and Oliveira (2018, Lemma A.2), we have

$$\begin{aligned}
\sup_{x\in\mathbb{R}}|\Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty>x|\mathcal{F}_n\right)-\Pr\left(\|\eta_{\mathcal{I}}\|_\infty>x|\mathcal{F}_n\right)| &\overset{\mathrm{p}}{\preccurlyeq}\|\mathbb{E}\left(\mathrm{Z}_{\mathcal{I}}\mathrm{Z}_{\mathcal{I}}^\top|\mathcal{F}_n\right)-\mathbb{E}\left(\eta_{\mathcal{I}}\eta_{\mathcal{I}}^\top|\mathcal{F}_n\right)\|_\infty\\
&=\|\mathbb{E}\left(\mathrm{Z}_{\mathcal{I}}\mathrm{Z}_{\mathcal{I}}^\top|\mathcal{F}_n\right)-W_n\|_\infty\\
&\le\|\mathbb{E}\left(\mathrm{Z}_{\mathcal{I}}\mathrm{Z}_{\mathcal{I}}^\top|\mathcal{F}_n\right)-W_n\|_\infty+\|V_n-W_n\|_\infty\\
&\overset{\mathrm{p}}{\preccurlyeq}\frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}=o_p(1),
\end{aligned}\qquad\text{(B.33)}$$

where the last inequality applies (B.27) and (B.148). In addition, by Belloni and Oliveira (2018, Corollary 3.1)

$$
\begin{aligned}
\Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x - \|\mathrm{N}_{\mathcal{I}} - \mathrm{Z}_{\mathcal{I}}\|_\infty | \mathcal{F}_n\right) &= \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + \|\mathrm{N}_{\mathcal{I}} - \mathrm{Z}_{\mathcal{I}}\|_\infty O_p(\sqrt{\log p}) \\
&= \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + o_p(1),
\end{aligned}
$$

where the second line applies Proposition B.6. where the first asymptotic equivalence applies the error bounds of the covariance matrices in Lemma B.12. Therefore,

$$
\begin{aligned}
\Pr\left(\|\mathrm{N}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) &\leq \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x - \|\mathrm{N}_{\mathcal{I}} - \mathrm{Z}_{\mathcal{I}}\|_\infty | \mathcal{F}_n\right) \\
&= \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + o_p(1).
\end{aligned} \tag{B.34}
$$

In a parallel way, we can show that the other side of the inequality in the first line of (B.34), and therefore

$$
\Pr\left(\|\mathrm{N}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) = \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x + \|\mathrm{N}_{\mathcal{I}} - \mathrm{Z}_{\mathcal{I}}\|_\infty | \mathcal{F}_n\right) = \Pr\left(\|\mathrm{Z}_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + o_p(1). \tag{B.35}
$$

Then Proposition B.7 follows by (B.33) and (B.35). □

B.2.3 Proof of Results in the Main Text

Proof of Theorem 2. Note that uniformly of all $x \in \mathbb{R}$,

$$
\begin{aligned}
\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > x\right) &\leq \Pr\left(\|\mathrm{N}_{\mathcal{I}}\|_\infty > x - \max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}|\right) \\
&\leq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x - \max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}| \,\middle|\, \mathcal{F}_n\right) + o_p(1) \\
&\leq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + O\left(\max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}| \log p\right) + o_p(1) \\
&= \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x | \mathcal{F}_n\right) + o_p(1)
\end{aligned}
$$

where the second inequality applies Proposition B.7, and the third inequality applies Corollary 3.1 of Belloni and Oliveira (2018), and the last step applies Proposition B.5. The other side of the inequality can be proved in a parallel way. Then (35) is verified. For the second conclusion, note that the conditional probability, as a random variable, is strictly bounded in $[0, 1]$. Then the second conclusion can be verified by taking $x = \mathrm{cv}_{\mathcal{G}}(\alpha)$ and utilizing the bounded convergence theorem. □

Proof of Theorem 2. Note that uniformly of all $x \in \mathbb{R}$,

$$\begin{aligned}
\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > x\right) &\le \Pr\left(\|\mathrm{N}_{\mathcal{I}}\|_\infty > x - \max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}|\right) \\
&\le \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x - \max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}| \,\middle|\, \mathcal{F}_n\right) + o_p(1) \\
&\le \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x|\mathcal{F}_n\right) + O\left(\max_{j\in[p]} |\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}| \log p\right) + o_p(1) \\
&= \Pr\left(\|\eta_{\mathcal{I}}\|_\infty > x|\mathcal{F}_n\right) + o_p(1)
\end{aligned}$$

where the second inequality applies Proposition B.7, and the third inequality applies Corollary 3.1 of Belloni and Oliveira (2018), and the last step applies Proposition B.5. The other side of the inequality can be proved in a parallel way. Then (35) is verified. For the second conclusion, note that the conditional probability, as a random variable, is strictly bounded in $[0, 1]$. Then the second conclusion can be verified by taking $x = \mathrm{cv}_{\mathcal{I}}(\alpha)$ and utilizing the bounded convergence theorem. □

Proof of Theorem 3. Define $j_{\max} = \max_{j\in\mathcal{I}} \left|\sqrt{n^{(\theta\wedge\gamma_j)}}(\beta_j - \beta_{0,j})\right|$. Following the proof of Theorem 2, we can show that the scaled standard error $\sqrt{n^{1+(\theta\wedge\gamma_{j_{\max}})}}\widehat{\omega}^{\mathrm{XD}}_{j_{\max},j_{\max}}$ converges in probability to a positive constant when $\gamma_{j_{\max}} = 0$, and converges in distribution to a stable law when $\gamma_{j_{\max}} = 1$. Define

$$\mathcal{W}_{\max} = \{\sqrt{n^{1+(\theta\wedge\gamma_{j_{\max}})}}\widehat{\omega}^{\mathrm{XD}}_{j_{\max},j_{\max}} \le (\log|\mathcal{I}|)^{1/4}\},$$

and we have $\Pr(\mathcal{W}_{\max}) \to 1$ as $n \to \infty$, since $|\mathcal{I}| \to \infty$. In addition, under $\mathcal{W}_{\max}$ and when $\beta^* \in \mathcal{U}_{\mathcal{I}}(c_0)$,

$$\begin{aligned}
\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty &\ge \max_{j\in\mathcal{I}} \left|\frac{\beta_j^* - \beta_{0,j}}{\widehat{\omega}^{\mathrm{XD}}_{j,j}}\right| - \|\mathrm{N}_{\mathcal{I}}\|_\infty \\
&\ge \left|\frac{\sqrt{n^{1+(\theta\wedge\gamma_{j_{\max}})}}(\beta^*_{j_{\max}} - \beta_{0,j_{\max}})}{\sqrt{n^{(\theta\wedge\gamma_{j_{\max}})}}\widehat{\omega}^{\mathrm{XD}}_{j_{\max},j_{\max}}}\right| - \|\mathrm{N}_{\mathcal{I}}\|_\infty \\
&\ge c_0(\log|\mathcal{I}|)^{3/4} - \|\mathrm{N}_{\mathcal{I}}\|_\infty.
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > \mathrm{cv}_{\mathcal{I}}(\alpha)\right)\\
\geq &\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > \mathrm{cv}_{\mathcal{I}}(\alpha), \mathcal{W}_{\max}\right)\\
\geq &\Pr\left(c_0(\log|\mathcal{I}|)^{3/4} > \mathrm{cv}_{\mathcal{I}}(\alpha) + \|\mathrm{N}_{\mathcal{I}}\|_\infty + \max_{j\in[p]}\left|\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}\right|, \mathcal{W}_{\max}\right)\\
\geq &\Pr\left(c_0(\log|\mathcal{I}|)^{3/4} > \mathrm{cv}_{\mathcal{I}}(\alpha) + \|\mathrm{N}_{\mathcal{I}}\|_\infty + \max_{j\in[p]}\left|\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}\right|\right) - \Pr(\mathcal{W}_{\max}^c)\\
\geq &\Pr\left(c_0(\log|\mathcal{I}|)^{3/4} > \mathrm{cv}_{\mathcal{I}}(\alpha) + \|\eta_{\mathcal{I}}\|_\infty + \max_{j\in[p]}\left|\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}\right| \middle| \mathcal{F}_n\right) + o_p(1)\\
\geq &\Pr\left(c_0(\log|\mathcal{I}|)^{3/4} > \mathrm{cv}_{\mathcal{I}}(\alpha) + \|\eta_{\mathcal{I}}\|_\infty \middle| \mathcal{F}_n\right) + O\left(\max_{j\in[p]}\left|\mathrm{Err}_{j,1} + \mathrm{Err}_{j,2}\right|\log p\right) + o_p(1)\\
\geq &\Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0(\log|\mathcal{I}|)^{3/4} - \mathrm{cv}_{\mathcal{I}}(\alpha)|\mathcal{F}_n\right) + o_p(1).
\end{aligned}
$$

The bounded convergence theorem applied to the conditional probability implies

$$
\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > \mathrm{cv}_{\mathcal{I}}(\alpha)\right) \geq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0\log|\mathcal{I}| - \mathrm{cv}_{\mathcal{G}}(\alpha)\right) + o(1). \tag{B.36}
$$

Recall that $\mathrm{cv}_{\mathcal{G}}(\alpha)$ is the $(1-\alpha)$th quantile of a maximum of $|\mathcal{I}|$ conditionally normal distribution with individual variance equaling one. Therefore, if we define

$$
\mathcal{CV} = \{\mathrm{cv}_{\mathcal{I}}(\alpha) \leq C\sqrt{\log|\mathcal{I}|}\}
$$

for some sufficiently large absolute constant C, we have $\Pr(\mathcal{CV}) = 1$. In addition, by Lemma 6 of Cai et al. (2014),

$$
\Pr\left(\|\eta_{\mathcal{I}}\|_\infty^2 - 2\log|\mathcal{I}| + \log\log|\mathcal{I}| < x|\mathcal{F}_n\right) \to \exp\left(-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right).
$$

Taking $x = \log p$, we have

$$
\Pr\left(\|\eta_{\mathcal{I}}\|_\infty^2 < 3\log|\mathcal{I}| - \log\log|\mathcal{I}| \middle| \mathcal{F}_n\right) \to 1.
$$

The bounded convergence theorem implies

$$
\Pr\left(\|\eta_{\mathcal{I}}\|_\infty^2 < 3\log|\mathcal{I}| - \log\log|\mathcal{I}|\right) \to 1.
$$

Therefore,

$$\begin{aligned}
\Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0(\log|\mathcal{I}|)^{3/4} - \mathrm{cv}_{\mathcal{G}}(\alpha)\right) &\geq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0(\log|\mathcal{I}|)^{3/4} - \mathrm{cv}_{\mathcal{G}}(\alpha), \mathcal{CV}\right) \\
&\geq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0(\log|\mathcal{I}|)^{3/4} - C\sqrt{\log|\mathcal{I}|}, \mathcal{CV}\right) \\
&\geq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty < c_0(\log|\mathcal{I}|)^{3/4} - C\sqrt{\log|\mathcal{I}|}\right) - \Pr(\mathcal{CV}^c) \\
&\geq \Pr\left(\|\eta_{\mathcal{I}}\|_\infty < \sqrt{3\log|\mathcal{I}| - \log\log|\mathcal{I}|}\right) - o(1) \to 1.
\end{aligned}$$

By (B.36), we have $\Pr\left(\|t_{\mathcal{I}}^{\mathrm{XD}}\|_\infty > \mathrm{cv}_{\mathcal{I}}(\alpha)\right) \to 1$. We complete the proof of Theorem 3. $\square$

B.3 Technical Lemmas

B.3.1 Technical Lemmas for Consistency

The lemmas in this section are used in not only the proofs for consistency of Slasso, but also the asymptotic distribution of the XDlasso. A mildly integrated (MI) instrumental variable $z_{j,t}$ defined in (21) is involved in the latter. Therefore, the lemmas in this section cover MI series that are more general than the settings in the main text.

Specifically, we suppose there is an additional group $\mathcal{G}^{(2)} = \{j \in [p] : \gamma_j = \gamma^{(2)}\}$, such that

$$x_{j,t} = \left(1 + \frac{c_j^*}{n^{\gamma^{(2)}}}\right) x_{j,t-1} + e_{j,t}$$

for $j \in \mathcal{G}^{(2)}$. The absolute constant $\gamma^{(2)} \in (0,1)$ and thus $x_{j,t}$ for any $j \in \mathcal{G}^{(2)}$ is MI. The notations $\mathcal{G}^{(1)} = \{j \in [p] : \gamma_j = 1\}$ and $\mathcal{G}^{(0)} = \{j \in [p] : \gamma_j = 0\}$ remains their original meanings in the main text, denoting the index sets for LUR and stationary regressors, respectively. Define $\gamma^{(0)} = 0$ and $\gamma^{(1)} = 1$.

Lemma B.1. *Under Assumptions 1-5,*

$$\sup_{j \in \mathcal{G}^{(k)}, t \in [n]} |x_{j,t}| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{n^{\gamma^{(k)}}} (\log p)^{3/2}. \tag{B.37}$$

Proof of Lemma B.1. For $\gamma^{(k)} = 1$, lemma B.1 follows by (B.12) of MS. For $\gamma^{(k)} = 0$, lemma B.1 follows by the sub-exponential tail.

