An Improved Inference for IV Regressions*

Liyu Dou[a], Pengjin Min[a], Wenjie Wang[b], and Yichong Zhang[a]

[a] *Singapore Management University, Singapore*
[b] *Nanyang Technological University, Singapore*

October 26, 2026

Abstract

Empirical applications of instrumental variables (IV) regressions often involve both low-dimensional and high-dimensional sets of instruments. In Angrist and Krueger (1991), the quarter-of-birth instruments generate a much larger set that includes many interaction-based technical instruments; in shift-share designs considered in Goldsmith-Pinkham, Sorkin, and Swift (2020), a single IV aggregates many underlying base instruments. To improve inference in such settings, this paper proposes a combination test that integrates low- and high-dimensional IV specifications. The test linearly combines a cluster-robust Wald statistic based on low-dimensional IVs with leave-cluster-out Lagrange multiplier (LM) and Anderson-Rubin (AR) statistics constructed from the many IVs. We establish the joint asymptotic normality of these statistics and show that the proposed test is asymptotically optimal within a broad class of tests. The procedure yields costless efficiency improvements, automatically adapts to weak identification of the many instruments, and provides a practical rule of thumb for assessing these gains.

Keywords: Many Weak Instruments, Shift-Share Instruments, Combination Test.
JEL Classification: C12, C36, C55

*I am especially grateful to my adviser, Yichong Zhang, for his continued guidance and support. I would like to thank my committee members, Jia Li, Liyu Dou and Wenjie Wang, for their valuable comments and suggestions. I would also like to thank Michal Kolesár, Taisuke Otsu, Andrew Patton, Peter C. B. Phillips, Mikkel Sølvsten and Liyang Sun for helpful discussions on this paper.

1 Introduction

Empirical applications of instrumental variables (IV) regressions in economics often involve several alternative sets of instruments, some with dimensionality proportional to the sample size and others considerably more parsimonious. A canonical example is Angrist and Krueger (1991), which estimates returns to schooling using three quarter-of-birth (QoB) dummies and their interactions with state- and/or year-of-birth dummies. Including fully interacted three-way interactions yields a total of 1,530 instruments (Mikusheva and Sun, 2022). As noted by Mikusheva and Sun (2024), such settings are common in empirical practice: one starts with a low-dimensional source of exogenous variation (e.g., the QoB dummies) and then constructs numerous 'technical' instruments to exploit a heterogeneous and/or nonlinear first stage.

A different but related class of examples arises in the context of shift-share IVs, which are now widely used in labor, public, development, macroeconomics, international trade, and finance.[1] In these examples, the set of many instruments emerges from the construction structure of shift-share IVs. This follows from the observation in Goldsmith-Pinkham et al. (2020) that, when the shares are exogenous, a shift-share IV can be interpreted as an aggregation of many underlying base instruments. For one example, in their empirical application following Card (2009), the shift-share IV used to estimate the (negative) inverse elasticity of substitution between immigrants and natives is constructed from 38 base instruments (the respective shares of immigrants from 38 home countries) in a sample of 124 observations.

This naturally raises the question of whether one can take the low-dimensional IV regression as the benchmark specification and combine it with its many-IV counterpart to enhance the power of IV inference essentially for free, in the sense that no additional identifying assumptions are imposed. In this paper, we provide a constructive solution for clustered settings with many clusters of bounded size and cluster fixed effects, including important panel data applications involving shift-share IVs.[2] Additionally, our procedure remains valid even when the many-IV specification is weakly identified, so empirical researchers need not be concerned about the size distortions and confidence interval under-coverage that may arise with many weak instruments.

[1] See, for instance, Bartik (1991), Blanchard, Katz, Hall, and Eichengreen (1992), Adao, Kolesár, and Morales (2019), Goldsmith-Pinkham et al. (2020), Borusyak, Hull, and Jaravel (2022), Borusyak, Hull, and Jaravel (2025), and references therein.

[2] See Section 3.2 for an illustrative example.

Specifically, we consider an optimal combined inference procedure based on three familiar test statistics for IV regressions: (i) a standard cluster-robust Wald statistic constructed from the low-dimensional IVs; (ii) a leave-cluster-out jackknife Lagrange multiplier (LM) statistic; and (iii) a leave-cluster-out jackknife Anderson–Rubin (AR) statistic. The LM and AR statistics are constructed using the many-IV specification, and the leave-cluster-out design removes the many-IV bias under within-cluster error dependence.

We first show that, under the null hypothesis and local alternatives, the Wald, LM, and AR statistics are jointly asymptotically normal, provided that the low-dimensional IVs strongly identify the parameter of interest, while allowing the many-IV specification to be weakly identified. Standard optimal testing theory then implies that, in the corresponding Gaussian limit experiment, the uniformly most powerful unbiased (UMPU) test rejects for large absolute values of an appropriate linear combination of the three limiting Gaussian statistics.[3] Our proposed test, as a function of the three statistics, is asymptotically equivalent to this UMPU test and is therefore weakly more powerful than each of the Wald, LM, and AR tests.

Importantly, the combination test adapts automatically to the strength of identification in the many-IV specification. When the parameter of interest is weakly identified under many IVs in the sense of Mikusheva and Sun (2022), the optimal linear combination places asymptotically negligible weight on the LM and AR components, so that the resulting test reduces to the Wald test. Consequently, the procedure remains valid and retains the same asymptotic power as the standalone Wald test, so there is no loss from incorporating the many-IV specification.

The notion of optimality in our study is defined relative to the class of tests based on the Wald, LM, and AR statistics and is used primarily to assess the improvement in inference within this class. We do not seek a globally optimal test by searching over all combinations of base IVs or by aggregating results across alternative low-dimensional specifications. Instead, we take the low-dimensional specification as the *benchmark* and combine its associated test statistic with those based on a *given* set of many IVs, thereby strengthening inference while remaining agnostic about both the identification strength of the many-IV specification and the functional relationship between the many and low-dimensional IVs. Nonetheless, the resulting combination

[3]See, for instance, Section 4.2 of Lehmann and Romano (2006) and Lemma 2.2 of Lim, Wang, and Zhang (2024a) for formal arguments.

test is asymptotically optimal in the sense of Müller (2011). We also emphasize that our goal is not to derive the "optimal" instrument in the many-IV setting. When the number of instruments is large, consistent estimation of the optimal instrument is generally impossible (without extra assumptions on its form), because estimation errors accumulate across instruments and induce many-IV bias.[4] Although jackknife-type debiasing can remove this bias, it also alters the estimator's asymptotic variance, thereby invalidating the classical optimality interpretation of the optimal IV.

The confidence interval implied by the combination test has the usual "estimator plus and minus a standard error times a critical value" form. Its center is an estimator that linearly combines a standard GMM estimator based on the low-dimensional IVs with a leave-cluster-out jackknife IV estimator based on many IVs and the AR statistic, using weights that capture both the relative identification strength from the two IV sets and the UMPU weights. The efficiency gain manifests itself as an almost surely shorter confidence interval. We measure this gain by the percentage reduction in the length of the resulting confidence interval relative to that of the Wald test based on the low-dimensional IVs. As an illustration, in the immigrant enclave application of Card (2009), the combination procedure reduces the length by between 5% and 17%. The magnitude of this reduction is determined primarily by the identification strengths from the two IV sets, together with the limiting correlations between the Wald and LM statistics, and between the LM and AR statistics. In Section 2, we translate these relationships into a practical rule of thumb based solely on the ratio of the standard errors of the low-dimensional and many-IV estimators, which can be read directly from reported regression tables prior to implementing our combination procedure. We illustrate this rule using the application of Card (2009).

Relation to the literature. This paper contributes to the large literature on many (weak) instruments,[5] and is particularly related to Lim et al. (2024a). Building on Andrews (2016), they propose a jackknife conditional linear combination (CLC) test

[4]See, for instance, Sections 3 and 4 of Mikusheva (2021) for detailed discussions on this issue.

[5]See, for instance, Morimune (1983), Bekker (1994), Donald and Newey (2001), Chao and Swanson (2005), Hansen, Hausman, and Newey (2008), Kuersteiner and Okui (2010), Anatolyev and Gospodinov (2011), Belloni, Chen, Chernozhukov, and Hansen (2012), Carrasco (2012), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Hausman, Newey, Woutersen, Chao, and Swanson (2012), Kolesár (2013), Carrasco and Tchuente (2015), Wang and Kaffo (2016), Kolesár (2018), Evdokimov and Kolesár (2018), Sølvsten (2020), Chao, Swanson, and Woutersen (2023), Boot and Nibbering (2024), Lim, Wang, and Zhang (2024b), Yap (2025), among others.

for independent data that is robust to weak identification, many instruments, and heteroskedasticity, combining jackknife AR, LM, and orthogonalized LM statistics to achieve good power across identification regimes. However, their analysis is confined to many-IV settings and does not apply directly to our framework. In contrast, we explicitly incorporate low-dimensional IVs and develop a unified procedure that efficiently combines estimation and inference from both low- and high-dimensional IV specifications, all within a clustered setting. Different from Lim et al. (2024a), our primary objective is to construct a theoretically grounded combination test that improves upon the conventional Wald test based solely on low-dimensional IVs.[6] We also note that, in a different context, Jiang, Li, Miao, and Zhang (2025) propose an optimal linear combination of adjusted and unadjusted estimators for the average treatment effect under covariate-adaptive randomization.

Our paper is also related to the literature on inference with many IVs and clustered data. Frandsen, Leslie, and McIntyre (2025) study leave-cluster-out jackknife versions of the IV estimator (CJIVE), and Ligtenberg (2025) develop weak-identification-robust procedures in clustered environments. We depart from this literature by not focusing solely on estimation or inference within the many-IV specification itself. Instead, we leverage the many-IV specification to enhance the power of low-dimensional Wald inference through an optimal combination of test statistics. Chao et al. (2023) and Kolesár, Min, Wang, and Zhang (2026) consider settings in which both the number of IVs and the number of control variables are allowed to be proportional to the sample size, posing additional challenges for cluster-robust inference. In this paper, we restrict attention to models with a fixed number of control variables and leave the case with both many IVs and many controls for future research.

Structure of the paper: Section 2 details the rationale behind the proposed rule of thumb and demonstrates its use in an empirical application; practitioners mainly interested in applications may focus on this section directly. Section 3 introduces the

[6]This paper may seem closely related to Qian and Schmidt (1999) because both titles feature the terms "improved" and "instrumental variables." While both analyses study efficiency gains from incorporating additional information beyond benchmark moment conditions, the contexts differ substantially. Qian and Schmidt (1999) analyze efficiency improvements of IV/GMM estimators using extra moment conditions that do not depend on the parameters of interest. In contrast, we focus on improving inference by exploiting potentially high-dimensional, additional moment conditions that explicitly depend on those parameters.

model and key preliminaries, while Section 4 develops the large-sample theory for the combination test and formalizes its theoretical properties. Section 5 presents the practical implementation of the combination test in an empirically relevant setting, employs simulations to evaluate its power properties, and illustrates the procedure in an additional empirical application. We conclude in Section 6. An additional case with weak low-dimensional IVs alongside strong many IVs is presented in the Appendix, together with additional simulation results that demonstrate it. All the proofs are collected in the Supplemental Appendix.