For $0 < \gamma^{(k)} < 1$, we follow the proof of (A.14) of GLMS. Recall that for any $j \in \mathcal{G}^{(k)}$, the regressor $x_{j,t}$ is an AR(1) process with the coefficient $\rho_j^* = 1 - c_j^*/n^{\gamma_j}$. The sequence $x_{j,t} = \sum_{s=1}^t \rho_j^{*s} e_{j,t-s}$ is a partial sum of α-mixing sup-exponential components $e_{j,t-s}$ weighted by ρ_j^s. Also, all regressors in the same group $\mathcal{G}^{(k)}$ shares the same degree of persistence

$\gamma_j = \gamma^{(k)}$. Define

$$a_{k,n} := \lfloor n^{\gamma^{(k)}} (\log p)^2 \rfloor.$$

Note that $\rho_j^s e_{j,t-s}$ is sub-exponential with an exponentially decaying α-mixing coefficient, and thus $\zeta_{j,t}$ is the partial sum of t observations from a sub-exponential and α-mixing time series. By MS24's Lemma B.2, there exists an absolute constant C such that

$$\sup_{j\in\mathcal{G}^{(k)}} \sup_{t\le a_{j,n}} |x_{j,t}| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{a_{k,n}\cdot \log p} = O\left[n^{\gamma^{(k)}/2} (\log p)^{3/2}\right]. \tag{B.38}$$

In addition, when $t > a_{k,n}$,

$$\begin{aligned} \sup_{j\in\mathcal{G}^{(k)}} |x_{j,t}| &\le \left|\sum_{s\le a_{k,n}} \rho_j^{*s} e_{j,t-s}\right| + \left|\sum_{a_{k,n}<s\le t} \rho_j^{*s} e_{j,t-s}\right| \\ &\le \left|\sum_{s\le a_{k,n}} \rho_j^{*s} e_{j,t-s}\right| + \rho_j^{a_n} \left|\sum_{0<s\le t-a_{k,n}} \rho_j^{*(s-a_{k,n})} e_{j,t-s+a_{k,n}}\right|. \end{aligned} \tag{B.39}$$

By the same arguments for (B.38), we bound the two sums on the right-hand side of (B.39) by

$$\sup_{j\in\mathcal{G}^{(k)}} \sup_{a_{k,n}<t\le n} \left|\sum_{s\le a_{k,n}} \rho_j^{*s} e_{j,t-s}\right| \overset{\mathrm{p}}{\preccurlyeq} n^{\gamma^{(k)}/2} (\log p)^{3/2}, \tag{B.40}$$

and

$$\sup_{j\in\mathcal{G}^{(k)}} \sup_{a_{k,n}<t\le n} \left|\sum_{0<s\le t-a_{k,n}} \rho_j^{*(s-a_{k,n})} e_{j,t-s+a_{k,n}}\right| \overset{\mathrm{p}}{\preccurlyeq} \sqrt{(n-a_{k,n})\cdot \log p}. \tag{B.41}$$

Besides, under the assumption $p \ge n^{\nu_1}$, the sequence

$$\sup_{j\in\mathcal{G}^{(k)}} \rho_j^{a_{k,n}} = \left(1 + c_j^*/n^{\gamma^{(k)}}\right)^{\lfloor n^{\gamma^{(k)}} (\log p)^2 \rfloor} = O\left(\exp\left(-|c_j^*|(\log p)^2\right)\right) = O\left(p^{-\underline{c}\log p}\right) \tag{B.42}$$

converges to zero faster than the reciprocal of any polynomial function of n. Thus, by (equation (B.41)) and (equation (B.42)),

$$\begin{aligned} \sup_{j\in\mathcal{G}^{(k)}} \sup_{a_{k,n}<t\le n} \rho_j^{a_{k,n}} \left|\sum_{0<s\le t-a_{k,n}} \rho_j^{s-a_{k,n}} e_{j,t-s+a_{k,n}}\right| &\overset{\mathrm{p}}{\preccurlyeq} p^{-\underline{c}\log p} \cdot \sqrt{(n-a_n)\cdot \log p} \\ &= o\left(n^{\tau/2} (\log p)^{3/2}\right). \end{aligned} \tag{B.43}$$

By (B.39), (B.40), and (B.43), we have

$$\sup_{j\in\mathcal{G}^{(k)}}\sup_{a_{k,n}<t\le n}|x_{j,t}| \overset{\mathrm{p}}{\precsim} Cn^{\gamma^{(k)}/2}(\log p)^{3/2}, \tag{B.44}$$

and then (B.37) follows by (B.38) and (B.44). We complete the proof of lemma B.1. $\square$

Lemma B.2. *Under Assumptions 1-5,*

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}(x_{j,t-1}e_{\ell,t}-\mathbb{E}(x_{j,t-1}e_{\ell,t}))\right| \overset{\mathrm{p}}{\precsim} \frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-\gamma^{(k)}}}}. \tag{B.45}$$

In addition, when $x_t^{(k)}$ is MI with $0<\gamma^{(k)}<1$, we have

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}(x_{j,t-1}e_{\ell,t}-\Lambda_{j,\ell})\right| \overset{\mathrm{p}}{\precsim} \frac{(\log p)^{\frac{2}{r}}}{n^{\gamma^{(k)}}}+\frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-\gamma^{(k)}}}}, \tag{B.46}$$

where $\Lambda_{j,\ell}=\sum_{d=1}^{\infty}\mathbb{E}(e_{j,t}e_{\ell,t-d})$.

Proof of Lemma B.2. **We first show (B.45).** Let $G=\lfloor(2c_\alpha^{-1}\log(np))^{1/r}\rfloor$. Notice we have the following decomposition

$$\sum_{t=1}^{n}x_{j,t-1}e_{\ell,t}=\sum_{t=1}^{G}x_{j,t-1}e_{\ell,t}+\sum_{t=G+1}^{n}e_{\ell,t}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}+\rho_j^{*G-1}\sum_{t=G+1}^{n}x_{j,t-G}e_{\ell,t}. \tag{B.47}$$

Therefore,

$$\sum_{t=1}^{n}(x_{j,t-1}e_{\ell,t}-\mathbb{E}(x_{j,t-1}e_{\ell,t}))=T_{1,n}^{(j,\ell)}+T_{2,n}^{(j,\ell)}+T_{3,n}^{(j,\ell)}+T_{4,n}^{(j,\ell)},$$

where

$$T_{1,n}^{(j,\ell)}=\sum_{t=1}^{G}x_{j,t-1}e_{\ell,t},$$

$$T_{2,n}^{(j,\ell)}=\sum_{t=G+1}^{n}e_{\ell,t}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}-\mathbb{E}\left(\sum_{t=G+1}^{n}e_{\ell,t}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}\right),$$

$$T_{3,n}^{(j,\ell)}=\rho_j^{*G-1}\sum_{t=G+1}^{n}x_{j,t-G}e_{\ell,t},$$

$$T_{4,n}^{(j,\ell)}=\sum_{t=1}^{n}\mathbb{E}(x_{j,t-1}e_{\ell,t})-\mathbb{E}\left(\sum_{t=G+1}^{n}e_{\ell,t}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}\right).$$

Define $T_i=\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}T_{i,n}^{(j,\ell)}$ for $i\in\{1,2,3,4\}$. We analyze one by one the four terms.

Bound T_1. Note that $\max_{\ell\in[p]}|\sum_{t=1}^{G}(|e_{\ell,t}|-\mathbb{E}|e_{\ell,t}|)|\overset{\mathrm{p}}{\preccurlyeq}\sqrt{G\log p}$ by the bound of the partial sum of weakly dependent time series. We deduce by the triangular inequality

$$\begin{aligned}T_1 &\le \max_{j\in\mathcal{G}^{(k)}}|x_{j,t-1}|\cdot\max_{\ell\in[p]}\sum_{t=1}^{G}|e_{\ell,t}|\\ &\overset{\mathrm{p}}{\preccurlyeq}\sqrt{n^{\gamma^{(k)}}(\log p)^3}\left[\max_{\ell\in[p]}\left|\sum_{t=1}^{G}(|e_{\ell,t}|-\mathbb{E}|e_{\ell,t}|)\right|+\max_{\ell\in[p]}\sum_{t=1}^{G}\mathbb{E}|e_{\ell,t}|\right]\\ &\overset{\mathrm{p}}{\preccurlyeq}\sqrt{n^{\gamma^{(k)}}(\log p)^3}\left(\sqrt{G\log p}+G\right)=o(n^{\frac{1+\gamma^{(k)}}{2}}),\end{aligned}\tag{B.48}$$

where the second row applies Lemma B.1.

Bound T_2. Note that

$$\begin{aligned}T_2 &= \max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\sum_{t=G+1}^{n}e_{\ell,t}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}-\mathbb{E}\left(\sum_{t=G+1}^{n}e_{\ell,t-1}\sum_{r=t-G+1}^{t-1}\rho_j^{*t-1-r}e_{j,r}\right)\right|\\ &\le \max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\sum_{d=1}^{G-1}\left|\sum_{t=G+1}^{n}\rho_j^{*d-1}\left(e_{\ell,t}e_{j,t-d}-\mathbb{E}(e_{\ell,t}e_{j,t-d})\right)\right|.\end{aligned}\tag{B.49}$$

Note that for any $t\in[n]$

$$\rho_j^{*t}\le\left(1+\frac{\bar{c}}{n}\right)^n\le \mathrm{e}^{\bar{c}},\tag{B.50}$$

where $\bar{c}>0$ is an absolute constant specified in Assumption 5. Therefore, $\rho_j^{*d-1}e_{j,t-d}$ is a stationary and strongly mixing sequence. By the same arguments to bound "T_{22}" on Page 8 of the online appendix to MS, we have

$$T_2=O_p(G^2\sqrt{n\log(p^2G^2)})=o(n^{\frac{1+\gamma^{(k)}}{2}}).$$

Bound T_3. Following the procedures to bound "T_3" starting from Page 9 of the online appendix to MS, we can show

$$T_3\overset{\mathrm{p}}{\preccurlyeq}\sqrt{n^{1+\gamma^{(k)}}}(\log p)^{2+\frac{1}{2r}}.$$

We only need to change the event $\mathcal{X}_t=\{\max_{j\in[p_x]}|X_{jt}|\le C_X\sqrt{n\log p}\}$ below (B.16) of MS into $\mathcal{X}_t^{(k)}=\{\max_{j\in[p_x]}|X_{jt}|\le C_X\sqrt{n^{\gamma^{(k)}}(\log p)^3}\}$, which is a high-probability event in view of (B.1).

Bound T_4. By the decomposition (B.47), we have

$$
\begin{aligned}
T_4 &= \max_{j\in\mathcal{G}^{(k)},\ell\in[p]} \left| \sum_{t=1}^{n} \mathbb{E}(x_{j,t-1}e_{\ell,t}) - \mathbb{E}\left(\sum_{t=G+1}^{n} e_{\ell,t} \sum_{r=t-G+1}^{t-1} \rho_j^{*t-1-r} e_{j,r} \right) \right| \\
&= \max_{j\in\mathcal{G}^{(k)},\ell\in[p]} \left| \mathbb{E}\left(\sum_{t=1}^{G} x_{j,t-1}e_{\ell,t} \right) + \mathbb{E}\left(\sum_{t=G+1}^{n} x_{j,t-G}e_{\ell,t} \right) \right|. \qquad \text{(B.51)}
\end{aligned}
$$

Note that $\mathbb{E}(e_{j,s}e_{\ell,t-1})$ is uniformly bounded for any $j,\ell\in[p]$ and $s,t\in[n]$. Therefore, by the fact that $\mathbb{E}(e_{j,s}e_{\ell,t-1}) = O(1)$ uniformly for all $j,\ell\in[p]$ and $s,t\in[n]$, we have

$$
\begin{aligned}
\left| \mathbb{E}\left(\sum_{t=1}^{G} x_{j,t-1}e_{\ell,t} \right) \right| &= \sum_{t=1}^{G}\sum_{s=0}^{t-1} \rho_j^{*t-1-s} \left|\mathbb{E}(e_{j,s}e_{\ell,t})\right| \\
&\leq \mathrm{e}^{\bar{c}} \sum_{t=1}^{G}\sum_{s=0}^{t-1} \left|\mathbb{E}(e_{j,s}e_{\ell,t})\right| = O(G^2), \qquad \text{(B.52)}
\end{aligned}
$$

where the second row applies (B.50). In addition,

$$
\left| \mathbb{E}\left(\sum_{t=G+1}^{n} x_{j,t-G}e_{\ell,t} \right) \right| = \left| \sum_{t=G+1}^{n} \sum_{s=1}^{t-G} \rho_j^{*s} \mathbb{E}\left(e_{j,t-G-s}e_{\ell,t}\right) \right| \qquad \text{(B.53)}
$$

To bound the right-hand side of (B.53), define the ρ-mixing coefficients of two generic σ-fields $\mathcal{A}$ and $\mathcal{B}$ as

$$
\rho(\mathcal{A},\mathcal{B}) := \sup_{X\in\mathcal{A},Y\in\mathcal{B}} \left|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y\right| / \sqrt{\mathbb{E}X^2\mathbb{E}Y^2} \quad \text{for } \mathbb{E}X^2, \mathbb{E}Y^2 < \infty. \qquad \text{(B.54)}
$$

For $d\in\mathbb{N}$, the ρ-mixing coefficients of $e = (e_t)_{t\in\mathbb{Z}}$ is defined as

$$
\rho(d) := \sup_{s\in\mathbb{Z}} \rho(\sigma((e_t)_{t\leq s}), \sigma((e_t)_{t\geq s+d})), \qquad \text{(B.55)}
$$

where $\sigma((e_t)_{t\leq s})$ is the σ-filed generated by $(e_t)_{t\leq s}$. By (B.4) of MS and $G = \lfloor (2c_\alpha^{-1}\log(np))^{1/r} \rfloor$,

$$
\begin{aligned}
\rho(G) &\leq C_\alpha \exp(-c_\alpha G^r) \\
&\leq C_\alpha \exp(-2\log(np)) = O((np)^{-2}).
\end{aligned}
$$

Therefore,

$$
\left|\mathbb{E}\left(e_{j,t-G-s}e_{\ell,t-1}\right)\right| \leq \rho(G)\sqrt{\mathbb{E}\left(e_{j,t-G-s}^2\right)\mathbb{E}\left(e_{j,t-G-s}^2\right)} = O((np)^{-2})
$$

uniformly for all $j, \ell \in [p]$ and $s, t \in [n]$. Then by (B.53),

$$\left|\sum_{t=G+1}^{n}\sum_{s=1}^{t-G}\rho_j^{*s}\mathbb{E}\left(e_{j,t-G-s}e_{\ell,t}\right)\right| \le \mathrm{e}^{\bar{c}}\left|\sum_{t=G+1}^{n}\sum_{s=1}^{t-G}\mathbb{E}\left(e_{j,t-G-s}e_{\ell,t}\right)\right|$$
$$= O(n^2 \cdot (np)^{-2}) = o(1). \tag{B.56}$$

By (B.51), (B.52), and (B.56),

$$T_4 = O(G^2). \tag{B.57}$$

where the second row applies Collecting the stochastic order of T_1, T_2, T_3, and T_4, we have

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}\left(x_{j,t-1}e_{\ell,t}-\mathbb{E}(x_{j,t-1}e_{\ell,t})\right)\right| \le & \frac{T_1+T_2+T_3+T_4}{n} \\ = & \frac{o_p(n^{\frac{1+\gamma^{(k)}}{2}})+o_p(n^{\frac{1+\gamma^{(k)}}{2}})+T_3+O(G^2)}{n} \\ \overset{\mathrm{p}}{\preccurlyeq} & \frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-\gamma^{(k)}}}}.\end{aligned}$$

We complete the proof of (B.46).

<u>**We then show (B.46).**</u> It suffices to show

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\mathbb{E}(x_{j,t-1}e_{\ell,t})-\Lambda_{\ell,j}\right)\right| = O\left(\frac{(\log p)^{\frac{2}{r}}}{n^{\gamma^{(k)}}}\right). \tag{B.58}$$

Note that

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\mathbb{E}(x_{j,t-1}e_{\ell,t})-\Lambda_{\ell,j}\right)\right| \le & T_4+\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\mathbb{E}\left(\frac{1}{n}\sum_{t=G+1}^{n}e_{\ell,t}\sum_{s=t-G+1}^{t-1}\rho_j^{*t-1-s}e_{j,s}\right)-\Lambda_{\ell,j}\right| \\ = & O(G^2)+\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\mathbb{E}\left(\frac{1}{n}\sum_{t=G+1}^{n}e_{\ell,t}\sum_{s=t-G+1}^{t-1}\rho_j^{*t-1-s}e_{j,s}\right)-\Lambda_{\ell,j}\right|,\end{aligned} \tag{B.59}$$

where the second step applies $T_4 = O(G^2)$ by (B.57). The terms inside the absolute value operator of the second term can be further decomposed as

$$\mathbb{E}\left(\frac{1}{n}\sum_{t=G+1}^{n} e_{\ell,t}\sum_{s=t-G+1}^{t-1}\rho_j^{*t-1-s}e_{j,s}\right)-\Lambda_{\ell,j}=\frac{1}{n}\sum_{t=G+1}^{n}\sum_{s=t-G+1}^{t-1}(\rho_j^{*t-1-s}-1)\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)$$
$$+\frac{1}{n}\sum_{t=G+1}^{n}\sum_{s=t-G+1}^{t-1}\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)-\Lambda_{\ell,j}. \tag{B.60}$$

The first term on the right-hand side of (B.60) is bounded by

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=G+1}^{n}\sum_{s=t-G+1}^{t-1}(\rho_j^{*t-1-s}-1)\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)\right|=O\left(\max_{t\in[n]}\sum_{s=t-G+1}^{t-1}\left|\rho_j^{*t-1-s}-1\right|\right)$$
$$=O\left(\max_{t\in[n]}\sum_{d=1}^{G-2}\left|\rho_j^{*d}-1\right|\right) \tag{B.61}$$

where the first line applies the uniform bound of $\mathbb{E}\left(e_{\ell,t-1}e_{j,s}\right)$. For any $1\le d\le G-2$ and $j\in\mathcal{G}^{(k)}$,

$$|\rho_j^{*d}-1|=\frac{|c_j^*|}{n^{\gamma_j}}\cdot|\rho_j^{*d-1}+\cdots+\rho_j^*+1|=O(G/n^{\gamma^{(k)}})$$

where the last step applies (B.50). Then by (B.61),

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\frac{1}{n}\sum_{t=G+1}^{n}\sum_{s=t-G+1}^{t-1}(\rho_j^{*t-1-s}-1)\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)\right|=O(G^2/n^{\gamma^{(k)}}), \tag{B.62}$$

which bounds the first term on the right-hand side of (B.60). For the second term,

$$\left|\frac{1}{n}\sum_{t=G+1}^{n}\sum_{s=t-G+1}^{t-1}\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)-\Lambda_{\ell,j}\right|\le\frac{1}{n}\left|\sum_{t=G+1}^{n}\left(\sum_{s=t-G+1}^{t-1}\mathbb{E}\left(e_{\ell,t}e_{j,s}\right)-\Lambda_{\ell,j}\right)\right|+(1-\frac{n-G}{n})|\Lambda_{\ell,j}|$$
$$=\frac{1}{n}\sum_{t=G+1}^{n}\sum_{d=G}^{\infty}\left|\mathbb{E}\left(e_{\ell,t}e_{j,t-d}\right)\right|+O(G/n). \tag{B.63}$$

Recall the definition of ρ-mixing coefficient in (B.54) and (B.55). By (B.4) of MS and $G=\lfloor(2c_\alpha^{-1}\log(np))^{1/r}\rfloor$,

$$\sum_{d=G}^{\infty}\left|\mathbb{E}\left(e_{\ell,t}e_{j,t-d}\right)\right|\le\sqrt{\mathbb{E}(e_{\ell,t}^2)\mathbb{E}(e_{j,t-d}^2)}\sum_{d=G}^{\infty}\rho(d)=O\left(\sum_{d=G}^{\infty}\exp(-c_\alpha d^r)\right)$$
$$=O\left(\exp(-c_\alpha G^r/2)\right)$$
$$=O((np)^{-1}),$$

where the second row applies (B.78) in MS. By (B.63),

$$\left| \frac{1}{n} \sum_{t=G+1}^{n} \sum_{s=t-G+1}^{t-1} \mathbb{E}\left(e_{\ell,t} e_{j,s}\right) - \Lambda_{\ell,j} \right| = O(G/n) + O((np)^{-1}) = O(G/n), \tag{B.64}$$

which bounds the second term on the right-hand side of (B.60). By (B.62) and (B.64),

$$\begin{aligned} \max_{j \in \mathcal{G}^{(k)}, \ell \in [p]} \left| \mathbb{E}\left(\frac{1}{n} \sum_{t=G+1}^{n} e_{\ell,t} \sum_{s=t-G+1}^{t-1} \rho_j^{*t-1-s} e_{j,s} \right) - \Lambda_{\ell,j} \right| &\overset{\mathrm{p}}{\preccurlyeq} O(G^2/n^{\gamma^{(k)}}) + O(G/n) \\ &= O(G^2/n^{\gamma^{(k)}}) = O((\log p)^{\frac{2}{r}}/n^{\gamma^{(k)}}). \end{aligned} \tag{B.65}$$

Thus, (B.58) is implied by (B.59) and (B.65). We complete the proof of (B.46) and end the proof of Lemma B.2. □

Lemma B.3. *Under Assumptions 1-5, when* $0 \le \gamma^{(k)} < 1$,

$$\left\| \frac{1}{n} \sum_{t=1}^{n} x_{t-1}^{(k)} e_{t-1}^{(k)\top} \right\|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} 1. \tag{B.66}$$

when $\gamma^{(k)} = 1$, *we have*

$$\left\| \frac{1}{n} \sum_{t=1}^{n} x_{t-1}^{(k)} e_{t-1}^{(k)\top} \right\|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} (\log p)^{1+\frac{1}{2r}}. \tag{B.67}$$

Proof of Lemma B.2. For stationary and MI regressors with $0 \le \gamma^{(k)} < 1$, (B.66) is a direct corollary of Lemma B.2. For LUR regressors with $\gamma^{(k)} = 1$, (B.67) is deduced in the proof of MS24's Theorem 3. □

Lemma B.4. *x Under Assumptions 1-5,*

$$\|n^{-1} \sum_{t=1}^{n} x_{t-1}^{(k)} x_{t-1}^{(m)\top} \|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} n^{(\gamma^{(k)} \wedge \gamma^{(m)})} (\log p)^{1+\frac{1}{2r}}.$$

Proof of Lemma B.4. For any $j \in \mathcal{G}^{(k)}$ and $\ell \in \mathcal{G}^{(m)}$,

$$x_{j,t-1} x_{\ell,t} = \rho_\ell^* x_{j,t-1} x_{\ell,t-1} + x_{j,t-1} e_{\ell,t}.$$

By $x_{j,t} x_{\ell,t} = \rho_j^* x_{j,t-1} x_{\ell,t} + e_{j,t} x_{\ell,t}$, we obtain

$$x_{j,t} x_{\ell,t} = \rho_j^* \left(\rho_\ell^* x_{j,t-1} x_{\ell,t-1} + x_{j,t-1} e_{\ell,t}\right) + e_{j,t} x_{\ell,t}. \tag{B.68}$$

Summing up both sides of (B.68) and using the fact that $\sum_{t=1}^n x_{j,t}x_{\ell,t} = \sum_{t=1}^n x_{j,t-1}x_{\ell,t-1} - x_{j,0}x_{\ell,0} + x_{j,n}x_{\ell,n}$, we deduce

$$\sum_{t=1}^n x_{j,t-1}x_{\ell,t-1} - x_{j,0}x_{\ell,0} + x_{j,n}x_{\ell,n} = \rho_j^* \left(\rho_\ell^* \sum_{t=1}^n x_{j,t-1}x_{\ell,t-1} + \sum_{t=1}^n x_{j,t-1}e_{\ell,t}\right) + \sum_{t=1}^n e_{j,t}x_{\ell,t}.$$

It can be further arranged into

$$(1-\rho_j^*\rho_\ell^*)\sum_{t=1}^n x_{j,t-1}x_{\ell,t-1} = (x_{j,0}x_{\ell,0} - x_{j,n}x_{\ell,n}) + \rho_j^* \sum_{t=1}^n x_{j,t-1}e_{\ell,t} + \sum_{t=1}^n e_{j,t}x_{\ell,t}. \tag{B.69}$$

By (B.1),

$$\sup_{j\in\mathcal{G}^{(k)}} \sup_{\ell\in\mathcal{G}^{(m)}} |x_{j,0}x_{\ell,0} - x_{j,n}x_{\ell,n}| \overset{\mathrm{p}}{\preccurlyeq} n^{\frac{\gamma^{(k)}+\gamma^{(m)}}{2}} (\log p)^3. \tag{B.70}$$

By (B.66) and the fact that $|\rho_j^*| \le 2$,

$$\sup_{j\in\mathcal{G}^{(k)}} \sup_{\ell\in\mathcal{G}^{(m)}} \left(\left|\rho_j^* \sum_{t=1}^n x_{j,t-1}e_{\ell,t}\right| + \left|\sum_{t=1}^n e_{j,t}x_{\ell,t}\right|\right) \overset{\mathrm{p}}{\preccurlyeq} n(\log p)^{1+\frac{1}{2r}}. \tag{B.71}$$

By (B.69), (B.70), and (B.71),

$$\begin{aligned}
&\sup_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} \left|(1-\rho_j^*\rho_\ell^*)\sum_{t=1}^n x_{j,t-1}x_{\ell,t-1}\right| \\
\le& \sup_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} |x_{j,0}x_{\ell,0} - x_{j,n}x_{\ell,n}| + \sup_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} \left(\left|\rho_j^* \sum_{t=1}^n x_{j,t-1}e_{\ell,t}\right| + \left|\sum_{t=1}^n e_{j,t}x_{\ell,t}\right|\right) \\
\overset{\mathrm{p}}{\preccurlyeq}& \, n^{\frac{\gamma^{(k)}+\gamma^{(m)}}{2}} (\log p)^2 + n(\log p)^{1+\frac{1}{2r}} \\
\le& 2n(\log p)^{1+\frac{1}{2r}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sup_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} \left|\sum_{t=1}^n x_{j,t-1}x_{\ell,t-1}\right| &\overset{\mathrm{p}}{\preccurlyeq} \frac{n(\log p)^{1+\frac{1}{2r}}}{\inf_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}}(1-\rho_j^*\rho_\ell^*)} \\
&= O(n^{1+\gamma^{(m)}\wedge\gamma^{(\ell)}}(\log p)^{1+\frac{1}{2r}}),
\end{aligned}$$

when the second inequality applies the fact that

$$1-\rho_j^*\rho_\ell^* = \frac{|c_j^*|}{n^{\gamma^{(k)}}} + \frac{|c_\ell^*|}{n^{\gamma^{(m)}}} - \frac{c_j^*c_\ell^*}{n^{\gamma^{(k)}+\gamma^{(m)}}} = O(n^{-(\gamma^{(m)}\wedge\gamma^{(\ell)})})$$

when $j \in \mathcal{G}^{(k)}, \ell \in \mathcal{G}^{(m)}$. This completes the proof of Lemma B.4. $\square$

Lemma B.5. *Suppose that $(1 + C_m(L))s = o(n \wedge p)$ as $n \to \infty$, and $k = 1$. Define $\ddot{\Sigma}^{(1)} = n^{-1}\sum_{t=1}^{n}(x_{t-1} - \bar{x})(x_{t-1} - \bar{x})^\top$. Then under Under Assumptions 1-5,*

$$\frac{\kappa_I(\ddot{\Sigma}^{(1)}, L, s)}{n} \geq \frac{c_\kappa}{L^2 s \log p} \tag{B.72}$$

holds w.p.a.1. for any $L \geq 1$.[B.2]

Proof of Lemma B.5. In this proof only, we simplify the notations $x_t^{(1)}$ and $e_t^{(1)}$ as x_t and e_t, given there is no ambiguity that we focus on the LUR case. Also, we simplify $\ddot{\Sigma}^{(1)}$ as $\ddot{\Sigma}$.