Notation. We write $[n] \equiv \{1, \ldots, n\}$ and $[G] \equiv \{1, \ldots, G\}$. Let A be an $n \times m$ matrix and let $\{n_g\}_{g\in[G]}$ be positive integers with $\sum_{g=1}^{G} n_g = n$. We denote by $A_{[g]}$ the g-th row-wise block of A, of dimension $n_g \times m$. When A is an $n \times n$ square matrix, we denote by $A_{[g,h]}$ the (g,h)-th block of A. For a positive semi-definite square matrix A, denote its largest and smallest eigenvalues by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. Let C be a generic positive constant independent of n, whose value may change from line to line. For brevity, we write $\sum_{g,h\in[G]^2, g\neq h} := \sum_{g\in[G]} \sum_{h\in[G], h\neq g}$.

2 Rule of Thumb and Empirical Illustration

In this section, we develop a practical rule of thumb that can be directly applied to reported estimates and standard errors from low-dimensional and many-IV regressions. We illustrate its empirical relevance using the application in Goldsmith-Pinkham et al. (2020, Section VII), which builds on Card (2009).

We quantify the efficiency gain from the combination test by the percentage reduction in the length of its confidence interval, relative to that of the Wald test, in large samples. Section 4.3 formally derives an analytical expression for this reduction as a function of the identification strengths of both the low-dimensional and many instrumental variables (IVs), as well as the limiting correlations between the Wald and LM statistics and between the LM and AR statistics. Because the percentage reduction is monotonically increasing in the absolute correlation between the LM and AR statistics, we further obtain a lower bound on the efficiency gain by setting this correlation to zero. The bound depends only on the correlation between the Wald and LM statistics (denoted as ρ_1) and on the ratio of the standard deviations of the GMM estimator based on low-dimensional IVs to the leave-cluster-out jackknife IV

estimator (CJIVE) based on many IVs (or, equivalently, on the relative strength of the many IVs to the low-dimensional IVs).

Figure 1 plots the lower bound as a function of the standard deviation ratio for different values of ρ_1, the limiting correlation between the Wald and LM statistics. Two observations emerge. First, for any fixed ρ_1, we show theoretically that whenever the standard deviation ratio exceeds ρ_1, the lower bound on efficiency gains increases with the ratio, reflecting the fact that the combination test exploits the additional precision provided by the LM statistic. Second, once the standard deviation ratio exceeds one, the lower bound decreases as ρ_1 increases, because the LM statistic then contributes relatively little additional information beyond the highly correlated Wald statistic. As a simple rule of thumb that only requires a back of envelope calculation based on the reported standard errors, we propose: for empirically plausible values of ρ_1 (between -0.7 and 0.7), whenever the standard error from the regression with low-dimensional IVs divided by that from the regression with many IVs is greater than 1.05, the corresponding confidence interval is reduced by at least 10%.[7] We further argue in Section 4.3 that the same rule of thumb can be applied to other many-IV estimators such as HLIML and HFUL developed by Hausman et al. (2012).

To illustrate the empirical relevance of efficiency gains and the associated rule of thumb, we implement the combination test in an empirical application that estimates the (negative) inverse elasticity of substitution between immigrants and natives, following Card (2009). As in Goldsmith-Pinkham et al. (2020, Section VII), we examine two separate sets of results by skill group: high school equivalent workers and college equivalent workers. The analysis is based on cross-sectional regressions for each skill group in 124 cities in 2000. The dependent variable is the residual log wage gap between immigrant and native men,[8] and the main regressor of interest is the log ratio of immigrant-to-native hours for both men and women within the same skill group. Here, the immigration shares from 38 origin-country groups in 1980 serve as the underlying base instruments. The one-dimensional Bartik instrument is constructed by aggregating these shares using the corresponding national immigrant inflows for each origin-country group between 1990 and 2000 as weights (see Section 3.2 below and

[7] The thresholds are 0.93 for 5% and 1.24 for 20%. In additional (unreported) plots for $\rho_1 \in [-0.99, 0.99]$, the confidence interval is at least 10% shorter whenever the standard deviation ratio exceeds 1.11 (with thresholds 1.05 for 5% and 1.25 for 20%).

[8] As noted in Card (2009, p.11, footnote 17), residual wages are wages after controlling for observed characteristics of the entire U.S. workforce.

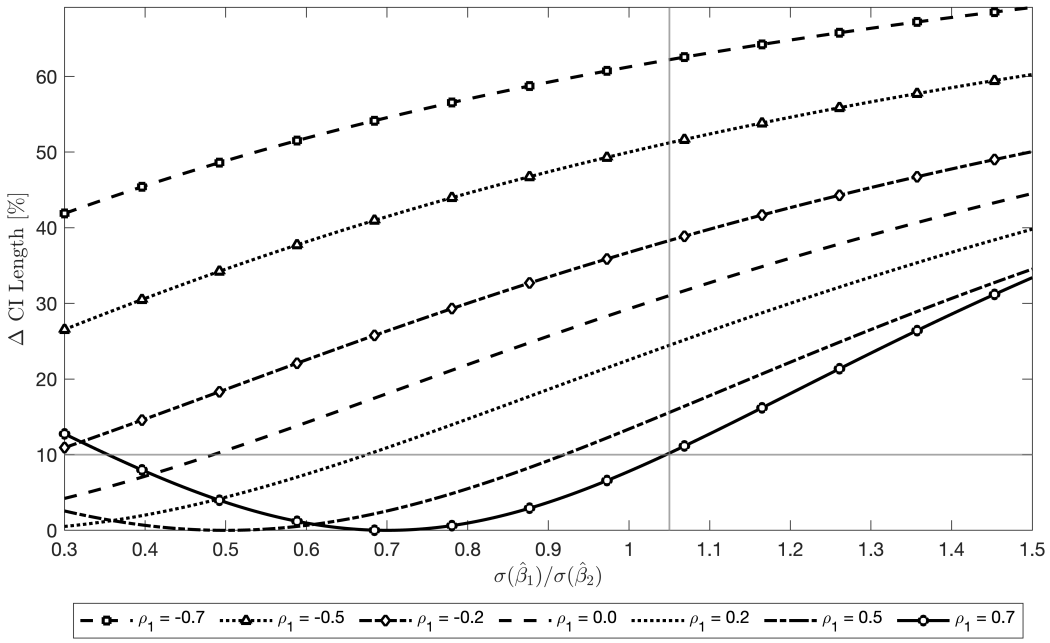


Figure 1: Theoretical lower bounds for percentage reduction in confidence interval length.

Notes: This figure plots the lower bound of the efficiency gain in (4.13) as a function of the standard deviation ratio $\sigma(\hat{\beta}_1)/\sigma(\hat{\beta}_2)$ for various values of ρ_1, the limiting correlation between the Wald and (leave-cluster-out jackknife) LM statistics. The horizontal axis shows the ratio of the standard deviations of $\hat{\beta}_1$ (the efficient GMM estimator using low-dimensional IVs) and $\hat{\beta}_2$ (the leave-cluster-out estimator using many IVs). The vertical axis shows the percentage reduction in confidence interval length.

Goldsmith-Pinkham et al. (2020) for further details). One important source of endogeneity that the instrument addresses is a positive labor demand shock that attracts immigrants to a location disproportionately relative to natives. In this case, the shock increases both relative earnings and relative labor supply, giving rise to endogeneity in the estimation of the elasticity of substitution. See also Goldsmith-Pinkham et al. (2020, Section VII) and Borusyak et al. (2025) for a detailed discussion of why identification in this application can reasonably be attributed to the exogeneity of the shares.

Table 1 reports the point estimates obtained from the two-stage least squares (TSLS) estimator using the Bartik instrument (the low-dimensional IV specification), $\hat{\beta}_1$, together with those from the leave-cluster-out estimator (the many-IV specifica-

tion), $\hat{\beta}_2$, based on the 38 base IVs. We present results separately for specifications that include and exclude city-level controls. Consistent with Goldsmith-Pinkham et al. (2020), the estimates are broadly similar within each skill group. As noted by Goldsmith-Pinkham et al. (2020, Section VII) and Borusyak et al. (2025, pp.199-201 and Online Appendix B.2), these findings suggest that treatment effect heterogeneity is unlikely to play an important role in this application. However, in every specification, the confidence intervals constructed from $\hat{\beta}_1$ (Wald CI) differ to some extent from those based on $\hat{\beta}_2$ (LM CI). This discrepancy is partly due to the different standard errors of the two estimators, which we exploit in our combination test to obtain strictly shorter confidence intervals across all specifications. For example, for workers with college equivalent skills, our confidence intervals are roughly 8% and 17% shorter than the Wald CIs in the specifications with and without controls, respectively.

Figure 2 displays the realized percentage reduction in confidence interval length achieved by our combination test, together with the asymptotic lower bounds for that reduction, analogous to Figure 1, but calculated using the specification-specific estimate $\hat{\rho}_1$ of the limiting correlation between the Wald and LM statistics. The theoretical results underlying the asymptotic lower bounds are presented in Section 4.3. Two observations are worth noting. First, allowing for finite-sample estimation error, the realized percentage reductions are generally close to or exceed their corresponding lower bounds across specifications, as predicted by our theory. Note also that these bounds are derived solely from the Wald and LM statistics, indicating that, in this specific empirical application, incorporating the AR statistic yields little additional efficiency gain. This aligns with our earlier theoretical discussion, which posits that efficiency gains increase monotonically with the absolute correlation between LM and AR statistics ρ_2, and, in fact, the corresponding consistent estimates $\hat{\rho}_2$ in Table 1 are not particularly large in almost every specification.

Second, Figure 2 clearly illustrates our proposed rule of thumb. All estimated $\hat{\rho}_1$ values fall within $[-0.7, 0.7]$. When the standard error ratio exceeds 1.05, the actual reduction in the length of the confidence interval is 17% (specification without controls for college equivalent workers), much above the rule-of-thumb benchmark 10% for its lower bound. In contrast, when the standard error ratio remains at or below 1.05, the improvements are modest, reflecting the converse of our rule of thumb. Nevertheless, even in these cases, our combination test can still deliver confidence intervals that are about 8% (specification with controls for college equivalent workers),

	College Equivalent Workers		High School Equivalent Workers	
	Yes	No	Yes	No
$\hat{\rho}_1$	0.602	0.447	0.409	0.574
$\hat{\rho}_2$	0.105	0.121	0.129	0.164
$\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$	0.953	1.105	0.694	0.720
$\hat{\beta}_1$	-0.078	-0.080	-0.037	-0.024
Wald CI	(-0.103, -0.053)	(-0.107, -0.052)	(-0.051, -0.023)	(-0.037, -0.011)
$\hat{\beta}_2$	-0.066	-0.058	-0.043	-0.030
LM CI	(-0.092, -0.040)	(-0.083, -0.033)	(-0.063, -0.024)	(-0.048, -0.012)
$\hat{\beta}^*$	-0.072	-0.064	-0.039	-0.025
Comb. CI	(-0.095, -0.049)	(-0.087, -0.041)	(-0.052, -0.026)	(-0.038, -0.013)

Table 1: Point estimates and confidence intervals: Immigrant enclave.