(a) We first impose the normality assumption $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, I_p)$. It implies $e_t \sim i.i.d.\ \mathcal{N}(0, \Omega_e)$ with $\Omega_e = \Phi_e \Phi_e^\top$. Note that for the LUR cases,

$$x_t - x_{t-1} = \frac{\mathbf{C}}{n} x_{t-1} + e_t$$

for any $t \geq 1$, where $\mathbf{C} = \mathbf{C}^{(k)}$. Define

$$e_t^\Delta = \begin{cases} \dfrac{\mathbf{C}}{n} x_{t-1} + e_t, & t \geq 1, \\ 0, & t = 0, \end{cases}$$

and note that $x_t = \sum_{s=1}^{t} e_s^\Delta$. Let R be an $n \times n$ lower triangular matrix of ones on and below the diagonal. Define $X = (x_0, x_1, \ldots, x_{n-1})^\top$, $e = (e_0, e_1, \ldots, e_{n-1})^\top$ and $e^\Delta = (e_0^\Delta, e_1^\Delta, \ldots, e_{n-1}^\Delta)^\top$. Note that $\underset{(n\times p)}{X} = \underset{(n\times n)}{R}\underset{(n\times p)}{e^\Delta}$. We decompose we write

$$\ddot{\Sigma} = n^{-1} X^\top X = n^{-1} e^{\Delta\top} R^\top R e^\Delta.$$

Define $\mathbf{J}_n = n^{-1} 1_n 1_n^\top$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_n \geq 0$ be the eigenvalues of $R^\top(\mathbf{I}_n - \mathbf{J}_n)R$ and $R^\top R$, respectively, ordered from large to small.

Let μ_ℓ be the ℓth largest singular value of the idempotent matrix $\mathbf{I}_n - \mathbf{J}_n$. Recall $\mathbf{1}(\cdot)$ is the indicator function, and obviously $\mu_\ell = \mathbf{1}(1 \leq \ell \leq n-1)$ for $\ell \in [n]$. Denote the ℓth eigenvalue values of $R^\top(\mathbf{I}_n - \mathbf{J}_n)R$ and $R^\top R$ be λ_ℓ and $\widetilde{\lambda}_\ell$, respectively. When $\ell \in [n-1]$, the first inequality of Eq.(15) in Merikoski and Kumar (2004, Theorem 9) gives $\lambda_\ell \geq \widetilde{\lambda}_{\ell+1}\mu_{n-1} = \widetilde{\lambda}_{\ell+1}$.

Following the technique used to prove Remark 3.5 in Zhang et al. (2019), which is also used for Theorem B.2 in Smeekes and Wijler (2021), we diagonalize $R(\mathbf{I}_n - \mathbf{J}_n)R^\top = V\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)V^\top$, where V is an orthonormal matrix. For any $\delta \in \mathbb{R}^p$, $\delta \neq 0$, the

[B.2] Here we use a generic $L \geq 1$ is useful for deducing the lower bound of $\widehat{\kappa}_D$ using $\widehat{\kappa}_I$.

quadratic form

$$\begin{aligned}\delta^\top \ddot{\Sigma}\delta &= \frac{1}{n}e^{\Delta\top}R^\top R e^\Delta = \frac{1}{n}\delta^\top e^{\Delta\top}V\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)V^\top e^\Delta\delta \\ &\geq \frac{1}{n}\delta^\top e^{\Delta\top}V_{\cdot[\ell]}\mathrm{diag}(\lambda_1,\cdots,\lambda_\ell)V_{\cdot[\ell]}^\top e^{(n)}\delta \geq \frac{\lambda_\ell}{n}\delta^\top e^{\Delta\top}V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e^\Delta\delta \\ &\geq \frac{\ell\widetilde{\lambda}_{\ell+1}}{n}\cdot\delta^\top\Gamma_\ell^\Delta\delta \end{aligned} \tag{B.73}$$

for any $\ell \in [n-1]$, where $V_{\cdot[\ell]}$ is the submatrix composed of the first ℓ columns of V and $\Gamma_\ell = \ell^{-1}e^{\Delta\top}V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e^\Delta$.

We first work with the first factor $\ell\lambda_\ell/n$ in (B.73). Smeekes and Wijler (2021) provide the exact formula of λ_ℓ:

$$\widetilde{\lambda}_{\ell+1} = \left[2\left(1-\cos\left(\frac{(2\ell+1)\pi}{2n+1}\right)\right)\right]^{-1} \text{ for all } \ell \in [n]. \tag{B.74}$$

A Taylor expansion of $\cos(x\pi)$ around $x=0$ yields

$$\widetilde{\lambda}_{\ell+1} = \left(\frac{(2\ell+1)\pi}{2n+1}\right)^2\left(1+o\left(\frac{\ell+1}{n}\right)\right) = \left(\frac{\ell\pi}{n}\right)^2\left(1+o\left(\frac{\ell}{n}\right)\right)$$

whenever $\ell = o(n)$. This implies

$$\frac{\ell\widetilde{\lambda}_{\ell+1}}{n} = \frac{n}{\pi^2\ell\left(1+o\left(\ell/n\right)\right)} \geq \frac{n}{2\pi^2\ell} \tag{B.75}$$

for $\ell = o(n)$ when n is sufficiently large.

Next, we focus on the second factor $\delta^\top\Gamma_\ell^\Delta\delta$ in (B.73). Define $X_{\mathbb{L}} := (0_p, X_0, X_1, \cdots, X_{t-2})^\top$. By definition, we have

$$e^\Delta = X_{\mathbb{L}}\frac{\mathbf{C}}{n} + e.$$

We deduce that

$$\begin{aligned}\delta^\top\Gamma_\ell^\Delta\delta &= \frac{\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} + \frac{\delta^\top\mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top X_{\mathbb{L}}\mathbf{C}^\top\delta}{n^2\ell} + \frac{2\delta^\top\mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{n\ell} \\ &\geq \frac{\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} + \frac{2\delta^\top\mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{n\ell}.\end{aligned}$$

Recall the generic inequality $2x^\top y \leq x^\top x + y^\top y$ for any vectors x and y of the same dimension.

Let $x = \frac{V_{\cdot[\ell]}^\top e\delta}{\sqrt{2}}$ and $y = \sqrt{2}n^{-1}\mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta$, we have

$$\frac{2\delta^\top \mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{n} \leq 0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta + \frac{2\mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top X_{\mathbb{L}}\mathbf{C}^\top}{n^2}.$$

It implies

$$\delta^\top \Gamma_\ell^\Delta \delta \geq \frac{0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} - \frac{2\delta^\top \mathbf{C}X_{\mathbb{L}}^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top X_{\mathbb{L}}\mathbf{C}\delta}{n^2\ell}.$$

In addition, $\lambda_{\max}(V_{\cdot[\ell]}V_{\cdot[\ell]}^\top) \leq \|V_{\cdot[\ell]}\|_2^2 \leq 1$, where the second inequality applied the fact that V is a unitary matrix. Therefore,

$$\begin{aligned}
\delta^\top \Gamma_\ell^\Delta \delta &\geq \frac{0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} - \frac{2\delta^\top \mathbf{C}X_{\mathbb{L}}^\top X_{\mathbb{L}}\mathbf{C}\delta}{n^2\ell} \\
&= \frac{0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} - \frac{2\delta^\top \mathbf{C}\sum_{t=1}^{n-1} X_{t-1}X_{t-1}^\top \mathbf{C}\delta}{n^2\ell} \\
&\geq \frac{0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} - \frac{2\delta^\top \mathbf{C}\sum_{t=1}^{n} X_{t-1}X_{t-1}^\top \mathbf{C}\delta}{n^2\ell} \\
&\geq \frac{0.5\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell} - \frac{2\delta^\top \mathbf{C}\widehat{\Sigma}\mathbf{C}\delta}{n\ell}.
\end{aligned} \tag{B.76}$$

Define $\Gamma_\ell = \dfrac{\delta^\top e^\top V_{\cdot[\ell]}V_{\cdot[\ell]}^\top e\delta}{\ell}$. By (B.73) and (B.76),

$$\delta^\top \ddot{\Sigma}\delta \geq \frac{n}{2\pi^2\ell}\left(0.5\delta^\top \Gamma_\ell \delta - 2n^{-1}\ell^{-1}\delta^\top \mathbf{C}\widehat{\Sigma}\mathbf{C}\delta\right). \tag{B.77}$$

We first lower bound the first term. Let $\ell = (16 + C_\ell)\cdot(s+m)\log p$ for some $C_\ell > 0$ to be determined later. Following the proof of (B.43) in MS24 utilizing the non-asymptotic bounds for Whishart matrices, we have

$$\delta^\top \Gamma_\ell \delta \geq C_\kappa \|\delta\|_2^2, \tag{B.78}$$

w.p.a.1, where the absolute constant C_κ not dependent on L or C_ℓ. We then bound the second term. Note that for any $\delta \in \mathcal{R}(L, s)$ such that for any $|\mathcal{M}| \leq s$ we have $\|\delta_{\mathcal{M}^c}\|_1 \leq L\|\delta_{\mathcal{M}}\|_1$,

$$\|\delta\|_1 \leq \|\delta_{\mathcal{M}}\|_1 + \|\delta_{\mathcal{M}^c}\|_1 \leq (1+L)\|\delta_{\mathcal{M}}\|_1 \leq (1+L)\sqrt{s}\|\delta\|_2. \tag{B.79}$$

Therefore,

$$\delta^\top \mathbf{C}\widehat{\Sigma}\mathbf{C}^\top \delta \leq (\|\delta\|_1)^2 \cdot \|\mathbf{C}\|_1^2 \cdot \|\widehat{\Sigma}\|_\infty \leq (1+L)^2 s\|\delta\|_2^2 \cdot \|\mathbf{C}\|_1^2 \cdot \|\widehat{\Sigma}\|_\infty.$$

By (B.4) for LURs,

$$\|\widehat{\Sigma}\|_\infty^2 \le \max_{j,t} |x_{j,t-1}|^2 \le C_{\sup} n \log p$$

w.p.a.1 for some absolute constant $C_{\sup}$. Therefore,

$$\begin{aligned}
\delta^\top \mathbf{C}\ddot{\Sigma}\mathbf{C}^\top \delta &\le (1+L)^2 s\|\delta\|_2^2 \cdot \|\mathbf{C}\|_1^2 \cdot \|\widehat{\Sigma}\|_\infty \\
&\le (1+L)^2 s\|\delta\|_2^2 \cdot 2^2 \cdot C_{\sup} n \log p \\
&\le 16L^2 \cdot s\|\delta\|_2^2 \cdot C_{\sup} n \log p,
\end{aligned}$$

where the second inequality applies $\|\mathbf{C}\|_1 \le \sup_{j\in[p]} |\rho_j^*| \le 2$, and the third inequality applies $L \ge 1$ and $\|\delta_{\mathcal{S}}\|_1 \le \sqrt{s}\|\delta\|_2$. Recall that $\ell = (16 + C_\ell) \cdot (s + m) \log p$. Let $C_\ell = 64 \cdot (1 \vee (L \cdot C_{\sup})) - 16$. Then

$$\delta^\top \mathbf{C}\ddot{\Sigma}\mathbf{C}^\top \delta \le \|\delta\|_2^2 \cdot \frac{16L \cdot C_{\sup}}{C_\kappa} s \cdot n \log p \le \|\delta\|_2^2 \cdot 0.25 C_\kappa n\ell. \tag{B.80}$$

Insert (B.78) and (B.80) into (B.77), we have

$$\delta^\top \ddot{\Sigma}\delta \ge \frac{n}{2\pi^2\ell}\left(0.5C_\kappa - 0.25C_\kappa\right)\|\delta\|_2^2 = \frac{nC_\kappa}{8\pi^2\ell}\|\delta\|_2^2.$$

By $\ell = (16 + C_\ell) \cdot (s + m) \log p$, $m = \lceil 4L\widetilde{C}/\widetilde{c}\rceil s$, and $L \ge 1$

$$\begin{aligned}
\frac{\delta^\top \widehat{\Sigma}\delta}{n\|\delta\|_2^2} &\ge \frac{C_\kappa}{8\pi^2(16 + C_\ell)\cdot(s+m)\log p} \\
&\ge \frac{C_\kappa}{8\pi^2(16 + C_\ell)\cdot(1 + \lceil 4L\widetilde{C}/\widetilde{c}\rceil)s\log p} \\
&\ge \frac{C_\kappa}{8\pi^2 \cdot 64(1 \vee (L\cdot C_{\sup}))\cdot(2\widetilde{C}/\widetilde{c})\cdot 8L \cdot s\log p} \\
&\ge \frac{C_\kappa}{8\pi^2 \cdot 64C_{\sup}\cdot(2\widetilde{C}/\widetilde{c})\cdot 8L^2 \cdot s\log p}
\end{aligned} \tag{B.81}$$

w.p.a.1. Then (B.72) holds with $\widetilde{c}_\kappa = \widetilde{c} \cdot C_\kappa / [8\pi^2 \cdot 64C_{\sup} \cdot 2\widetilde{C} \cdot 8]$.

(b) We then extend the result to non-normal errors. Recall that

$$x_{j,t} = \sum_{s=1}^{t} \rho_j^{t-s} e_{j,s} = \sum_{k=1}^{p} \Phi_{j,k} \sum_{s=1}^{t} \rho_j^{t-s}\varepsilon_{k,s}.$$

The random component $\sum_{s=1}^{t} \rho_j^{t-s}\varepsilon_{k,s}$ is a partial sum of α-mixing variables, and $\varepsilon_{k,s}$ is independent of $\varepsilon_{j,t}$ whenever $k \ne j$. Also note that ρ_j^{t-s}. Following the technique of By the

Gaussian approximation, there exists a sequence of Brownian motion such that

$$\sup_{j\in[p],t\in[n]} \frac{1}{\sqrt{n}}\left|\sum_{s=0}^{t-1}\rho_j^{t-s}\varepsilon_{j,s}-\sum_{s=0}^{t-1}\rho_j^{t-s}\eta_{j,s}\right| \overset{\mathrm{p}}{\preccurlyeq} \frac{\log p}{n^{1/4}}.$$

Define $\xi_{j,t}=\sum_{s=1}^{t}\rho_j^{t-s}\eta_{j,s}$ satisfying $\xi_{j,t}=\rho_j\xi_{j,t-1}+\eta_{j,s}$ Therefore,

$$\sup_{j\in[p],t\in[n]}|x_{j,t}-\xi_{j,t}|\le \sup_{j\in[p]}\left|\sum_{k=1}^{p}\Phi_{j,k}\right|\cdot\left|\sum_{s=0}^{t-1}\rho_j^{t-s}\varepsilon_{j,s}-\sum_{s=0}^{t-1}\rho_j^{t-s}\eta_{j,s}\right| = O_p\left(n^{1/4}\log p\right). \quad \text{(B.82)}$$

Let $\ddot{\Upsilon}=n^{-1}\sum_{t=1}^{n}(\xi_{t-1}-\bar{\xi})(\xi_{t-1}-\bar{\xi})^\top$.

$$\delta^\top\ddot{\Sigma}\delta\ge\delta^\top\ddot{\Upsilon}\delta-\left|\delta^\top(\ddot{\Sigma}-\ddot{\Upsilon})\delta\right| \quad \text{(B.83)}$$

Notice that $\widehat{\Upsilon}$ is the Gram matrix of the LUR processes ζ_t with normally distributed errors. The procedures as in Part (a) bounds the first term on the right-hand side of the above expression

$$\delta^\top\ddot{\Upsilon}\delta\ge\frac{c'_\kappa}{L^2 s\log p}n\|\delta\|_2^2 \quad \text{(B.84)}$$

w.p.a.1 for some absolute constant c'_κ. We move on to the second term

$$\begin{aligned}\left|\delta^\top(\widehat{\Sigma}-\widehat{\Upsilon})\delta\right| &\le \|\delta\|_1^2\|\ddot{\Sigma}-\ddot{\Upsilon}\|_\infty \le (1+L)^2 s\|\delta\|_2^2\|\ddot{\Sigma}-\ddot{\Upsilon}\|_\infty \\ &\le 4L^2 s\|\delta\|_2^2\|\ddot{\Sigma}-\ddot{\Upsilon}\|_\infty \quad \text{(B.85)}\end{aligned}$$

whenever $L\ge 1$. Since $X_t=\sum_{s=0}^{t}e_s=\Phi_e\sum_{s=0}^{t}\varepsilon_s=\Phi_e\xi_{t-1}$, it follows that

$$\|\ddot{\Sigma}-\ddot{\Upsilon}\|_\infty \;\le\; C_L^2\left(\|n^{-1}\sum_{t=1}^{n}(x_{t-1}x_{t-1}^\top-\zeta_{t-1}\zeta_{t-1}^\top)\|_\infty+\|\bar{x}\bar{x}^\top-\bar{\xi}\bar{\xi}^\top\|_\infty\right).$$

Following the proof of Part (b) in Proposition MS24, we can show under (B.82) that

$$\|\ddot{\Sigma}-\ddot{\Upsilon}\|_\infty=O_p\left(n^{3/4+\nu'}\sqrt{\log p}\right)$$

for any arbitrary small absolute value ν'. Inserting the above expression into (B.85), we have

$$\frac{|\delta^\top(\ddot{\Sigma}-\ddot{\Upsilon})\delta|}{n\|\delta\|_2^2}\le 4L^2 s\cdot O_p\left(n^{-1/4+\nu'}\sqrt{\log p}\right)=o_p\left(\frac{L^{-2}}{s\log p}\right) \quad \text{(B.86)}$$

given the condition $s^2L^4(\log p)^{3/2}=o(n^{1/4-\nu'})$ in the Proposition. (B.84) and (B.86) then provide

$$\frac{\delta^\top \ddot{\Sigma}\delta}{n\|\delta\|_2^2} \geq \frac{c'_\kappa}{L^2 s\log p} - o_p\left(\frac{L^{-2}}{s\log p}\right) \geq \frac{c_\kappa}{L^2 s\log p}$$

w.p.a.1 when n is large enough, where $c_\kappa = 0.5c'_\kappa$. □

Define the long-run covariance $\Omega^{(1)} = \sum_{d=-\infty}^{\infty} \mathbb{E}(e_{t-d}^{(1)} e_t^{(1)\top})$, and

$$\Sigma^{(1)} = \int_{-\infty}^{\infty} \mathrm{e}^{\tau \mathbf{C}^{(1)}} \Omega^{(1)} \mathrm{e}^{\tau \mathbf{C}^{(1)}} \mathrm{d}\tau.$$

Also, recall that $\Sigma^{(0)} = \mathbb{E}(x_t^{(0)} x_t^{(0)\top})$.

Proposition B.8. *Under Assumptions 1-5, for $k = 0$*

$$\frac{1}{n^{\theta\wedge k}}\|\widehat{\Sigma}^{(k)} - \Sigma^{(k)}\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{n^\theta}\mathbf{1}\{k=1\} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\theta\wedge k}}}. \tag{B.87}$$

Therefore, there exists a absolute constant $c_\kappa^{(0)}$ such that

$$\kappa_{\mathbf{I}}(\widehat{\Sigma}^{(0)}, 3, s) \geq c_\kappa^{(0)} \tag{B.88}$$

w.p.a.1.

Proof of Proposition B.8. When $k = 0$, i.e. $x_t^{(k)}$ is stationary, (B.87) follows standard concentration inequalities for weakly dependent time series, like (B.30) in MS.

When $k = 1$, define $\mathbf{R}^{(1)} = \mathrm{diag}(\{\rho_j^*\}_{j\in\mathcal{G}^{(1)}})$, and $\mathbf{K}^{(1)}$ as the $p_1^2 \times p_1^2$ commutation matrix. We follow Magdalinos and Phillips (2009) to decompose $\mathrm{vec}(\widehat{\Sigma}^{(1)})$ as

$$\begin{aligned}
&(\mathbf{I}_{p_k^2} - \mathbf{R}^{(1)} \otimes \mathbf{R}^{(1)})\mathrm{vec}(\widehat{\Sigma}^{(1)}) \\
=&\frac{1}{n}\left(\mathrm{vec}(x_n^{(1)} x_n^{(1)\top}) - \mathbf{R}^{(1)} \otimes \mathbf{R}^{(1)}\mathrm{vec}(x_0^{(1)} x_0^{(1)\top})\right) \\
&+ (\mathbf{I}_{p_k^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_k} \otimes \mathbf{R}^{(1)})\frac{1}{n}\sum_{t=1}^n \mathrm{vec}(x_{t-1}^{(1)} e_t^{(1)\top}) + \frac{1}{n}\sum_{t=1}^n \mathrm{vec}(e_t^{(1)} e_t^{(1)\top}).
\end{aligned}$$

By $\mathbf{I}_{p_k^2} - \mathbf{R}^{(1)} \otimes \mathbf{R}^{(1)} = -n^{-\theta}(\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k}) - n^{-2\theta}\mathbf{C}^{(1)} \otimes \mathbf{C}^{(1)}$, we further deduce

$$
\begin{aligned}
& - (\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k}) \frac{\mathrm{vec}(\widehat{\Sigma}^{(1)})}{n^\theta} \\
=& \frac{1}{n}\left(\mathrm{vec}(x_n^{(1)} x_n^{(1)\top}) - \mathbf{R}^{(1)} \otimes \mathbf{R}^{(1)} \mathrm{vec}(x_0^{(1)} x_0^{(1)\top})\right) + \frac{\mathbf{C}^{(1)} \otimes \mathbf{C}^{(1)}}{n^{2\theta}} \mathrm{vec}(\widehat{\Sigma}^{(1)}) \\
& + (\mathbf{I}_{p_k^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_k} \otimes \mathbf{R}^{(1)}) \frac{1}{n} \sum_{t=1}^n \mathrm{vec}(x_{t-1}^{(1)} e_t^{(1)\top}) + \frac{1}{n} \sum_{t=1}^n \mathrm{vec}(e_t^{(1)} e_t^{(1)\top}).
\end{aligned} \tag{B.89}
$$

By Lemma B.1,

$$
\|\frac{1}{n}\left(\mathrm{vec}(x_n^{(1)} x_n^{(1)\top}) - \mathbf{R}^{(1)} \otimes \mathbf{R}^{(1)} \mathrm{vec}(x_0^{(1)} x_0^{(1)\top})\right)\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^3}{n^{1-\theta}}, \tag{B.90}
$$

and

$$
\|\frac{\mathbf{C}^{(1)} \otimes \mathbf{C}^{(1)}}{n^{2\theta}} \mathrm{vec}(\widehat{\Sigma}^{(1)})\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{n^\theta (\log p)^3}{n^{2\theta}} = \frac{(\log p)^3}{n^\theta}. \tag{B.91}
$$

Define $\Lambda^{(1)} = \sum_{d=0}^\infty e_t^{(1)} e_{t-d}^{(1)\top}$. By (B.46) in Lemma B.3,

$$
\|(\mathbf{I}_{p_k^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_k} \otimes \mathbf{R}^{(1)}) \frac{1}{n} \sum_{t=1}^n \mathrm{vec}(x_{t-1}^{(1)} e_t^{(1)\top} - \Lambda^{(1)})\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}}. \tag{B.92}
$$

By standard concentration inequalities for stationary time series like (B.30) in MS,

$$
\|\frac{1}{n} \sum_{t=1}^n \mathrm{vec}(e_t^{(1)} e_t^{(1)\top} - \Omega^{(1)})\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \sqrt{\frac{\log p}{n}}, \tag{B.93}
$$

where $\Omega^{(1)} = \mathbb{E}(e_t^{(1)} e_t^{(1)\top})$. Therefore, by (B.89), (B.90), (B.91), (B.92), and (B.93) we have

$$
-(\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_1}) \frac{\mathrm{vec}(\widehat{\Sigma}^{(1)})}{n^{\gamma^{(1)}}} = (\mathbf{I}_{p_1^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_1} \otimes \mathbf{R}^{(1)}) \mathrm{vec}(\Lambda^{(1)}) + \mathrm{vec}(\Omega^{(1)}) + g_{1,n}, \tag{B.94}
$$

where

$$
\begin{aligned}
\|g_{1,n}\|_\infty &\overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^3}{n^{1-\theta}} + \frac{(\log p)^3}{n^\theta} + \frac{(\log p)^{\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{2+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}} + \sqrt{\frac{\log p}{n}} \\
&= O\left(\frac{(\log p)^{3+\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}}\right).
\end{aligned} \tag{B.95}
$$

In addition,

$$
(\mathbf{I}_{p_k^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_k} \otimes \mathbf{R}^{(1)}) \mathrm{vec}(\Lambda^{(1)}) = \mathrm{vec}(\Lambda^{(1)} + \Lambda^{(1)\top}) + g_{2,n} \tag{B.96}
$$

where

$$\|g_{2,n}\|_\infty = \|(\mathbf{I}_{p_k^2} + \mathbf{K}^{(1)})(\mathbf{I}_{p_k} \otimes (\mathbf{R}^{(1)} - \mathbf{I}_{p_k}))\mathrm{vec}(\Lambda^{(1)})\|_\infty = O(n^{-\theta}). \tag{B.97}$$

Not that the long-run covariance $\Theta^{(1)} = \Lambda^{(1)} + \Lambda^{(1)\top} + \Omega^{(1)}$. Then by (B.95) and (B.97),

$$\| - (\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k})\frac{\mathrm{vec}(\widehat{\Sigma}^{(1)})}{n^\theta} - \mathrm{vec}(\Theta^{(1)})\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}}.$$

Notice that $\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k}$ is diagonal, with the diagonal entries uniformly bounded from above and below by some absolute constants. Therefore,

$$\left\|\frac{\mathrm{vec}(\widehat{\Sigma}^{(1)})}{n^\theta} + (\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k})^{-1}\mathrm{vec}(\Theta^{(1)})\right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}}$$

Define $c_j^{(1)}$ as the j-th diagonal entry of $\mathbf{C}^{(1)}$. Observe that

$$\begin{aligned}(\mathbf{I}_{p_k} \otimes \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \otimes \mathbf{I}_{p_k})^{-1}\mathrm{vec}(\Theta^{(1)}) &= \int_0^\infty \mathrm{e}^{\tau\mathbf{C}^{(1)}} \otimes \mathrm{e}^{\tau\mathbf{C}^{(1)}}\mathrm{d}\tau \cdot \mathrm{vec}(\Theta^{(1)}) \\ &= \mathrm{vec}(\int_0^\infty \mathrm{e}^{\tau\mathbf{C}^{(1)}} \cdot \Theta^{(1)} \cdot \mathrm{e}^{\tau\mathbf{C}^{(1)}}\mathrm{d}\tau).\end{aligned}$$

Therefore,

$$\left\|n^{-\theta} \cdot \widehat{\Sigma}^{(1)} - \int_0^\infty \mathrm{e}^{\tau\mathbf{C}^{(1)}} \cdot \Theta^{(1)} \cdot \mathrm{e}^{\tau\mathbf{C}^{(1)}}\mathrm{d}\tau\right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{n^\theta} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\theta}}},$$

which completes the proof. □

B.3.2 Technical Lemmas for Inference

Let $\zeta_{j,t}$ be an AR(1) process satisfying

$$\zeta_{j,t} = \rho_z\zeta_{j,t-1} + e_{j,t} \tag{B.98}$$

with a zero initial value. Therefore, $\zeta_{j,t}$ is an MI process with degree of persistence θ, thereby satisfying