Notes: This table reports estimation and inference results for the immigrant enclave example using the Card (2009) dataset, separately for college and high school equivalent workers. Columns labeled "Yes" include city-level controls; those labeled "No" do not. Point estimates are obtained from the standard TSLS estimator with the one-dimensional Bartik instrument, $\hat{\beta}_1$, and from the leave-cluster-out estimator, $\hat{\beta}_2$, which uses the 38 base IVs. "Wald CI" and "LM CI" are confidence intervals based on $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. The estimator $\hat{\beta}^*$, defined in Section 4.3, is the combined estimator for β, essentially the midpoint of the confidence interval in (4.11), labeled "Comb. CI" in the table. $\hat{\rho}_1$ and $\hat{\rho}_2$ are estimates of the asymptotic correlation between the Wald and LM statistics, and between the LM and AR statistics, respectively. Finally, $\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$ is the ratio of the standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$. All numbers are rounded to three decimal places.

5% (specification without controls for high school equivalent workers), and 5% (specification with controls for high school equivalent workers) shorter. In Section 5, we further consider the return-to-education application of Angrist and Krueger (1991), in which our approach delivers even larger efficiency gains: our confidence intervals are $15\% - 23\%$ shorter than their conventional Wald counterparts based solely on the three QoB instruments.

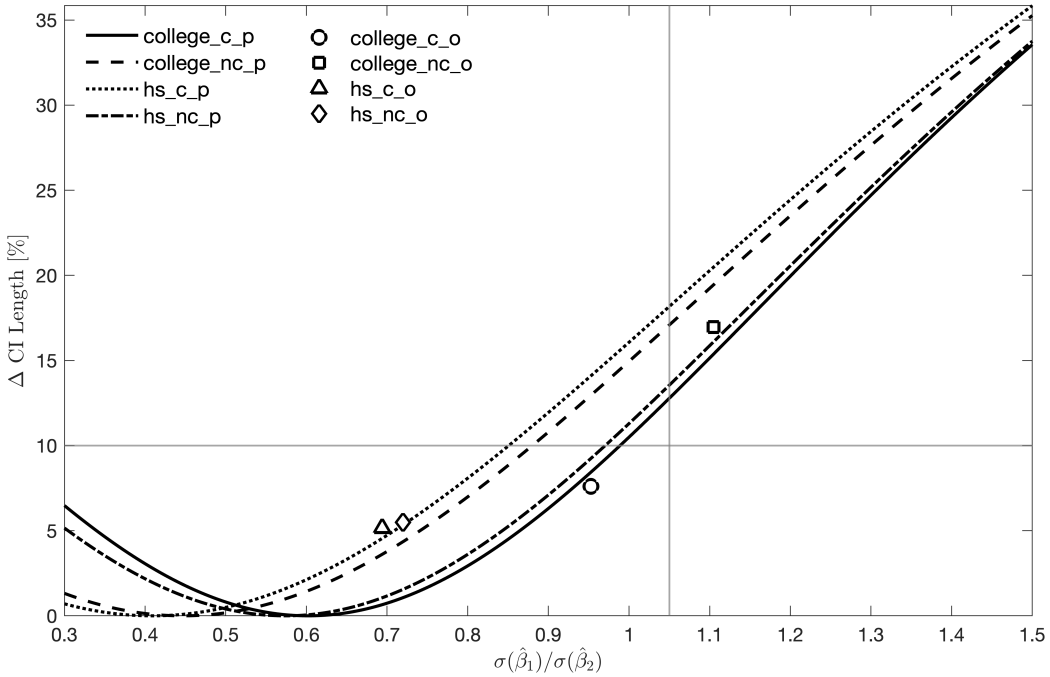


Figure 2: Realized percentage reduction in confidence interval length: Immigrant enclave.

Notes: This figure plots, for each specification in the immigrant enclave example, the observed percentage reduction in confidence interval length (Combined CI vs. Wald CI, as in Table 1, labeled "o" in the legends) against the standard error ratio $\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$ from Table 1. It also shows the theoretical lower bound on this reduction (labeled "p"), analogous to Figure 1, but computed using the specification-specific estimate $\hat{\rho}_1$ from Table 1. "college" refers to college equivalent workers and "hs" to high school equivalent workers; "c" indicates specifications with controls and "nc" those without. The figure is based on the unrounded numerical values underlying Table 1.

3 Model and Preliminaries

3.1 Setup

We consider a clustered dataset with G clusters and denote the size of the g-th cluster as n_g for $g \in [G]$. We index observations by units followed by clusters. Denote $I_g = [N_{g-1} + 1, \cdots, N_g]$, where $N_0 = 0$, $N_g = \sum_{g'=0}^{g} n_{g'}$, and $N_G = n$. Then, $\{I_g\}_{g\in[G]}$ forms a partition of $[n]$, and if $i \in I_g$, this means that the i-th observation belongs to

the g-th cluster. We then consider a linear IV regression with clustered data:

$$\tilde{Y}_{i,g} = \tilde{X}_{i,g}\beta + W_{i,g}^{\top}\gamma + \tilde{e}_{i,g}, \tag{3.1}$$

where we denote $\tilde{Y}_{i,g} \in \mathbb{R}$, $\tilde{X}_{i,g} \in \mathbb{R}$, and $W_{i,g} \in \mathbb{R}^{d_w}$ as an outcome variable, an endogenous regressor, and exogenous regressors, respectively. We assume that d_w is fixed with respect to the sample size n. Further denote $\tilde{Z}_{i,g} \in \mathbb{R}^K$ as the IVs for $\tilde{X}_{i,g}$, where K, the dimension of $\tilde{Z}_{i,g}$, diverges to infinity with the sample size.

In addition to the K-dimensional many instruments $\tilde{Z}_{i,g}$, we assume the availability of a second set of low-dimensional instruments $\tilde{z}_{i,g} \in \mathbb{R}^{d_z}$, where d_z is fixed with respect to the sample size n. Note that our framework accommodates, as a special case, settings in which $\tilde{z}_{i,g}$ form a subset of the K-dimensional instruments $\tilde{Z}_{i,g}$.

The first-stage equation can be written as

$$\tilde{X}_{i,g} = \tilde{\Pi}_{i,g} + \tilde{V}_{i,g}, \tag{3.2}$$

where $\tilde{\Pi}_{i,g} = \mathbb{E}(\tilde{X}_{i,g}|\{\tilde{Z}_{j,g}, \tilde{z}_{j,g}, W_{j,g}\}_{j\in I_g})$ is not assumed to be linear in $\tilde{Z}_{i,g}$, $\tilde{z}_{j,g}$ and $W_{i,g}$. We assume that $\left\{\tilde{e}_{i,g}, \tilde{V}_{i,g}\right\}_{i\in I_g, g\in[G]}$ are independent between clusters, but allow them to have a general dependence structure within each cluster.

Throughout the paper, we treat the many instruments $\tilde{Z}_{i,g}$, the controls $W_{i,g}$, and the low-dimensional instruments $\tilde{z}_{i,g}$ as deterministic, and assume

$$\mathbb{E}(\tilde{e}_{i,g}) = 0, \qquad \mathbb{E}(\tilde{V}_{i,g}) = 0.$$

This is equivalent to conducting all subsequent analyses conditionally on

$$(\tilde{Z}_{i,g}, W_{i,g}, \tilde{z}_{i,g})_{i\in I_g,\, g\in[G]},$$

while maintaining cross-cluster independence together with the exogeneity conditions

$$\mathbb{E}(\tilde{e}_{i,g} \mid \{\tilde{Z}_{i,g}, W_{i,g}, \tilde{z}_{i,g}\}_{i\in I_g,\, g\in[G]}) = 0, \qquad \mathbb{E}(\tilde{V}_{i,g} \mid \{\tilde{Z}_{i,g}, W_{i,g}, \tilde{z}_{i,g}\}_{i\in I_g,\, g\in[G]}) = 0.$$

If researchers wish to include cluster fixed effects, they can obtain (3.1) by first demeaning the outcome, endogenous regressor, controls, and instruments within each cluster. Such a practice is also recommended by Djogbenou, MacKinnon, and Nielsen (2019, Section 2.2) and MacKinnon, Nielsen, and Webb (2023, Section 3.2).

Let $\tilde{Y}$, $\tilde{X}$, $\tilde{\Pi}$, W, $\tilde{Z}$, $\tilde{z}$ be $n\times 1$, $n\times 1$, $n\times 1$, $n\times d_w$, $n\times K$, and $n\times d_z$-dimensional vectors and matrices formed by $\tilde{Y}_{i,g}$, $\tilde{X}_{i,g}$, $\tilde{\Pi}_{i,g}$, $W_{i,g}$, $\tilde{Z}_{i,g}$ and $\tilde{z}_{i,g}$, respectively. More specifically, $\tilde{Y}$ is constructed by stacking up $\tilde{Y}_{i,g}$ across $i \in I_g$ followed by $g \in [G]$, and similarly for $\tilde{X}$, $\tilde{\Pi}$, W, $\tilde{Z}$ and $\tilde{z}$. We then partial out W from $\tilde{Y}$, $\tilde{X}$, $\tilde{Z}$, and $\tilde{z}$ so that the model in (3.1) and (3.2) can be written in a vector form as

$$Y = X\beta + e, \quad X = \Pi + V, \tag{3.3}$$

where $Y = M_W\tilde{Y}$, $X = M_W\tilde{X}$, $\Pi = M_W\tilde{\Pi}$, $e = M_W\tilde{e}$, $V = M_W\tilde{V}$, $M_W = I_n - P_W$, $P_W = W(W^\top W)^{-1}W^\top$, and I_n denotes an $n\times n$ identity matrix. We further denote $Z = M_W\tilde{Z}$ and $z = M_W\tilde{z}$.

The null and alternative hypotheses studied are $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$. We focus on the model with a scalar endogenous variable for two reasons. First, in many empirical applications of IV regressions, there is only one endogenous variable.[9] Second, as we assume that at least the low-dimensional IVs provide strong identification, our results can be extended to testing of scalar restrictions with multiple endogenous variables by applying standard subvector inference methods, without appealing to a projection-based weak-identification-robust inference approach.[10]

Throughout the paper, we focus on the homogeneous treatment effect setting and

[9]For example, 101 out of 230 specifications in Andrews, Stock, and Sun (2019)'s sample and 1,087 out of 1,359 in Young (2022)'s sample feature one endogenous regressor and one IV. Similarly, Lee, McCrary, Moreira, and Porter (2022) find that 61 out of 123 IV papers published in *AER* between 2013 and 2019 use a single IV. Our setting further allows for the overidentified case with one endogenous regressor and multiple IVs. In addition, empirical researchers can often generate many IVs by using polynomials or interactions based on their low-dimensional base IVs and control variables, in the same spirit as Angrist and Krueger (1991). Then, it is possible to achieve efficiency improvement using our combination procedure.