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]} |\zeta_{j,t}| \overset{\mathrm{p}}{\preccurlyeq} n^{\frac{\theta}{2}}(\log p)^{3/2}. \tag{B.99}$$

Further define

$$\psi_{j,t} = \sum_{s=1}^{t} \rho_z^{t-s}x_{j,s-1}. \tag{B.100}$$

The instrument $z_{j,t}$ has the following decompositions

$$z_{j,t} = \zeta_{j,t} - (1-\rho_j^*)\psi_{j,t} \tag{B.101}$$

and

$$z_{j,t} = x_{j,t} - (1-\rho_z)\psi_{j,t}. \tag{B.102}$$

Lemma B.6. *Under the conditions in Theorem 2, for $k, m \in \{0,1\}$ we have*

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]} |\psi_{j,t}| \overset{\mathrm{p}}{\preccurlyeq} n^{\frac{k\vee\theta}{2}+(\theta\wedge k)}(\log p)^{3/2}, \tag{B.103}$$

$$\max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} \left|\sum_{t=1}^{n}\psi_{j,t-1}x_{\ell,t-1}\right| \overset{\mathrm{p}}{\preccurlyeq} n^{1+\frac{k\vee\theta}{2}+(\theta\wedge k)+\frac{m}{2}}(\log p)^{3}, \tag{B.104}$$

and

$$\sup_{j\in\mathcal{G}^{(k)}} \left|\sum_{t=1}^{n}\psi_{j,t-1}u_t\right| \overset{\mathrm{p}}{\preccurlyeq} n^{\frac{1}{2}+\frac{k\vee\theta}{2}+(\theta\wedge k)}(\log p)^{2+\frac{1}{2r}}. \tag{B.105}$$

Proof of Lemma B.6. For (B.103), by definition in (B.100),

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},t\in[n]} |\psi_{j,t}| &\leq \max_{t\in[n]}\left|\sum_{s=1}^{t}\rho_z^{t-s}\right| \max_{j\in\mathcal{G}^{(k)},t\in[n]}|x_{j,t-1}| \\ &\leq \frac{n^\theta}{|c_z|}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|x_{j,t-1}| \overset{\mathrm{p}}{\preccurlyeq} n^{\theta+\frac{k}{2}}(\log p)^{3/2}. \end{aligned}\tag{B.106}$$

where the last inequality applies Lemma B.1. In addition, we can reorganize the expression of $\psi_{j,t}$ as

$$\begin{aligned}\psi_{j,t} &= \sum_{s=1}^{t}\sum_{r=0}^{s-1}\rho_z^{t-s}\rho_j^{*r}e_{j,s-1-r} = \sum_{r=0}^{t-1}\rho_j^{*r}\sum_{s=r+1}^{t}\rho_z^{t-s}e_{j,s-1-r} = \sum_{r=0}^{t-1}\rho_j^{*r}\left(\sum_{d=1}^{t-r}\rho_z^{t-r-d}e_{j,d-1}\right) \\ &= O(n^{\gamma^{(k)}})\max_{j\in\mathcal{G}^{(k)},t\in[n]}\left|\sum_{d=1}^{t-r}\rho_z^{t-r-d}e_{j,d-1}\right| \end{aligned}\tag{B.107}$$

and thus

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]} |\psi_{j,t}| \leq O(n^k)\max_{j\in\mathcal{G}^{(k)},t\in[n]}\left|\sum_{d=1}^{t-r}\rho_z^{t-r-d}e_{j,d-1}\right|.$$

Note that $\{\sum_{d=1}^{t}\rho_z^{t-d}e_{j,d-1}\}_{t\in[n]}$ is an MI process with degree of persistence equaling θ. By the arguments the prove Lemma B.1 we have

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]}\max_{j\in\mathcal{G}^{(k)},t\in[n]}\left|\sum_{d=1}^{t}\rho_z^{t-d}e_{j,d-1}\right| \overset{\mathrm{p}}{\precsim} Cn^{\frac{\theta}{2}}(\log p)^{3/2}.$$

Therefore, by (B.107)

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t}| \overset{\mathrm{p}}{\precsim} Cn^{k+\frac{\theta}{2}}(\log p)^{3/2}(\log p)^{3/2}. \tag{B.108}$$

By (B.106) and (B.108),

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t}| &\overset{\mathrm{p}}{\precsim} \left(n^{\theta+\frac{k}{2}}(\log p)^{3/2}\right)\wedge\left(n^{k+\frac{\theta}{2}}(\log p)^{3/2}\right)\\ &=n^{\frac{k\vee\theta}{2}+(\theta\wedge k)}(\log p)^{3/2},\end{aligned}$$

which verifies (B.103). (B.104) follows by

$$\max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}}\left|\sum_{t=1}^{n}\psi_{j,t-1}x_{\ell,t-1}\right| \overset{\mathrm{p}}{\precsim} n\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t}|\cdot\max_{j\in\mathcal{G}^{(m)},t\in[n]}|x_{j,t}|$$

and the inequalities (B.103) and (B.1).

For (B.105), by definition of $\psi_{j,t}$ in (B.100) we have the recursive formula $\psi_{j,t}=\rho_z\psi_{j,t-1}+x_{j,t-1}$. Take $G=\lfloor(2c_\alpha^{-1}\log(np))^{1/r}\rfloor$ as in the proof of (B.45) in Lemma B.2. We thus have the following decomposition.

$$\sum_{t=1}^{n}\psi_{j,t-1}u_t=\sum_{t=1}^{G}\psi_{j,t-1}u_t+\sum_{d=0}^{G-2}\rho_z^d\sum_{t=G+1}^{n}u_tx_{j,t-d-1}+\rho_z^{G-1}\sum_{t=G+1}^{n}\psi_{j,t-G}u_t. \tag{B.109}$$

For the first term of the right-hand side, we have

$$\max_{j\in\mathcal{G}^{(k)}}|\sum_{t=1}^{G}\psi_{j,t-1}u_t| \overset{\mathrm{p}}{\precsim} G\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t-1}|\max_{t\in[n]}|u_t| \overset{\mathrm{p}}{\precsim} n^{\frac{k\vee\theta}{2}+(\theta\wedge k)}(\log p)^{5/2}, \tag{B.110}$$

where the last inequality applies (B.103) and the sub-exponential distribution of u_t. For the second term, by Proposition B.3

$$\max_{j\in\mathcal{G}^{(k)}}|\sum_{d=0}^{G-2}\rho_z^d\sum_{t=G+1}^{n}u_tx_{j,t-d-1}| \overset{\mathrm{p}}{\precsim} G\sqrt{n^{1+k}}(\log p)^{2+\frac{1}{2r}}. \tag{B.111}$$

We then bound the third term. Following the arguments to bound "T_3" in the proof of (B.45) in Lemma B.2 (See also the proof of Proposition B.2 in MS), by the upper bound of $\psi_{j,t}$ given as (B.103), we have

$$\left|\rho_z^{G-1}\sum_{t=G+1}^{n}\psi_{j,t-G}u_t\right| \overset{\mathrm{p}}{\precsim} n^{\frac{1}{2}+\frac{k\vee\theta}{2}+(\theta\wedge k)}(\log p)^{2+\frac{1}{2r}}, \tag{B.112}$$

Then (B.105) follows by (B.109), (B.110), (B.111), and (B.112). We complete the proof of Lemma B.6. $\square$

Lemma B.7. *Under the conditions in Theorem 2, we have*

$$\max_{j\in\mathcal{G}^{(k)},t\in[n]}|z_{j,t}| \overset{\mathrm{p}}{\precsim} n^{\frac{k\wedge\theta}{2}}(\log p)^{3/2}, \tag{B.113}$$

$$\max_{j\in\mathcal{G}^{(k)},\ell\in[p]}\left|\sum_{t=1}^{n}z_{j,t-1}e_{\ell,t}\right| \overset{\mathrm{p}}{\precsim} n+n^{1-\frac{k\vee\theta}{2}+k\wedge\theta}(\log p)^{3/2}, \tag{B.114}$$

and

$$\max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}}\left|\sum_{t=1}^{n}z_{j,t-1}x_{\ell,t-1}\right| \overset{\mathrm{p}}{\precsim} (n^{\theta\wedge k\wedge m}+n^{-\frac{k\vee\theta}{2}+k\wedge\theta+\frac{m}{2}})\cdot n(\log p)^{3+\frac{1}{2r}}. \tag{B.115}$$

Proof of Lemma B.7. For (B.7), when $\theta\le\gamma_j$, by (B.101) we have

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|z_{j,t}| &\le \max_{j\in\mathcal{G}^{(k)},t\in[n]}|\zeta_{j,t}|+\frac{\bar{c}}{n^{\gamma^{(k)}}}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t}|\\ &\overset{\mathrm{p}}{\precsim} n^{\frac{\theta}{2}}(\log p)^{3/2}+n^{\frac{k}{2}+\theta-k}(\log p)^{3/2}\\ &= n^{\frac{\theta}{2}}(\log p)^{3/2}+n^{\frac{\theta}{2}-\frac{k-\theta}{2}}(\log p)^{3/2}\le 2n^{\frac{\theta}{2}}(\log p)^{3/2}.\end{aligned} \tag{B.116}$$

When $\gamma_j<\theta$, by (B.102)

$$\begin{aligned}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|z_{j,t}| &\le \max_{j\in\mathcal{G}^{(k)},t\in[n]}|x_{j,t}|+\frac{|c_z|}{n^{\theta}}\max_{j\in\mathcal{G}^{(k)},t\in[n]}|\psi_{j,t}|\\ &\overset{\mathrm{p}}{\precsim} n^{\frac{k}{2}}(\log p)^{3/2}+n^{\frac{\theta}{2}+k-\theta}(\log p)^{3/2}\\ &= n^{\frac{k}{2}}(\log p)^{3/2}+n^{\frac{k}{2}-\frac{\theta-k}{2}}(\log p)^{3/2}\le 2n^{\frac{\theta}{2}}(\log p)^{3/2}.\end{aligned} \tag{B.117}$$

We then have (B.7) by combining (B.116) and (B.117).

For (B.114), by (B.101), when $k = 1$

$$\max_{j\in\mathcal{G}^{(1)},\ell\in[p]} |\sum_{t=1}^{n} z_{j,t-1}e_{\ell,t}| \le \max_{j\in\mathcal{G}^{(1)},\ell\in[p]} |\sum_{t=1}^{n} \zeta_{j,t-1}e_{\ell,t}| + \frac{\bar{c}}{n} \max_{j\in\mathcal{G}^{(1)},\ell\in[p]} |\sum_{t=1}^{n} \psi_{j,t-1}e_{\ell,t}|$$
$$\overset{\mathrm{p}}{\preccurlyeq} n + n^{\frac{1}{2}+\theta}(\log p)^3, \tag{B.118}$$

where in the last step the upper bound of the first term applies (B.46) bounded the cross-product between and MI process and a stationary component, and the bound of the second term applies (B.104) with $m = 0$ as $e_{\ell,t}$ is stationary. Similarly, we can deduce by (B.102) that when $k = 0$

$$\max_{j\in\mathcal{G}^{(0)},\ell\in[p]} |\sum_{t=1}^{n} z_{j,t-1}e_{\ell,t}| \overset{\mathrm{p}}{\preccurlyeq} n + n^{1-\frac{\theta}{2}}(\log p)^3. \tag{B.119}$$

Then (B.114) follows by (B.118) and (B.119).

For (B.115), by (B.101), when $k = 0$,

$$\max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} |\sum_{t=1}^{n} z_{j,t-1}x_{\ell,t-1}| \le \max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} |\sum_{t=1}^{n} \zeta_{j,t-1}x_{\ell,t-1}| + \bar{c} \max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} |\sum_{t=1}^{n} \psi_{j,t-1}x_{\ell,t-1}|$$
$$\overset{\mathrm{p}}{\preccurlyeq} n^{1+(\theta\wedge m)}(\log p)^{1+\frac{1}{2r}} + n^{\frac{1}{2}+\theta+\frac{m}{2}}(\log p)^3 \tag{B.120}$$

where the second step applies (B.4) that bounds the cross-product of MI regressors, and (B.104). Similarly, when $k = 0$ we deduce by (B.102) that

$$\max_{j\in\mathcal{G}^{(k)},\ell\in\mathcal{G}^{(m)}} |\sum_{t=1}^{n} z_{j,t-1}x_{\ell,t-1}| \le n^{1+(k\wedge m)}(\log p)^{1+\frac{1}{2r}} + n^{1-\frac{\theta}{2}+k+\frac{m}{2}}(\log p)^3. \tag{B.121}$$

Then (B.115) follows by (B.120) and (B.121). □

Lemma B.8. *Under the conditions in Theorem 2, we have*

$$\sqrt{n^{\theta\wedge k}} \overset{\mathrm{p}}{\preccurlyeq} \min_{j\in\mathcal{H}^{(k)}} \widehat{\tau}_j \overset{\mathrm{p}}{\preccurlyeq} \max_{j\in\mathcal{H}^{(k)}} \widehat{\tau}_j \overset{\mathrm{p}}{\preccurlyeq} \sqrt{n^{\theta\wedge k}}. \tag{B.122}$$

Proof of Lemma B.8. When $k = 1$, by decomposition (B.101)

$$(z_{j,t} - \bar{z}_j)^2 = (\zeta_{j,t} - \bar{\zeta}_j)^2 + \frac{c_j^*}{n}(\zeta_{j,t} - \bar{\zeta}_j)(\psi_{j,t} - \bar{\psi}_j) + \frac{c_j^{*2}}{n^2}(\psi_{j,t} - \bar{\psi}_j)^2 \tag{B.123}$$

Recall that $\zeta_{j,t}$ is an MI process with degree of persistence θ. By (B.103) and (B.37), we have

$$\max_{j\in\mathcal{G}^{(k)}}|\frac{1}{n}\sum_{t=1}^{n}(\zeta_{j,t}-\bar{\zeta}_j)(\psi_{j,t}-\bar{\psi}_j)|\overset{\mathrm{p}}{\preccurlyeq} n^{\frac{1}{2}+\theta+\frac{\theta}{2}}(\log p)^3=n^{\frac{3\theta}{2}+\frac{1}{2}}(\log p)^{4.5}.$$

Therefore,

$$\max_{j\in\mathcal{G}^{(k)}}|\frac{1}{n}\sum_{t=1}^{n}\frac{c_j^*}{n}(\zeta_{j,t}-\bar{\zeta}_j)(\psi_{j,t}-\bar{\psi}_j)|\overset{\mathrm{p}}{\preccurlyeq} n^{\frac{3\theta}{2}-\frac{1}{2}}(\log p)^{4.5}=o(n^{\theta}). \tag{B.124}$$

By (B.6),

$$\max_{j\in\mathcal{G}^{(k)}}|\frac{1}{n}\sum_{t=1}^{n}\frac{c_j^{*2}}{n}(\psi_{j,t}-\bar{\psi}_j)^2|\overset{\mathrm{p}}{\preccurlyeq} n^{2\theta-1}(\log p)^{3/2}=o(n^{\theta}). \tag{B.125}$$

By (B.123), (B.124), and (B.125), we have

$$\max_{j\in\mathcal{G}^{(k)}}|\frac{1}{n}\sum_{t=1}^{n}(z_{j,t}-\bar{z}_j)^2-\frac{1}{n}\sum_{t=1}^{n}(\zeta_{j,t}-\bar{\zeta}_j)^2|=o_p(n^{\theta}). \tag{B.126}$$

By (B.2),

$$n^{\theta}\overset{\mathrm{p}}{\preccurlyeq}\min_{j\in\mathcal{G}^{(k)}}\frac{1}{n}\sum_{t=1}^{n}(\zeta_{j,t}-\bar{\zeta}_j)^2\overset{\mathrm{p}}{\preccurlyeq}\max_{j\in\mathcal{G}^{(k)}}\frac{1}{n}\sum_{t=1}^{n}(\zeta_{j,t}-\bar{\zeta}_j)^2\overset{\mathrm{p}}{\preccurlyeq} n^{\theta}, \tag{B.127}$$

and thus (B.122) follows by (B.126) and (B.127).

When $\gamma=0$, (B.122) can be verified in a parallel way utilizing the decomposition (B.102). We complete the proof of Lemma B.8. □

$Q_n=\mathrm{diag}(\{n^{\phi_j}\}_{j\in[p]})$. Additionally define

$$r_{j,t}^*=z_{j,t}-x_{-j,t}^{\top}\varphi^{[j]*},\ \text{ and } r_t^*=(r_{1,t}^*,r_{2,t}^*,\ldots,r_{p,t}^*)^{\top} \tag{B.128}$$

as the true error term in the auxiliary regression.

Lemma B.9. *Under the conditions in Theorem 2, we have*

$$\mathbb{E}\left(\max_{j\in[p],t\in[n]}\left|\frac{r_{j,t-1}^*u_t}{\sqrt{n\Pi_{j,j}}}\right|^{\varpi}\middle|\{r_{t-1}^*\}_{t\in[n]}\right)=O_p\left(\left|\frac{(\log p)^5}{n}\right|^{\varpi/2}\right)$$

for any fixed constant $\varpi>0$.

Proof of Lemma B.9. By the decomposition (B.135),

$$r_{j,t}^*=v_{j,t}-x_{-j,t}^{\top}\varphi^{[j]*}+O\left(\frac{1}{n^{\theta\vee\gamma_j}}\right)\psi_{j,t},$$

where $v_{j,t}$ is an MI process with degree of persistence $\theta \wedge \gamma_j$. Recall from the definition of $\varphi^{[j]*}$ in Section A that $\varphi_\ell^{[j]*}$ is nonzero if and only if $x_{\ell,t}$ shares the same degree of persistence with $v_{j,t}$. Therefore, $v_{j,t} - x_{-j,t}^\top \varphi^{[j]*}$ is a linear combination of MI processes with degree of persistence $\theta \wedge \gamma_j$. Therefore

$$\max_{j\in[p],t\in[n]} \frac{|v_{j,t} - x_{-j,t}^\top \varphi^{[j]*}|}{\sqrt{n^{\theta\wedge\gamma_j}}} \overset{\mathrm{p}}{\preccurlyeq} (\log p)^{3\varpi/2}. \tag{B.129}$$

Combining (B.103) and (B.129), we have

$$\max_{j\in[p],t\in[n]} \frac{|r_{j,t}^*|}{\sqrt{n^{\theta\wedge\gamma_j}}} \overset{\mathrm{p}}{\preccurlyeq} (\log p)^{3\varpi/2}. \tag{B.130}$$

For simplicity, use $\mathbb{E}_r(\cdot)$ to denote $\mathbb{E}(\cdot|\{r_{t-1}^*\}_{t\in[n]})$. By the sub-exponential distribution of u_t, we have $\mathbb{E}(\max_{t\in[n]} |u_t|^\varpi) = O((\log p)^\varpi)$. Therefore, the conditional mean $\mathbb{E}(\max_{t\in[n]} |u_t|^\varpi|\{r_{t-1}^*\}_{t\in[n]})$ is a random variable with a mean value of order $O((\log p)^\varpi)$, and thus

$$\mathbb{E}(\max_{t\in[n]} |u_t|^\varpi|\{r_{t-1}^*\}_{t\in[n]}) = O_p((\log p)^\varpi). \tag{B.131}$$

Recall that $\Pi_{j,j} = n^{\theta\wedge\gamma_j}\widetilde{\Pi}_{j,j}$, where $\widetilde{\Pi}_{j,j}$ is bounded from above and below by some absolute constants uniformly for all $j \in [p]$. Therefore, by (B.130) and (B.131), we have

$$\begin{aligned}
\mathbb{E}_r\left(\max_{j\in[p],t\in[n]} \left|\frac{r_{j,t}^* u_t}{\sqrt{n\Pi_{j,j}}}\right|^\varpi\right) &\le \sqrt{n^{-\varpi/2} \max_{j\in[p],t\in[n]} \left|\frac{r_{j,t}^*}{\sqrt{n^{\theta\wedge\gamma_j}\widetilde{\Pi}_{j,j}}}\right|^{2\varpi} \max_{j\in[p],t\in[n]} \mathbb{E}_r\left|u_t\right|^{2\varpi}} \\
&= O_p\left(\left|\frac{(\log p)^5}{n}\right|^{\varpi/2}\right).
\end{aligned}$$

We complete the proof of Lemma B.9. □

Recall that $\widehat{r}_{j,t}$ is the residual of the auxiliary LASSO regression (23). Define

$$\check{r}_{j,t} = \widehat{\tau}_j \widehat{r}_{j,t}, \text{ and } \check{r}_t = (\check{r}_{1,t}, \check{r}_{2,t}, \ldots, \check{r}_{p,t})^\top \tag{B.132}$$

as the residual standardized by the sample s.d. of the instrument, denoted as $\widehat{\tau}_j$ in (22), so that $r_{j,t}^*$ is a population truth of $\check{r}_{j,t}$. Finally, define

$$\Pi_n = \frac{1}{n}\sum_{t=1}^n r_{t-1}^* r_{t-1}^{*\top}, \ \widehat{\Pi}_n = \frac{1}{n}\sum_{t=1}^n \check{r}_{t-1}\check{r}_{t-1}^\top.$$

Lemma B.10. *Under the conditions in Theorem 2, there exists a nonrandom matrix* Π *such that*

$$\left\|Q_n^{-1/2}(\Pi - \Pi_n)Q_n^{-1/2}\right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^3}{\sqrt{n^{\delta_{\min}}}}, \tag{B.133}$$

and

$$\left\|Q_n^{-1/2}(\Pi - \widehat{\Pi}_n)Q_n^{-1/2}\right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}, \tag{B.134}$$

where $\delta_{\min} = \theta \wedge (1-\theta)$.

Proof of Lemma B.10. **We first prove (B.133).** By decomposition (B.101) and decomposition (B.101), we can summarize that

$$z_{j,t} = v_{j,t} + O\left(\frac{1}{n^{\theta \vee \gamma_j}}\right)\psi_{j,t} \tag{B.135}$$

where

$$v_{j,t} = \zeta_{j,t}\mathbf{1}\{\theta < \gamma_j\} + x_{j,t}\mathbf{1}\{\theta > \gamma_j\} \tag{B.136}$$

is an AR(1) process with innovation $e_{j,t}$ and AR coefficient either $\rho_z\mathbf{1}\{\theta < \gamma_j\} + \rho_j\mathbf{1}\{\theta > \gamma_j\}$. Recall that $\phi_j = \theta \wedge \gamma_j$. Therefore, $v_{j,t}$ is MI or stationary with degree of persistence $\phi_j = \theta \wedge \gamma_j$. Define

$$w_{j,t}^* = v_{j,t} - x_{-j,t}^\top \varphi^{[j]*}$$

with $z_{j,t}$ replaced by $v_{j,t}$ in the definition of $r_{j,t}^*$ in (B.128). The proof will consist of the following steps:

(I) Prove

$$\max_{j,\ell\in[p]}\left|n^{-1}\sum_{t=1}^n \frac{r_{j,t-1}^* r_{\ell,t-1}^*}{\sqrt{n^{\phi_j+\phi_\ell}}} - n^{-1}\sum_{t=1}^n \frac{w_{j,t-1}^* w_{\ell,t-1}^*}{\sqrt{n^{\phi_j+\phi_\ell}}}\right| \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^3}{\sqrt{n^{\delta_{\min}}}}.$$

(II) Show that there exists a $\Pi = (\Pi_{j,\ell})$ such that

$$\max_{j,\ell\in[p]}\left|n^{-1}\sum_{t=1}^n \frac{w_{j,t-1}^* w_{\ell,t-1}^*}{\sqrt{n^{\phi_j+\phi_\ell}}} - \Pi_{j,\ell}\right| \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{\sqrt{n^{\delta_{\min}}}}.$$

Then (B.133) follows by the results of (I) and (II), and the triangular inequality.

Proof of (I). Recall that $\phi_j = \theta \wedge \gamma_j = \theta$ for all $j \in \mathcal{G}^{(1)}$ and $\phi_j = 0$ for $j \in \mathcal{G}^{(0)}$. By the definition of $\varphi^{[j]*}$ in Section A, we have $\varphi_\ell^{[j]*} \neq 0$ if and only if $\phi_\ell = \phi_j$. Therefore, $w_{j,t}^* = v_{j,t} - x_{-j,t}^\top \varphi^{[j]*}$ is a linear combination of MI variables with degree of persistence ϕ_j.

Define $\phi^{(m)} := m \wedge \theta$. By the upper bound of L_1-norm for $\varphi^{[j]*}$ in Assumption 6, we have

$$\sup_{j\in\mathcal{G}^{(m)},t\in[n]} |w^*_{j,t}| \overset{\text{p}}{\preccurlyeq} \sqrt{n^{\phi^{(m)}}}(\log p)^{3/2},$$

following (B.1) for the upper bound of MI regressors, which implies

$$\sup_{j\in[p],t\in[n]} \frac{|w^*_{j,t}|}{\sqrt{n^{\phi_j}}} \le \sup_{m\in[M]} \sup_{j\in\mathcal{G}^{(m)},t\in[n]} \frac{|w^*_{j,t}|}{\sqrt{n^{\phi^{(m)}}}} \overset{\text{p}}{\preccurlyeq} (\log p)^{3/2}. \tag{B.137}$$

In addition,

$$\begin{aligned}\sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t} - w^*_{j,t}|}{\sqrt{n^{\phi_j}}} &= O\left(\sup_{j\in[p],t\in[n]} \frac{|\psi_{j,t}|}{n^{\theta\vee\gamma_j}\sqrt{n^{\phi_j}}}\right) \\ &\overset{\text{p}}{\preccurlyeq} \sup_{j\in[p],t\in[n]} \frac{n^{\frac{\theta\vee\gamma_j}{2}+\phi_j}(\log p)^{3/2}}{n^{\theta\vee\gamma_j}\sqrt{n^{\phi_j}}} \\ &= \sup_{j\in[p],t\in[n]} \frac{(\log p)^{3/2}}{\sqrt{n^{(\theta\vee\gamma_j)-(\theta\wedge\gamma_j)}}} \le \sqrt{\frac{(\log p)^3}{n^{\delta_{\min}}}},\end{aligned} \tag{B.138}$$

where the first line applies (B.135), and the second line applies (B.103). Then by (B.137) and (B.138), we have

$$\sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t}|}{\sqrt{n^{\phi_j}}} \le \sup_{j\in[p],t\in[n]} \frac{|w^*_{j,t}|}{\sqrt{n^{\phi_j}}} + \sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t} - w^*_{j,t}|}{\sqrt{n^{\phi_j}}} \overset{\text{p}}{\preccurlyeq} (\log p)^{3/2} \tag{B.139}$$

by (B.137) and (B.138). Therefore,

$$\begin{aligned}&\max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{r^*_{j,t-1}r^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}} - n^{-1}\sum_{t=1}^n \frac{w^*_{j,t-1}w^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}}\right| \\ &\le \max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{(r^*_{j,t-1}-w^*_{j,t-1})r^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}}\right| + \max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{w^*_{j,t-1}(r^*_{\ell,t-1}-w^*_{\ell,t-1})}{\sqrt{n^{\phi_j+\phi_\ell}}}\right| \\ &\le \sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t} - w^*_{j,t}|}{\sqrt{n^{\phi_j}}} \sup_{\ell\in[p],t\in[n]} \frac{|r^*_{\ell,t}|}{\sqrt{n^{\phi_\ell}}} + \sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t} - w^*_{j,t}|}{\sqrt{n^{\phi_j}}} \sup_{\ell\in[p],t\in[n]} \frac{|w^*_{j,t}|}{\sqrt{n^{\phi_j}}} \\ &\overset{\text{p}}{\preccurlyeq} \frac{(\log p)^3}{\sqrt{n^{\delta_{\min}}}},\end{aligned}$$

where the last inequality applies (B.137), (B.138), and (B.139). We complete the proof of (I).