[10]For weak-identification-robust subvector inference, in general, one may use a projection approach (Dufour and Taamouti, 2005) after implementing inference on the whole vector of endogenous variables. However, the projection approach typically leads to conservative inference. Alternative methods provide a power improvement under a fixed number of instruments (e.g., see Guggenberger, Kleibergen, Mavroeidis, and Chen (2012), Andrews (2017), Guggenberger, Kleibergen, and Mavroeidis (2019, 2021), and Wang and Doko Tchatoka (2018)). However, whether they can be applied to the setting of many weak instruments with clustered data is unclear. In addition, our method could potentially be extended to the case in which the identification strength provided by many IVs is mixed for the endogenous variables, by following the approach of Chao et al. (2012). We leave these extensions for future research.

study how a set of many instruments can improve inference based on a standard low-dimensional IV regression. In principle, when treatment effects are heterogeneous, a similar approach could be used to construct an optimal convex combination of the low- and high-dimensional IV estimators that minimizes asymptotic variance. Such an extension would require addressing several additional issues. In particular, under heterogeneous treatment effects, different sets of instruments generally identify different IV estimands. For each estimand to admit a causal interpretation, Kolesár (2013) and Blandhol, Bonney, Mogstad, and Torgovitsky (2026) show that the corresponding reduced-form equation must be (approximately) correctly specified and that a monotonicity condition must hold. Accordingly, the extension would require a framework in which these conditions are satisfied simultaneously for both the low- and high-dimensional IV specifications. We believe that developing such a framework is an interesting direction for future research, and leave it to subsequent work. Finally, we note that the analysis in Goldsmith-Pinkham et al. (2020, Section VII) and Borusyak et al. (2025, pp. 199-201 and Online Appendix B.2) indicates that treatment effect heterogeneity is unlikely to be a major concern in the immigrant enclave application of Card (2009) discussed in Section 2. The analysis in Mikusheva and Sun (2022, 2024) and in our Section 5.3 points to the same conclusion for the return-to-education application of Angrist and Krueger (1991).

3.2 Examples

Returns to Education in Angrist and Krueger (1991). As discussed in the Introduction, researchers often begin with a low-dimensional set of base instruments $\tilde{z}_{i,g}$, such as the three quarter-of-birth (QoB) dummies, and then generate a large number of additional 'technical' instruments by interacting $\tilde{z}_{i,g}$ with control variables, such as state- and year-of-birth dummies in Angrist and Krueger (1991). In Section 5.3, we follow Mikusheva and Sun (2022) to construct $\tilde{Z}_{i,g}$, and use these instruments to improve inference for the benchmark specification based solely on the three QoB instruments.

Immigrant Enclaves in Card (2009) and Goldsmith-Pinkham et al. (2020). A prominent example of the low-dimensional instrument $\tilde{z}_{i,g}$ is the widely used shift-share instrument. As noted by Goldsmith-Pinkham et al. (2020), under the identification strategy that treats the shares as exogenous, the one-dimensional shift-share

IV can be interpreted as a weighted average of many underlying base instruments.

For the immigrant enclave application of Card (2009) discussed in Section 2, the structural equation for each skill group (either high school or college equivalent) can be written as

$$\tilde{Y}_g = \tilde{X}_g\beta + W_g^\top\gamma + \tilde{e}_g,$$

where g indexes cities, and there is one observation per city, so $n_g = 1$ in this application. $\tilde{Y}_g$, $\tilde{X}_g$, and W_g denote the residual log wage gap between immigrant and native men, the log ratio of immigrant to native hours, and the controls, respectively. The shift-share instrument is given by

$$\tilde{z}_g = \sum_{k\in[K]} \tilde{s}_{1980,g,k}\, h_k,$$

where $\tilde{s}_{1980,g,k} = (N_{1980,g,k}/N_{1980,k}) \times (1/P_{2000,g})$, $N_{1980,g,k}$ denotes the number of immigrants from source country k residing in city g in 1980, $N_{1980,k}$ denotes the total number of immigrants from country k in the United States (US) in 1980, and $P_{2000,g}$ denotes the population of city g in 2000. Therefore, $\tilde{s}_{1980,g,k}$ is a population-normalized measure of city g's 1980 share of the nationwide immigrant population from country k. Additionally, h_k denotes the number of immigrants arriving in US from k between 1990 and 2000, and $K = 38$ is the number of source countries.

As pointed out by Goldsmith-Pinkham et al. (2020), when shares are exogenous, any individual share or linear combination of shares is itself a valid instrument. An alternative specification therefore replaces the shift-share instrument $\tilde{z}_g$ with K-dimensional underlying instruments $\tilde{Z}_g = \{\tilde{s}_{1980,g,k}\}_{k\in[K]}$. Under this interpretation, $\tilde{z}_g$ is simply a particular linear aggregation of $\tilde{Z}_g$, with $\{h_k\}_{k\in[K]}$ serving as the aggregation weights. As a robustness check, estimation results with alternative IV estimators based on $\tilde{Z}_g$ are reported alongside the benchmark specification based on $\tilde{z}_g$ (see Table 6 in their paper).

Inverse labor supply elasticity in Goldsmith-Pinkham et al. (2020). Consider the canonical application of estimating the inverse labor supply elasticity with panel data (see Sections I and VI of Goldsmith-Pinkham et al. (2020)). Observations are clustered at the location level (e.g., U.S. commuting zones) and observed over a

short panel of T periods. The structural equation is

$$\breve{Y}_{t,g} = \breve{X}_{t,g}\beta + \breve{\mathcal{W}}_{t,g}^{\top}\gamma + \breve{e}_{t,g}. \tag{3.4}$$

Here, g indexes locations and t indexes time. The variable $\breve{Y}_{t,g}$ denotes wage growth, $\breve{X}_{t,g}$ denotes employment growth, and $\breve{\mathcal{W}}_{t,g}$ collects controls. In our notation, G is the number of locations, $n_g = T$ for every $g \in [G]$, and $n = G \times T$.

The shift-share instrument is given by

$$\breve{z}_{t,g} = \sum_{k\in[\bar{K}]} \breve{s}_{0,g,k}\, h_{t,k},$$

where $\breve{s}_{0,g,k}$ denotes the employment share of industry k in location g during an initial period (e.g., 1980), $h_{t,k}$ is the national growth rate of industry k over period $t \in [T]$, and $\bar{K}$ is the number of industries. As suggested by Goldsmith-Pinkham et al. (2020), under the assumption that each initial industry share is exogenous, an alternative specification replaces $\breve{z}_{t,g}$ with a set of $K = \bar{K} \times T$-dimensional instruments

$$\breve{Z}_{t,g} = \{\breve{s}_{0,g,k} \times D_{t'}\}_{k\in[\bar{K}],\, t'\in[T]}, \tag{3.5}$$

where $D_{t'}$ denotes the time dummy for period t'. In particular, Goldsmith-Pinkham et al. (2020, Online Appendix D) shows that $\breve{z}_{t,g}$ can be obtained by aggregating $\breve{Z}_{t,g}$, using $\{h_{t',k}\}_{k\in[\bar{K}],\, t'\in[T]}$ as weights. We note that the time dummies in (3.5) are purely a technical device and do not themselves introduce additional identifying assumptions. Specifically, their role is to allow the same predetermined industry share $\breve{s}_{0,g,k}$ to receive different weights across time periods, thereby accommodating time variation in the national industry growth rates (i.e., the shifts).[11]

The empirical application reported in Section VI of their paper considers 228 industries and three time periods (1980-1990, 1990-2000, and 2000-2010). Note that (3.4) may include location (cluster) and time fixed effects.[12] Specifically, let $\breve{\mathcal{W}}_{t,g} =$

[11] Goldsmith-Pinkham et al. (2020, p.2594) uses an example with two industries and two time periods to illustrate the role of time dummies in the many-instrument representation.

[12] In this case, we can use $(\bar{K}-1)\times(T-1)$-dimensional interacted instruments to avoid perfect multicollinearity. With 228 industries and three periods in the current application, location fixed effects make one of the three time interactions redundant for each industry, while time fixed effects make one industry redundant in each of the two remaining periods.

$(\breve{W}_{t,g}^\top, D_g^\top)^\top$, where $\breve{W}_{t,g}$ collects the time dummies and other time-varying controls, and D_g denotes the location dummies, respectively. To fit the framework of Section 3.1, we first partial out the location fixed effects by demeaning the wage growth, the employment growth, the instruments, and the time-varying controls within each location. Denoting the demeaned variables by $\tilde{Y}_{t,g}$, $\tilde{X}_{t,g}$, $\tilde{Z}_{t,g}$, $\tilde{z}_{t,g}, W_{t,g}$, and $\tilde{e}_{t,g}$, the model can be written as

$$\tilde{Y}_{t,g} = \tilde{X}_{t,g}\beta + W_{t,g}^\top \gamma + \tilde{e}_{t,g},$$

which is of the form given in (3.1).

Finally, we emphasize that our objective is not to modify or reinterpret the identifying assumptions underlying IV applications, including shift-share IV designs. Rather, we take the identifying assumptions adopted in empirical applications as given. Conditional on these assumptions, whenever a low-dimensional IV specification and its many-IV counterpart identify the same parameter, our procedure can be employed to improve inference.

3.3 Test Statistics

Under our setting, it is possible to conduct inference directly based on the low-dimensional IVs. Specifically, given a $d_z \times d_z$ positive definite weighting matrix $\hat{A}_n$, the generalized method of moments (GMM) estimator can be written as

$$\hat{\beta}_1 = (X^\top z \hat{A}_n z^\top X)^{-1}(X^\top z \hat{A}_n z^\top Y). \tag{3.6}$$

It is also possible to construct test statistics using the K-dimensional many IVs. Denote $P = Z(Z^\top Z)^{-1}Z^\top$ as the projection matrix of Z. Then, the leave-cluster-out jackknife IV estimator of β is denoted as $\hat{\beta}_2$ and defined as

$$\begin{aligned}\hat{\beta}_2 &= \left(\sum_{g,h\in[G]^2, g\neq h} X_{[g]}^\top P_{[g,h]} X_{[h]}\right)^{-1} \left(\sum_{g,h\in[G]^2, g\neq h} X_{[g]}^\top P_{[g,h]} Y_{[h]}\right) \\ &= \left(X^\top (P-\bar{P})X\right)^{-1}\left(X^\top (P-\bar{P})Y\right),\end{aligned} \tag{3.7}$$

where $\bar{P}$ is the block diagonal matrix corresponding to P such that the g-th block on its diagonal is $P_{[g,g]}$. This estimator is the CJIVE estimator proposed by Frandsen et al. (2025), which extends the JIVE estimator for independent data studied by

Angrist, Imbens, and Krueger (1999), Chao et al. (2012), and Mikusheva and Sun (2022), among others. The key technical innovation of our paper is to use the CJIVE estimator to improve the conventional low-dimensional GMM IV estimator $\hat{\beta}_1$, while permitting weak identification of the many-IV specification. This requires a joint asymptotic analysis of $(\hat{\beta}_1, \hat{\beta}_2)$, in particular the consistent estimation of the variance-covariance matrix of the two estimators.

Given $\hat{\beta}_1$ and $\hat{\beta}_2$, we define the estimator $\hat{\beta}$ as

$$\hat{\beta} = \frac{\ddot{\Phi}_2^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}} \times \hat{\beta}_1 + \frac{\dot{\Phi}_1^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}} \times \hat{\beta}_2, \tag{3.8}$$

where $\dot{\Phi}_1$ and $\ddot{\Phi}_2$ are the variance estimators for $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, to be defined later in Section 4.1. We show that $\hat{\beta}$ is consistent whenever either the low-dimensional IV estimator $\hat{\beta}_1$ or the many-IV estimator $\hat{\beta}_2$ is consistent (i.e., either the low-dimensional IVs or the many IVs provide strong identification for β). The estimator $\hat{\beta}$ is therefore robust in the sense that its construction does not require researchers to determine ex ante which of the two preliminary estimators is consistent.

The construction of $\hat{\beta}$ is crucial for our analysis. Although our main analysis assumes that the low-dimensional instruments strongly identify the parameter, we also consider in Appendix A the opposite case, in which the low-dimensional instruments are weak and the many instruments are strong. This robust estimator accommodates both cases, providing consistent residuals without requiring researchers to determine in advance which specification provides consistent estimation.

We then use the robust estimator $\hat{\beta}$ to re-estimate the variances associated with $\hat{\beta}_1$ and $X^\top(P-\bar{P})e$, denoted by $\hat{\Phi}_1$ and $\hat{\Sigma}$, respectively, and defined in Section 4.1. These variance estimates are used to construct the Wald statistic (with low-dimensional IVs) and the leave-cluster-out jackknife LM statistic (with many IVs):

$$T(\beta_0) = \frac{(X^\top z \hat{A}_n z^\top X)^{-1} X^\top z \hat{A}_n z^\top e(\beta_0)}{\sqrt{\hat{\Phi}_1}} = \frac{\hat{\beta}_1 - \beta_0}{\sqrt{\hat{\Phi}_1}}, \tag{3.9}$$

$$LM(\beta_0) = \frac{X^\top (P - \bar{P}) e(\beta_0)}{\sqrt{\hat{\Sigma}}}, \tag{3.10}$$

where $e(\beta_0) = Y - X\beta_0$.

Lastly, as pointed out by Hausman et al. (2012), Lim et al. (2024a), and Mikusheva

and Sun (2024), it is possible to use the jackknife AR statistic to further improve the efficiency of the jackknife LM statistic. In the current setting with clustered data, we define the leave-cluster-out jackknife AR statistic as

$$AR = \frac{\hat{e}^\top (P - \bar{P})\hat{e}}{\sqrt{\hat{\Upsilon}}}, \tag{3.11}$$

where $\hat{\Upsilon}$ is a consistent variance estimator for the numerator defined later, and $\hat{e} = Y - X\hat{\beta}$.

We use the consistent estimator $\hat{\beta}$ to construct our tests for two reasons. First, using the null value β_0 to form the AR statistic and combining it with the LM statistic can produce a non-monotonic power curve, even under strong identification (i.e., a power ditch against certain alternatives; see Andrews (2016) and Lim et al. (2024a) for related discussions). Constructing the AR statistic with $\hat{\beta}$ avoids this issue. Consequently, the AR statistic in (3.11) does not depend on β_0 but serves as a normalized estimator of zero, used solely to improve the efficiency of our procedure. Second, the combination test relies on correlations between $T(\beta_0)$ and $LM(\beta_0)$ and between $LM(\beta_0)$ and AR, whose consistent estimation requires a consistent β estimator. This consistency is crucial to ensure proper size control, especially when the many-IV specification is weakly identified.

In the next section, we illustrate how to optimally combine $T(\beta_0)$, $LM(\beta_0)$, and AR. The corresponding variance estimators $(\dot{\Phi}_1, \ddot{\Phi}_2, \widehat{\Phi}_1, \hat{\Sigma}, \hat{\Upsilon})$ are introduced later in Section 4.1.

3.4 Combination Test

Given the three test statistics $(T(\beta_0), LM(\beta_0), AR)$, we seek to combine them in a theoretically justified way that can improve on the Wald test based only on the low-dimensional IVs. The key insight of our paper is that under certain local alternative $\beta - \beta_0 = \delta d_n$ with some deterministic sequence $d_n \downarrow 0$, we have the following joint limiting distribution:

$$\begin{pmatrix} T(\beta_0) \\ LM(\beta_0) \\ AR \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} a_1\delta \\ a_2\delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix} \right) \tag{3.12}$$

for some a_1, a_2, ρ_1 and ρ_2 to be defined later. In this limiting problem, the UMPU level-α test for the default null hypothesis $\delta = 0$ against two-sided alternatives, which are solely based on the limiting three-dimensional normal random vector, can be obtained by invoking standard hypothesis-testing results (see, for example, Section 4.2 of Lehmann and Romano (2006)), and is stated in Proposition 3.1 below.

Proposition 3.1. *Suppose that one observes $(\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3)$, which follows the limiting distribution in (3.12) with $\rho_1^2 + \rho_2^2 < 1$ and wants to test $\mathcal{H}_0 : \delta = 0$ against $\mathcal{H}_1 : \delta \neq 0$ for known values of $(a_1, a_2, \rho_1, \rho_2)$, then the UMPU level-$\alpha$ test rejects if*

$$\left(\frac{b_1 \tilde{\mathcal{N}}_1 + b_2 \tilde{\mathcal{N}}_2 + b_3 \tilde{\mathcal{N}}_3}{\sqrt{b_1^2 + b_2^2 + b_3^2}} \right)^2 \geqslant \mathbb{C}_\alpha,$$

where $\mathbb{C}_\alpha$ is the $(1-\alpha)$ percentile of a chi-squared random variable with one degree of freedom,

$$\begin{pmatrix} \tilde{\mathcal{N}}_1 \\ \tilde{\mathcal{N}}_2 \\ \tilde{\mathcal{N}}_3 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \\ \mathcal{N}_3 \end{pmatrix}, \text{ and } \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}.$$

The corresponding power function for the UMPU test is

$$\mathbb{P}\left(\chi_1^2 \left(\delta^2 \frac{(1-\rho_2^2)a_1^2 - 2\rho_1 a_1 a_2 + a_2^2}{1 - \rho_1^2 - \rho_2^2} \right) \geq \mathbb{C}_\alpha \right),$$

where $\chi_1^2(\lambda)$ is a noncentral chi-squared with noncentrality λ and one degree of freedom.

In light of this optimal testing result in the limiting problem, one may wish to propose implementing the following test:

$$\phi_n^o = \mathbf{1}\left\{ (\omega_1 T(\beta_0) + \omega_2 LM(\beta_0) + \omega_3 AR)^2 \geqslant \mathbb{C}_\alpha \right\} \equiv \phi^*\left(T(\beta_0), LM(\beta_0), AR \right), \quad (3.13)$$

$$\text{where} \quad \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \frac{1}{\sqrt{b_1^2 + b_2^2 + b_3^2}} \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad (3.14)$$

and then investigate its asymptotic justification.

However, the parameters a_1, a_2, ρ_1, ρ_2 are usually unknown and need to be estimated. In addition, it turns out that the weights $(\omega_1, \omega_2, \omega_3)$ are invariant to the scale normalization of (b_1, b_2, b_3), and thus, (a_1, a_2). Therefore, to construct the UMPU test, it suffices to consistently estimate $\alpha_1 = a_1/\sqrt{a_1^2 + a_2^2}$ and $\alpha_2 = a_2/\sqrt{a_1^2 + a_2^2}$ along with ρ_1 and ρ_2.

Given the consistent estimators $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\rho}_1, \hat{\rho}_2)$ for $(\alpha_1, \alpha_2, \rho_1, \rho_2)$ specified in Section 4.1, we then implement the feasible version of the combination test:

$$\phi_n^* = \mathbf{1}\left\{(\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR)^2 \geqslant \mathbb{C}_\alpha\right\}, \tag{3.15}$$

$$\text{where } \begin{pmatrix}\hat{\omega}_1\\ \hat{\omega}_2\\ \hat{\omega}_3\end{pmatrix} = \frac{1}{\sqrt{\hat{b}_1^2 + \hat{b}_2^2 + \hat{b}_3^2}} \times \begin{pmatrix}1 & \hat{\rho}_1 & 0\\ \hat{\rho}_1 & 1 & \hat{\rho}_2\\ 0 & \hat{\rho}_2 & 1\end{pmatrix}^{-1/2} \begin{pmatrix}\hat{b}_1\\ \hat{b}_2\\ \hat{b}_3\end{pmatrix}, \begin{pmatrix}\hat{b}_1\\ \hat{b}_2\\ \hat{b}_3\end{pmatrix} = \begin{pmatrix}1 & \hat{\rho}_1 & 0\\ \hat{\rho}_1 & 1 & \hat{\rho}_2\\ 0 & \hat{\rho}_2 & 1\end{pmatrix}^{-1/2} \begin{pmatrix}\hat{\alpha}_1\\ \hat{\alpha}_2\\ 0\end{pmatrix}. \tag{3.16}$$

4 Large-Sample Theory

In this section, we investigate the asymptotic behavior of our combination test. We begin by stating and discussing general assumptions about the data-generating process and the identification strength of both the low-dimensional and many IVs. We then establish the asymptotic efficiency properties of the combination test and, finally, compare its efficiency to that of the conventional Wald test based solely on low-dimensional IVs via the limiting length ratio of their confidence intervals.

4.1 General Assumptions

As in Chao et al. (2012), we treat $\tilde{Z}$ and W as fixed. This is equivalent to treating them as random and repeating all the analyses in the paper by conditioning on them. For the data-generating process, we impose the following assumptions.