Proof of (II). Recall that $w^*_{j,t} = v_{j,t} - x^\top_{-j,t}\varphi^{[j]*}$ is a linear combination of MI variables with degree of persistence ϕ_j. Therefore, when $\phi_\ell \neq \phi_j$,by (B.4)

$$\max_{j,\ell\in[p]} |n^{-1}\sum_{t=1}^{n}\frac{w^*_{j,t-1}w^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}}| \leq \max_{j,\ell\in[p]} \frac{n^{\phi_j\wedge\phi_\ell}(\log p)^{1+\frac{1}{2r}}}{\sqrt{n^{\phi_j+\phi_\ell}}} = \frac{(\log p)^{1+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}, \tag{B.140}$$

and thus (II) follows by taking $\Pi_{j,\ell}=0$. When $\phi_j=\phi_\ell$, recall that $\phi_j=\theta\wedge\gamma_j$. Therefore,

$$w^*_{j,t-1}w^*_{\ell,t-1} = v_{j,t-1}v_{\ell,t-1} - v_{j,t-1}x^\top_{-\ell,t}\varphi^{[\ell]*} - v_{\ell,t-1}x^\top_{-j,t}\varphi^{[j]*} + \varphi^{[\ell]*\top}x_{-\ell,t-1}x^\top_{-j,t}\varphi^{[j]*}.$$

Without loss of generality, suppose $j,\ell\in\mathcal{G}^{(k)}$ so that $\gamma_j=\gamma_\ell=k$. By (B.87) and the definition of $\varphi^{[j]*}$ in Section A, we have

$$\max_{j,\ell\in\mathcal{G}^{(k)}}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{w^*_{jt-1}w^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}} - \widetilde{\Pi}_{j,\ell}\right| \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^{3+\frac{2}{r}}}{n^{\phi(k)}} + \frac{(\log p)^{3+\frac{1}{2r}}}{\sqrt{n^{1-\phi(k)}}} \leq \frac{(\log p)^{3+\frac{2}{r}}}{\sqrt{n^{\delta_{\min}}}}, \tag{B.141}$$

$$\text{where } \widetilde{\Pi}_{j,\ell} = \Sigma^{(k)}_{j,\ell} - \Sigma^{(k)}_{j,-\ell}(\Sigma^{(k)}_{-\ell,-\ell})^{-1}\Sigma^{(k)}_{-\ell,\ell} - \Sigma^{(k)}_{\ell,-j}(\Sigma^{(k)}_{-j,-j})^{-1}\Sigma^{(k)}_{-j,j} + \Sigma^{(k)\top}_{-\ell,\ell}(\Sigma^{(k)}_{-\ell,-\ell})^{-1}\Sigma^{(k)}_{-\ell,-j}(\Sigma^{(k)}_{-j,-j})^{-1}\Sigma^{(k)}_{-j,j}.$$

When $j=\ell$, $\widetilde{\Pi}_{j,\ell}$ is simplified as

$$\widetilde{\Pi}_{j,j} = \Sigma^{(k)}_{j,j} - \Sigma^{(k)}_{j,-j}(\Sigma^{(k)}_{-j,-j})^{-1}\Sigma^{(k)}_{-j,j}. \tag{B.142}$$

Then (II) is implied by (B.140) with $\Pi_{j,\ell}=0$, and taking the maximum over $k\in[K]$ of (B.141) with

$$\Pi_{j,\ell} = n^{\phi_j+\phi_\ell}\widetilde{\Pi}_{j,\ell}. \tag{B.143}$$

We complete the proof of (B.133).

We next prove (B.133). By the triangular inequality if suffices to show

$$\left\|Q_n^{-1/2}(\Pi_n-\widehat{\Pi}_n)Q_n^{-1/2}\right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}. \tag{B.144}$$

Define $\check{\varphi}^{[j]} = \widehat{\tau}_j\widehat{\varphi}^{[j]}$, where $\widehat{\tau}_j$ is the sample s.d. of the IV in (22), and $\widehat{\varphi}^{[j]}$ is the auxiliary

LASSO estimator in (23). Then

$$
\begin{aligned}
\max_{j\in[p],t\in[n]} \frac{|r^*_{j,t} - \check{r}^*_{j,t}|}{\sqrt{n^{\phi_j}}} &= \max_{j\in[p],t\in[n]} \frac{|x^\top_{-j,t}(\check{\varphi}^{[j]} - \varphi^{[j]*})|}{\sqrt{n^{\theta\wedge\gamma_j}}} \\
&\le \max_{j\in[p],t\in[n]} \frac{\|D^{-1}_{-j}x_{-j,t}\|_\infty \cdot \widehat{\tau}_j \|D_{-j}(\widehat{\varphi}^{[j]} - \varphi^{[j]*})\|_1}{\sqrt{n^{\theta\wedge\gamma_j}}} \\
&\overset{\mathrm{p}}{\preccurlyeq} (\log p)^2 \max_{k\in[K]} \max_{j\in\mathcal{G}^{(k)},t\in[n]} \frac{\widehat{\tau}_j \|D_{-j}(\widehat{\varphi}^{[j]} - \varphi^{[j]*})\|_1}{\sqrt{n^{\theta\wedge\gamma_j}}} \\
&\overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{\frac{19}{2}+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}},
\end{aligned} \tag{B.145}
$$

where the third row applies (B.1) and Proposition B.2, and the last inequality applies Lemma B.10 and Proposition B.4. Therefore,

$$
\begin{aligned}
&\max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{r^*_{j,t-1}r^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}} - n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}\check{r}_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}} \right| \\
&\le \max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{2(r^*_{j,t-1} - \check{r}_{j,t-1})r^*_{\ell,t-1}}{\sqrt{n^{\phi_j+\phi_\ell}}} \right| + \max_{j,\ell\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{(r^*_{j,t-1} - \check{r}_{j,t-1})(r^*_{\ell,t-1} - \check{r}_{\ell,t-1})}{\sqrt{n^{\phi_j+\phi_\ell}}} \right| \\
&\le \sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t-1} - \check{r}_{j,t-1}|}{\sqrt{n^{\phi_j}}} \sup_{\ell\in[p],t\in[n]} \frac{|r^*_{\ell,t}|}{\sqrt{n^{\phi_\ell}}} + \left(\sup_{j\in[p],t\in[n]} \frac{|r^*_{j,t-1} - \check{r}_{j,t-1}|}{\sqrt{n^{\phi_j}}} \right)^2 \\
&\overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}},
\end{aligned}
$$

where the last inequality applies (B.139) and (B.145). We complete the proof of (B.144) and thus (B.134) is verified. We end the proof of Lemma B.10. □

Define

$$
D_\Pi = \mathrm{diag}(\Pi),\ \widehat{D}_{\Pi,n} = \mathrm{diag}(\widehat{\Pi}_n)
$$

Lemma B.11. *Under the conditions in Theorem 2, we have*

$$
\left\| Q_n^{1/2}(\widehat{D}^{-1/2}_{\Pi,n} - D^{-1/2}_\Pi) \right\|_\infty \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}, \tag{B.146}
$$

where $\delta_{\min} = \theta \wedge (1-\theta)$.

Proof of Lemma 1. Note that

$$
\begin{aligned}
\left\| Q_n^{1/2}(\widehat{D}_{\Pi,n}^{-1/2} - D_{\Pi}^{-1/2}) \right\|_{\infty} &= \max_{j\in[p]} \frac{\sqrt{n^{\phi_j}} |n^{-1}\sum_{t=1}^n \check{r}_{j,t-1}^2 - \Pi_{j,j}|}{\left(\sqrt{n^{-1}\sum_{t=1}^n \check{r}_{j,t-1}^2} + \sqrt{\Pi_{j,j}}\right)\sqrt{n^{-1}\sum_{t=1}^n \check{r}_{j,t-1}^2 \Pi_{j,j}}} \\
&\leq \max_{j\in[p]} \frac{\left| n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}} - \widetilde{\Pi}_{j,j} \right|}{\left(\sqrt{n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}}} + \sqrt{\widetilde{\Pi}_{j,j}}\right)\sqrt{n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}} \widetilde{\Pi}_{j,j}}} \\
&\leq \frac{\left\| Q_n^{-1/2}(\Pi - \widehat{\Pi}_n) Q_n^{-1/2} \right\|_{\infty}}{\min_{j\in[p]}\left(\sqrt{n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}}} + \sqrt{\widetilde{\Pi}_{j,j}}\right)\sqrt{n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}} \widetilde{\Pi}_{j,j}}}.
\end{aligned}
$$

By (B.141), we have

$$
\min_{j\in[p]} n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}} \geq \min_{j\in[p]} \widetilde{\Pi}_{j,j} - \max_{j\in[p]} \left| n^{-1}\sum_{t=1}^n \frac{\check{r}_{j,t-1}^2}{n^{\phi_j}} - \widetilde{\Pi}_{j,j} \right| \geq 0.5 \min_{j\in[p]} \widetilde{\Pi}_{j,j}
$$

with a sufficiently large n. Therefore, by (B.134)

$$
\left\| Q_n^{1/2}(\widehat{D}_{\Pi,n}^{-1/2} - D_{\Pi}^{-1/2}) \right\|_{\infty} \leq \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}} \Big/ \left(0.5(\min_{j\in[p]} \widetilde{\Pi}_{j,j})^{1.5}\right) = O\left(\frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}}\right).
$$

We complete the proof of Lemma (B.11). □

Define

$$
V = D_{\Pi}^{-1/2}\Pi D_{\Pi}^{-1/2},\ V_n = D_{\Pi}^{-1/2}\Pi_n D_{\Pi}^{-1/2},\ W_n = \widehat{D}_{\Pi,n}^{-1/2}\widehat{\Pi}_n \widehat{D}_{\Pi,n}^{-1/2}.
$$

Lemma B.12. *Under the conditions in Theorem 2, we have*

$$
\|V - V_n\|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} \frac{(\log p)^3}{\sqrt{n^{\delta_{\min}}}} \tag{B.147}
$$

$$
\|V_n - W_n\|_{\infty} \overset{\mathrm{p}}{\preccurlyeq} \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}} \tag{B.148}
$$

Proof of Lemma B.12. Note that each diagonal entry in D_{Π} is $\Pi_{j,j} = n^{\phi_j}\widetilde{\Pi}_{j,j}$ according to

(B.143), where $\widetilde{\Pi}_{j,j}$ defined in (B.142) has an $O(1)$ order.

$$\|D_\Pi^{-1}Q_n\|_\infty \le \max_{j\in[p]} \frac{n^{\phi_j}}{n^{\phi_j}\widetilde{\Pi}_{j,j}} = O(1). \tag{B.149}$$

Then

$$\|V - V_n\|_\infty \le \|Q_n^{-1/2}(\Pi - \Pi_n)Q_n^{-1/2}\|_\infty \cdot \|D_\Pi^{-1/2}Q_n^{1/2}\|_\infty \overset{\mathrm{p}}{\precsim} \frac{(\log p)^3}{\sqrt{n^{\delta_{\min}}}},$$

where the last inequality applies (B.133).

For (B.148), note that by $\Pi_{j,\ell} = \sqrt{n^{\phi_j+\phi_\ell}}\widetilde{\Pi}_{j,\ell}$ according to (B.143), we have

$$\|Q_n^{-1/2}\Pi Q_n^{-1/2}\|_\infty = O(1). \tag{B.150}$$

By (B.134) we further have

$$\|Q_n^{-1/2}\widehat{\Pi}_n Q_n^{-1/2}\|_\infty \overset{\mathrm{p}}{\precsim} 1. \tag{B.151}$$

By (B.149) and (B.11), the triangular inequality yields

$$\|\widehat{D}_{\Pi,n}^{-1}Q_n\|_\infty \overset{\mathrm{p}}{\precsim} 1. \tag{B.152}$$

We thus have

$$\begin{aligned}
\|V_n - W_n\|_\infty =& \|D_\Pi^{-1/2}\Pi D_\Pi^{-1/2} - \widehat{D}_{\Pi,n}^{-1/2}\widehat{\Pi}_n\widehat{D}_{\Pi,n}^{-1/2}\|_\infty \\
=& \|(D_\Pi^{-1/2} - \widehat{D}_{\Pi,n}^{-1/2})Q_n^{1/2}\|_\infty \|Q_n^{-1/2}\Pi Q_n^{-1/2}Q_n^{1/2}D_\Pi^{-1/2}\|_\infty \\
&+ \|\widehat{D}_{\Pi,n}^{-1/2}Q_n^{1/2}Q_n^{-1/2}\widehat{\Pi}_n Q_n^{-1/2}\|_\infty \|Q_n^{1/2}(\widehat{D}_{\Pi,n}^{-1/2} - D_\Pi^{-1/2})\|_\infty + \\
&+ \|\widehat{D}_{\Pi,n}^{-1/2}Q_n^{1/2}\|_\infty^2 \|Q_n^{-1/2}(\widehat{\Pi}_n - \Pi)Q_n^{-1/2}\|_\infty \\
=& O(1)\|(D_\Pi^{-1/2} - \widehat{D}_{\Pi,n}^{-1/2})Q_n^{1/2}\|_\infty + O(1)\|Q_n^{-1/2}(\widehat{\Pi}_n - \Pi)Q_n^{-1/2}\|_\infty \\
\overset{\mathrm{p}}{\precsim}& \frac{s^2(\log p)^{11+\frac{1}{2r}}}{\sqrt{n^{\delta_{\min}}}},
\end{aligned}$$

where the third row applies (B.149), (B.150), (B.151), and (B.152), and the last inequality applies (B.134) and (B.146). We complete the proof of Lemma B.12. □