Assumption 1. *The following conditions hold when n is sufficiently large:*

1. $\max_{i\in I_g, g\in[G]} \mathbb{E}(\tilde{e}_{i,g}^4 + \tilde{V}_{i,g}^4) \leqslant C < \infty$;
2. $\max_{g\in[G]} n_g \leqslant C < \infty$;

3. *Let*

$$\Omega_g = \mathbb{E}\left[\begin{pmatrix} \tilde{e}_{[g]}\tilde{e}_{[g]}^\top & \tilde{e}_{[g]}\tilde{V}_{[g]}^\top \\ \tilde{V}_{[g]}\tilde{e}_{[g]}^\top & \tilde{V}_{[g]}\tilde{V}_{[g]}^\top \end{pmatrix}\right], \quad g \in [G],$$

then

$$0 < \frac{1}{C} \leqslant \min_{g\in[G]} \lambda_{\min}(\Omega_g) \leqslant \max_{g\in[G]} \lambda_{\max}(\Omega_g) \leqslant C < \infty;$$

4.

$$0 < \frac{1}{C} \leqslant \lambda_{\min}\left(\frac{1}{n}\sum_{i\in I_g, g\in[G]} z_{i,g}z_{i,g}^\top\right) \leqslant \lambda_{\max}\left(\frac{1}{n}\sum_{i\in I_g, g\in[G]} z_{i,g}z_{i,g}^\top\right) \leqslant C < \infty,$$

$$0 < \frac{1}{C} \leqslant \lambda_{\min}\left(\frac{1}{n}\sum_{i\in I_g, g\in[G]} W_{i,g}W_{i,g}^\top\right) \leqslant \lambda_{\max}\left(\frac{1}{n}\sum_{i\in I_g, g\in[G]} W_{i,g}W_{i,g}^\top\right) \leqslant C < \infty,$$

and

$$\max_{i\in I_g, g\in[G]} \left\{\left|\tilde{\Pi}_{i,g}\right| + \left|\Pi_{i,g}\right|\right\} \leqslant C < \infty, \quad \max_{i\in I_g, g\in[G]} \left\{\left\|z_{i,g}\right\|_2 + \left\|W_{i,g}\right\|_2\right\} = o(\sqrt{n});$$

5. *There exists a sequence of non-random positive definite matrices A_n such that*

$$A_n^{-1/2}\hat{A}_n A_n^{-1/2} \xrightarrow{p} I_{d_z},$$

and $\lambda_{\max}(A_n)/\lambda_{\min}(A_n) \leqslant C < \infty$.

Remark 4.1. Assumption 1.1 is a standard condition on the moments of error terms. Assumption 1.2 restricts the cluster size to be bounded, which incorporates cross-sectional and short-panel data structures. It is also possible to extend our analysis to the case with divergent cluster sizes, especially when the within-cluster dependence is weak. However, when the cluster size is allowed to diverge and there is strong within-cluster dependence, the convergence rates of various IV estimators depend on the identification strength, the within-cluster dependence, the cluster sizes, and the number of clusters in a complicated way. We leave the investigation in this direction for future research. Assumption 1.3 ensures that the error covariance matrix is non-singular for each cluster. Assumption 1.4 is a mild condition for the design matrix. Assumption 1.5 states that the weighing matrix $\hat{A}_n$ converges in probability to some non-random positive definite matrix, which is standard in the GMM setup. We

note that for two-stage least squares (TSLS), Assumption 1.5 is actually implied by Assumption 1.4.

For the low-dimensional IVs, we focus on the case of strong identification strength in the main text. We further investigate the case in which the low-dimensional IVs have weak identification strength in Supplemental Appendix A, and show that, under certain mild conditions, when the many IVs provide strong identification, our optimal combination test still asymptotically controls size.

With strong identification, the asymptotic variance of $\hat{\beta}_1$ is

$$\Phi_1 = (\Pi^\top \acute{\Pi})^{-1} \Psi (\Pi^\top \acute{\Pi})^{-1},$$

where $\Omega = \mathbb{E} \sum_{g \in [G]} \left(z_{[g]}^\top \tilde{e}_{[g]} \right) \left(z_{[g]}^\top \tilde{e}_{[g]} \right)^\top$, $\acute{\Pi} = z A_n z^\top \Pi$, and $\Psi = \Pi^\top z A_n \Omega A_n z^\top \Pi$. The cluster-robust variance estimator $\widehat{\Phi}_1$ for the Wald statistic with low-dimensional IVs is

$$\widehat{\Phi}_1 = (X^\top \acute{X})^{-1} \hat{\Psi} (X^\top \acute{X})^{-1}, \tag{4.1}$$

where $\hat{\Omega} = \sum_{g \in [G]} \left(z_{[g]}^\top \hat{e}_{[g]} \right) \left(z_{[g]}^\top \hat{e}_{[g]} \right)^\top$, $\acute{X} = z \hat{A}_n z^\top X$, and $\hat{\Psi} = X^\top z \hat{A}_n \hat{\Omega} \hat{A}_n z^\top X$. The initial estimator $\dot{\Phi}_1$ for Φ_1 used in the computation of $\hat{\beta}$ in (3.8) is defined in the same way as $\widehat{\Phi}_1$, except that $\hat{e}_{[g]} = Y_{[g]} - X_{[g]} \hat{\beta}$ is replaced by $\dot{e}_{[g]} = Y_{[g]} - X_{[g]} \hat{\beta}_1$.

We make the following assumptions regarding the inference with low-dimensional IVs.

Assumption 2. *Let* $r_n = \left\| z^\top \Pi \right\|_2$, *then* $\sqrt{n}/r_n \to 0$.

Remark 4.2. Assumption 2 ensures that $z^\top \Pi$, the deterministic component of $z^\top X$, dominates $z^\top V$, its random component. This condition is therefore key for the consistency of $\hat{\beta}_1$ for β, that is, it ensures strong identification by the low-dimensional IVs. Throughout the paper, we focus on this strong identification case for the low-dimensional IVs while allowing the many-IV specification to be either weakly or strongly identified. In Section 4.2, we show that the optimal combination test adapts to the identification strength of the many IVs. In particular, it controls size asymptotically and retains nontrivial power even when $\hat{\beta}_2$ is inconsistent. In Section A of the Supplemental Appendix, we consider the opposite scenario in which the low-dimensional IV specification is weakly identified while the many-IV specification is

strongly identified. We provide mild conditions under which the exact inference procedure proposed in the paper still controls asymptotic size under the null.

If the many-IV-based identification is strong, similar to Chao et al. (2012), we can show that the asymptotic variance of $\hat{\beta}_2$ is

$$\Phi_2 = \left(\Pi^\top (P - \bar{P})\Pi\right)^{-1} \Sigma \left(\Pi^\top (P - \bar{P})\Pi\right)^{-1},$$

where

$$\Sigma = \mathbb{E}\left(\sum_{g,h\in[G]^2, g\neq h} \Pi_{[g]}^\top P_{[g,h]} \left(\sum_{k\in[G]} M_{W,[h,k]} \tilde{e}_{[k]}\right)\right)^2 + \mathbb{E}\left(\sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)^2$$

is the asymptotic variance of $X^\top (P - \bar{P})e$. A natural estimator for Φ_2 is thus

$$\hat{\Phi}_2 = \left(X^\top (P - \bar{P})X\right)^{-1} \hat{\Sigma} \left(X^\top (P - \bar{P})X\right)^{-1}, \tag{4.2}$$

where $\hat{\Sigma}$ is a consistent estimator of Σ. Such an estimator is proposed in Chao et al. (2012) for the case with independent data. Here, in addition to extending to clustered data, we need to account for the fact that W has already been partialled out, whereas in Chao et al. (2012) the coefficients for W are also estimated. Therefore, some adjustments are required, as in Matsushita and Otsu (2024). For that purpose, define $Q = M_W(P - \bar{P})M_W$, and let $\bar{Q}$ be the block diagonal matrix corresponding to Q such that the g-th block on its diagonal is $Q_{[g,g]}$. Our variance estimator is similar to the one in Chao et al. (2012) but with $P - \bar{P}$ replaced by $Q - \bar{Q}$, i.e.,

$$\hat{\Sigma} = \sum_{g\in[G]} \left(\sum_{h\in[G], h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \hat{e}_{[g]}\right)^2 + \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \hat{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \hat{e}_{[g]}\right). \tag{4.3}$$

The initial estimator $\ddot{\Phi}_2$ for Φ_2 used in the computation of $\hat{\beta}$ in (3.8) is defined in the same way as $\hat{\Phi}_2$, except that $\hat{e}_{[g]} = Y_{[g]} - X_{[g]}\hat{\beta}$ is replaced by $\ddot{e}_{[g]} = Y_{[g]} - X_{[g]}\hat{\beta}_2$.

Last, for the jackknife AR statistic, the variance estimator is given by

$$\hat{\Upsilon} = 2 \sum_{g,h\in[G]^2, g\neq h} \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right)^2, \tag{4.4}$$

which is consistent for the asymptotic variance of $\hat{e}^\top(P-\bar{P})\hat{e}$, given by

$$\Upsilon = \mathbb{E}\left(\sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)^2.$$

We make the following assumptions regarding the inference with many IVs.

Assumption 3. *1. $K \to \infty$ as $n \to \infty$ such that $\limsup_{n\to\infty} K/n \leqslant C < 1$;*

2. rank$(P) = K$ and $\max_{g\in[G]} \lambda_{\max}\left(P_{[g,g]}\right) \leqslant C < 1$;

3. Let $\hat{\Pi} = M_W(P-\bar{P})\Pi = Q\tilde{\Pi}$ and $\bar{\Pi} = (Q-\bar{Q})\tilde{\Pi}$, then

$$\max_{i\in I_g, g\in[G]} \left\{\left|\hat{\Pi}_{i,g}\right| + \left|\bar{\Pi}_{i,g}\right|\right\} \leqslant C < \infty,$$

and

$$\tilde{\Pi}^\top\tilde{\Pi} \leq C\Pi^\top\Pi, \;\; \hat{\Pi}^\top\hat{\Pi} \geq \Pi^\top\Pi/C, \;\; \Pi^\top(P-\bar{P})\Pi \geq \Pi^\top\Pi/C,$$

when n is large enough;

4. For all sufficiently large n,

$$\left|corr\left(\sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{V}_{[h]}, \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\right| \leqslant C < 1.$$

Remark 4.3. Assumption 3.1 allows the dimension of many IVs K to be proportional to the sample size n. Assumption 3.2 is similar to the standard condition that $\max_{1\leq i\leq n} P_{ii} \leqslant C < 1$ in the literature on many instruments with independent data, and the restriction that $\text{rank}(P) = K$ will exclude redundant columns from Z. Assumption 3.3 holds in general if $\tilde{\Pi}^\top\tilde{\Pi}$, $\Pi^\top\Pi$ and $\Pi^\top P\Pi$ are of the same order. Assumption 3.4 excludes perfect correlations between the two quadratic forms. Note that we do not impose any restriction on the identification strength of the many-IV specification, allowing $\Pi^\top\Pi/\sqrt{K}$ to be bounded, which corresponds to weak identification (Mikusheva and Sun, 2022).

4.2 Asymptotic Efficiency Properties of Combination Test

We now investigate the asymptotical properties of ϕ_n^* when the low-dimensional IVs are strong, in the sense that Assumption 2 holds. This allows us to define the local alternative according to the asymptotic variance of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the limiting covariance structure of the component statistics, from which the joint limiting distribution of $(T(\beta_0), LM(\beta_0), AR)$ can be derived. The results with weak low-dimensional IVs are given in Supplemental Appendix A. The formal regularity condition is stated as follows.

Assumption 4. *The following limits exist:*

$$\rho_1 = \lim_{n\to\infty} \frac{1}{\sqrt{\Psi\Sigma}} \sum_{g\in[G]} \mathbb{E}\left[\left(\acute{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right],$$
$$\rho_2 = \lim_{n\to\infty} \frac{2}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\right],$$

with $\rho_1^2 + \rho_2^2 < 1$.

Remark 4.4. Under many instruments, the asymptotic expansion of $LM(\beta_0)$ includes both linear and quadratic functions of the errors $(\tilde{e}, \tilde{V})$, whereas AR depends only on a quadratic function of $\tilde{e}$. The linear and quadratic components are asymptotically normal and uncorrelated, and therefore asymptotically independent. Since the Wald statistic $T(\beta_0)$ involves only linear functions of the errors, it is asymptotically uncorrelated with AR. Finally, ρ_1 and ρ_2 denote, respectively, the correlation between the linear components of $T(\beta_0)$ and $LM(\beta_0)$, and the correlation between the quadratic components of $LM(\beta_0)$ and AR.

The following theorem establishes the joint distribution of the three test statistics above under the local alternative.

Theorem 4.1. *Under Assumptions 1–4, suppose that there exists a deterministic sequence* $d_n \downarrow 0$ *such that* $d_n \Phi_1^{-1/2} \to a_1$, $d_n \Phi_2^{-1/2} \to a_2$, *and* $\beta - \beta_0 = \delta d_n$ *for some fixed* δ, *where* $a_1 \geq 0$, $a_2 \geq 0$ *and* $a_1^2 + a_2^2 > 0$. *Then we have the joint limiting distribution (3.12) for* $(T(\beta_0), LM(\beta_0), AR)^\top$.

Remark 4.5. The existence of the sequence d_n is ensured by Assumption 2. In particular, we may define $d_n = \min\left(\Phi_1^{1/2}, \Phi_2^{1/2}\right)$. Under the strong identification of

the low-dimensional IVs in Assumption 2, we have $\Phi_1^{1/2} = O\left(\frac{\sqrt{n}}{r_n}\right) = o(1)$, which immediately implies that $d_n = \min\left(\Phi_1^{1/2}, \Phi_2^{1/2}\right) = o(1)$, regardless of the order of Φ_2.

Remark 4.6. The joint normality established in Theorem 4.1 holds even when the many-IV specification is weakly identified in the sense of Mikusheva and Sun (2022), that is, when $\Pi^\top \Pi / \sqrt{K}$ is bounded. This result follows from two observations. First, the estimator $\hat{\beta}$ used to construct the variance estimators $\hat{\Phi}_2$ and $\hat{\Upsilon}$ for the LM and AR statistics remains consistent due to the strong identification of the low-dimensional IVs and its robustness property by the design in (3.8).

Second, under weak many-IV identification, the quadratic components of the LM and AR statistics dominate their asymptotic behavior and yield asymptotic normality as long as $K \to \infty$. In this regime, we have $\rho_1 = 0$ (as the Wald statistic is asymptotically uncorrelated with the quadratic components), $d_n = \sqrt{n}/r_n$, and $\Phi_2^{-1} = O(1)$, which further imply that $a_2 = 0$ under Assumption 2.

To implement the optimal test ϕ_n^* defined in (3.15), we still need to estimate α_1, α_2, ρ_1, and ρ_2. For α_1 and α_2, we propose the following estimators:

$$\hat{\alpha}_1 = \frac{\sqrt{\hat{\Phi}_2}}{\sqrt{\hat{\Phi}_1 + \hat{\Phi}_2}}, \quad \text{and} \quad \hat{\alpha}_2 = \frac{\sqrt{\hat{\Phi}_1}}{\sqrt{\hat{\Phi}_1 + \hat{\Phi}_2}}. \tag{4.5}$$

In addition, let $\hat{X} = M_W(P - \bar{P})X$. For ρ_1 and ρ_2, we propose the following estimators:

$$\hat{\rho}_1 = \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}} \sum_{g \in [G]} \left[\left(\hat{X}_{[g]}^\top \hat{e}_{[g]} \right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]} \right) \right], \quad \text{and} \tag{4.6}$$

$$\hat{\rho}_2 = \frac{2}{\sqrt{\hat{\Sigma}\hat{\Upsilon}}} \sum_{g,h \in [G]^2, g \neq h} \left[\left(X_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \right]. \tag{4.7}$$

By combining Proposition 3.1 and Theorem 4.1, and invoking the approach developed by Müller (2011), we obtain a precise sense of asymptotic optimality for our proposed test ϕ_n^*, which is formalized in the following Theorem 4.2.

Theorem 4.2. *Let $\mathcal{M}$ denote the set of data generating processes m that satisfy the conditions of Theorem 4.1 pointwise for all $\delta \in \mathbb{R}$. Suppose that one wants to test*

$\mathcal{H}_0 : \delta = 0$ *against* $\mathcal{H}_1 : \delta \neq 0$. *Then, for the class* $\mathfrak{C}$ *of tests* ϕ_n *satisfying that*

$$\lim_{n\to\infty} \mathbb{E}[\phi_n] \leqslant \alpha \quad \text{for all } m \in \mathcal{M}, \delta = 0, \tag{4.8}$$

$$\liminf_{n\to\infty} \mathbb{E}[\phi_n] \geqslant \alpha \quad \text{for all } m \in \mathcal{M}, \delta \neq 0, \tag{4.9}$$

we have $\phi_n^* \in \mathfrak{C}$, *and, for any* $\delta_1 \neq 0$ *and any* $\phi_n \in \mathfrak{C}$,

$$\lim_{n\to\infty} \mathbb{E}[\phi_n] \leqslant \lim_{n\to\infty} \mathbb{E}[\phi_n^*] \quad \text{for all } m \in \mathcal{M}, \delta = \delta_1. \tag{4.10}$$

Moreover, for the test $\tilde{\phi}_n = \mathbf{1}\{T^2(\beta_0) \geqslant \mathbb{C}_\alpha\}$, *we have* $\tilde{\phi}_n \in \mathfrak{C}$, *and for any* δ *and all* $m \in \mathcal{M}$,

$$\lim_{n\to\infty} \mathbb{E}\left[\tilde{\phi}_n\right] = \lim_{n\to\infty} \mathbb{E}[\phi_n^*] \quad \text{if and only if} \quad a_2 = \rho_1 a_1.$$

Remark 4.7. Theorem 4.2 shows that, under local alternatives, ϕ_n^* attains the asymptotic efficiency bound within the class of tests that remain asymptotically unbiased and valid for all data generating processes inducing the same weak limit for $(T(\beta_0), LM(\beta_0), AR)^\top$ as in Theorem 4.1. This class of tests includes, in particular, the Wald and jackknife LM tests based solely on the low-dimensional and many IVs, respectively. It also includes the HLIM and HFUL-based tests of Hausman et al. (2012), both of which are asymptotically equivalent to linear combinations of the LM and AR tests under many IVs.[13] Furthermore, as we will discuss more in detail in Section 4.3, the weights in (3.14) also yield an optimally efficient combined estimator of $(\hat{\beta}_1, \hat{\beta}_2, AR)$ that achieves the minimal asymptotic variance.

However, our optimal test ϕ_n^* may not dominate tests that cannot be expressed directly as functions of $(T(\beta_0), LM(\beta_0), AR)$, such as the sup-score test of Belloni et al. (2012) and the ridge-regularized AR test of Dovì, Kock, and Mavroeidis (2024). Indeed, it is possible to construct data generating processes under which either the optimal combination test, the sup-score test, or the ridge-regularized AR test achieves the highest power. We establish the notion of optimality in Theorem 4.2 primarily to provide guidance for constructing tests that improve upon the conventional Wald test, rather than to identify a globally optimal procedure. That said, one could potentially

[13] E.g., see the discussions below Theorem 4.2 in Lim et al. (2024a).

combine our ϕ_n^* test with alternative tests such as the sup-score test, in the spirit of Navjeevan (2024), to obtain more powerful inference.

Remark 4.8. Theorem 4.2 also clarifies the necessary and sufficient condition under which the combination test does not deliver a strict power gain over the Wald test $T(\beta_0)$, namely $a_2 = \rho_1 a_1$. Recall that a_1 and a_2 represent the orders of the concentration parameters for the low-dimensional and many IVs, respectively. As shown below, even when we allow either of them to be zero—thereby covering situations where one IV estimator dominates the other in terms of convergence rate—the condition $(a_2 = \rho_1 a_1)$ is still seldom met. Put differently, one should generally anticipate strictly more powerful inference when using our test.

In particular, when $a_1 = 0$ and $a_2 > 0$, corresponding to the case when the identification strength of many IVs is larger than that of the low-dimensional IVs (i.e., the convergence rate of $\hat{\beta}_2$ is faster than that of $\hat{\beta}_1$), we obtain a strict power improvement for all values of ρ_1 and ρ_2 satisfying $\rho_1^2 + \rho_2^2 < 1$. Conversely, when $a_1 > 0$ and $a_2 = 0$, meaning that low-dimensional IVs provide stronger identification than many IVs, we still achieve a strict power gain as long as $\rho_1 \neq 0$. When $a_1 > 0$ and $a_2 > 0$, that is, when the two sets of IVs have identification strengths of the same order, strict power improvement is ensured, provided that $\rho_1 \neq a_2/a_1$. Indeed, at $\rho_1 = a_2/a_1$, the sufficient statistic for δ derived from the joint limiting distribution of the three component statistics (cf. (3.12)) becomes independent of the limiting Gaussian observations associated with the LM and AR statistics, so it is not surprising that combining the Wald statistic with them does not yield a more powerful inference.

Remark 4.9. As long as the low-dimensional IVs provide strong identification, the optimal combination test ϕ_n^* does not lose asymptotic power for any degree of identification strength of the many IVs. In this sense, the efficiency gains delivered by the combination test are essentially a "free lunch". Under local alternatives, the weak convergence result in Theorem 4.1 holds uniformly, regardless of whether the many IVs are strong or weak. In particular, when the many IVs are weak so that $a_2 = \rho_1 = 0$ (as the quadratic term in $LM(\beta_0)$ dominates the linear term), the second part of Theorem 4.2 shows that the combination test asymptotically reduces to the Wald test, implying no efficiency loss from combining. Moreover, as shown below, the combination test remains consistent against fixed alternatives irrespective of the identification strength of the many IVs.

Finally, for any fixed alternative, both $T(\beta_0)$ and $LM(\beta_0)$ are consistent and, by construction, avoid the issue of non-monotonic power (noted at the end of Section 3.3) when their corresponding set of IVs is strong. Hence, it is reasonable to anticipate that our combined test will retain these desirable properties, a result that we formalize in the theorem below. However, we emphasize once more that these results remain valid regardless of the strength of the many IVs.

Theorem 4.3. *Suppose that Assumptions 1-4 hold. Then, under $\beta - \beta_0 = \delta$ for some fixed $\delta \neq 0$, we have $\lim_{n\to\infty} \mathbb{E}[\phi_n^*] = 1$.*

4.3 Quantifying Efficiency Improvement in Large Samples

We measure the efficiency improvement of the combination test over the conventional Wald test that uses only low-dimensional IVs by the percentage reduction in the asymptotic length of the resulting confidence interval. Recall from Remark 4.9 that, if the many-IV specification is weakly identified, the combination test is asymptotically equivalent to the Wald test. Consequently, the associated confidence interval takes the usual "estimator plus and minus a standard error times a critical value" form: $\left[\hat{\beta}_1 - \sqrt{\hat{\Phi}_1} \times \sqrt{\mathbb{C}_\alpha},\ \hat{\beta}_1 + \sqrt{\hat{\Phi}_1} \times \sqrt{\mathbb{C}_\alpha}\right]$, where $\hat{\beta}_1$ and $\hat{\Phi}_1$ are as defined above, and $\sqrt{\mathbb{C}_\alpha}$ is the standard normal critical value. In this case, the combination test remains valid and has the same efficiency as the conventional Wald test.

When the many IVs provide strong identification, the confidence interval associated with our combination test can also be expressed in the familiar "estimator plus and minus a standard error times a critical value" form. To see this, observe that under strong identification by many IVs, the component LM statistic can be represented as follows:

$$LM(\beta_0) = \frac{X^\top (P - \bar{P}) e(\beta_0)}{\sqrt{\hat{\Sigma}}} = \frac{\hat{\beta}_2 - \beta_0}{\sqrt{\hat{\Phi}_2}} + o_P(1),$$

where the $o_P(1)$ term comes from the fact that under strong identification,

$$\text{sign}(X^\top (P - \bar{P}) X) \xrightarrow{p} 1.$$

Inserting this into (3.15) gives the following form of our combination test:

$$\phi_n^* = \mathbf{1}\left\{\left(\hat{\omega}_1\frac{\hat{\beta}_1-\beta_0}{\sqrt{\widehat{\Phi}_1}}+\hat{\omega}_2\frac{\hat{\beta}_2-\beta_0}{\sqrt{\widehat{\Phi}_2}}+\hat{\omega}_2 o_P(1)+\hat{\omega}_3 AR\right)^2 \geqslant \mathbb{C}_\alpha\right\}.$$

The resulting confidence interval is asymptotically equivalent to

$$CI^* = \left[\hat{\beta}^* - \frac{1}{\left(\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}+\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}\right)}\sqrt{\mathbb{C}_\alpha},\ \hat{\beta}^* + \frac{1}{\left(\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}+\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}\right)}\sqrt{\mathbb{C}_\alpha}\right], \tag{4.11}$$

where $\hat{\beta}^*$ is a combined estimator of β,

$$\hat{\beta}^* = \frac{\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}}{\left(\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}+\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}\right)}\hat{\beta}_1 + \frac{\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}}{\left(\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}+\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}\right)}\hat{\beta}_2 + \frac{\hat{\omega}_3 AR}{\left(\hat{\omega}_1/\sqrt{\widehat{\Phi}_1}+\hat{\omega}_2/\sqrt{\widehat{\Phi}_2}\right)}.$$

The intuition behind the combined estimator is fundamentally efficiency-driven. First, $\widehat{\Phi}_1$ and $\widehat{\Phi}_2$ estimate the asymptotic variances of $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, so for given weights $(\hat{\omega}_1, \hat{\omega}_2)$ the combined estimator $\hat{\beta}^*$ assigns greater weight to the estimator with the smaller asymptotic variance. Second, although the AR statistic is asymptotically centered at zero and therefore does not affect the location of the combined estimator, its correlation with the many-IV-based estimator $\hat{\beta}_2$ allows it to reduce the variance of the combined estimator. This mechanism parallels the construction of the HLIML and HFUL estimators in Hausman et al. (2012), which can be more efficient than the JIVE estimator. Third, the estimated weights $(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)$ are consistent for the population weights $(\omega_1, \omega_2, \omega_3)$ in (3.14) associated with the UMPU test. Finally, the optimal weights defined in (3.14) solve the following problem:

$$\min_{\omega_1,\omega_2,\omega_3} \frac{1}{(a_1\omega_1+a_2\omega_2)^2} \quad \text{s.t.} \quad (\omega_1,\omega_2,\omega_3)\begin{pmatrix}1 & \rho_1 & 0\\ \rho_1 & 1 & \rho_2\\ 0 & \rho_2 & 1\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\\ \omega_3\end{pmatrix} = 1.$$

The quadratic constraint ensures that CI^* attains the correct asymptotic coverage, while the objective corresponds to the asymptotic variance of the combined estimator

$\hat{\beta}^*$, since

$$\frac{1}{d_n^2\left(\omega_1/\sqrt{\Phi_1}+\omega_2/\sqrt{\Phi_2}\right)^2} \to \frac{1}{(a_1\omega_1+a_2\omega_2)^2}.$$

Thus, the optimal weights in (3.14), originally motivated by the UMPU testing problem, also yield an optimally efficient combined estimator of $(\hat{\beta}_1, \hat{\beta}_2, AR)$ that achieves the minimal asymptotic variance.

A direct implication of the confidence interval in (4.11) is that, asymptotically, the percentage reduction in its length relative to the confidence interval based on the conventional Wald test can be derived analytically as follows,

$$\begin{aligned}
&\operatorname*{plim}_{n\to\infty} 1 - \frac{1/\left(\hat{\omega}_1/\sqrt{\hat{\Phi}_1}+\hat{\omega}_2/\sqrt{\hat{\Phi}_2}\right)}{\sqrt{\hat{\Phi}_1}} \\
=&1-\sqrt{\frac{(1-\rho_1^2-\rho_2^2)}{(1-\rho_2^2)-2\rho_1 a_2/a_1+(a_2/a_1)^2}} \qquad (4.12)\\
\geqslant&1-\sqrt{\frac{(1-\rho_1^2)a_1^2}{a_1^2-2\rho_1 a_1 a_2+a_2^2}} = 1-\sqrt{\frac{(1-\rho_1^2)}{(1-\rho_1^2)+(\rho_1-a_2/a_1)^2}}. \qquad (4.13)
\end{aligned}$$

As equation (4.12) shows, the efficiency gain is primarily driven by the relative identification strength of the low-dimensional and many IVs, summarized by

$$\frac{a_2}{a_1} = \lim_{n\to\infty}\sqrt{\frac{\Phi_1}{\Phi_2}},$$

as well as by the limiting correlations between the Wald and LM statistics, denoted by ρ_1, and between the LM and AR statistics, denoted by ρ_2. In particular, consistent with Theorem 4.2, when $a_2/a_1 = \rho_1$, the combination test ϕ_n^* yields no efficiency improvement, and its confidence interval is asymptotically of the same length as that based on $\tilde{\phi}_n$. This case includes the weak many-IV scenario, where $\rho_1 = a_2 = 0$.

By contrast, whenever $a_2/a_1 \neq \rho_1$, the resulting confidence interval is strictly shorter, implying improved efficiency. Moreover, the efficiency gain in (4.12) is monotonically increasing in $|\rho_2|$, which leads to the lower bound reported in (4.13) by setting $\rho_2 = 0$. This lower bound can also be interpreted as the CI length reduction

achieved by optimally combine $\hat{\beta}_1$ and $\hat{\beta}_2$ only.[14] Figure 1 plots this bound as a function of a_2/a_1, the ratio of the standard deviations of $\hat{\beta}_1$ and $\hat{\beta}_2$, for various values of ρ_1.

In practice, a_2/a_1 can be estimated by the ratio of any consistent estimator of Φ_1 and Φ_2 under strong identification.[15] This yields a simple rule of thumb, discussed in Section 2: for empirically plausible values of ρ_1 between -0.7 and 0.7, if the ratio of the reported standard errors $\sqrt{\dot{\Phi}_1}$ and $\sqrt{\ddot{\Phi}_2}$ exceeds 1.05, then the associated confidence interval shortens by at least 10%.

In some applications (e.g., the replication of Card (2009) by Goldsmith-Pinkham et al. (2020, Table 6)), researchers report alternative many-IV estimators, such as HFUL, instead of the JIVE estimator ($\hat{\beta}_2$) considered above. Let $\tilde{\beta}_2$ and $\tilde{\Phi}_2$ denote the alternative estimator and its variance, respectively, and let $\tilde{\rho}_1$ denote the correlation between the GMM estimator $\hat{\beta}_1$ and $\tilde{\beta}_2$. By the same argument, if $\tilde{\rho}_1 \in [-0.7, 0.7]$ (or $\tilde{\rho}_1 \in [-0.99, 0.99]$), the confidence interval $\widetilde{CI}$, obtained by optimally combining $\hat{\beta}_1$ and $\tilde{\beta}_2$ using our procedure, is at least 10% shorter than the conventional Wald interval whenever the ratio $\sqrt{\Phi_1/\tilde{\Phi}_2}$ exceeds 1.05 (or 1.1). Moreover, if $\tilde{\beta}_2$ is asymptotically equivalent to a combination of the JIVE estimator $\hat{\beta}_2$ and the AR statistic, as is the case for HLIML and HFUL, then Theorem 4.2 implies that our optimal confidence interval (CI^*) is weakly shorter than $\widetilde{CI}$ and therefore at least 10% shorter than the conventional Wald interval. Hence, the rule of thumb is not specific to the JIVE estimator $\hat{\beta}_2$, but applies more broadly to many-IV estimators asymptotically

[14] The optimal reduction in this case is $\lim_{n\to\infty} 1 - \frac{1/\left(\tilde{\omega}_1/\sqrt{\Phi_1}+\tilde{\omega}_2/\sqrt{\Phi_2}\right)}{\sqrt{\Phi_1}} = 1 - \frac{1}{(\tilde{\omega}_1+\tilde{\omega}_2 a_2/a_1)}$, where the optimal weights are computed by

$$(\tilde{\omega}_1, \tilde{\omega}_2) = \underset{\omega_1,\omega_2}{\arg\min} \frac{1}{(a_1\omega_1 + a_2\omega_2)^2} \quad \text{s.t.} \quad (\omega_1, \omega_2)\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = 1.$$

By direct calculation, we can show that

$$1 - \frac{1}{(\tilde{\omega}_1 + \tilde{\omega}_2 a_2/a_1)} = 1 - \sqrt{\frac{(1-\rho_1^2)}{(1-\rho_1^2) + (\rho_1 - a_2/a_1)^2}}.$$

[15] If the many-IV specification is weakly identified, then $\Phi_1 = o(1)$ and $1/\Phi_2 = O(1)$, so that $a_2/a_1 = 0$. Suppose we estimate Φ_1 and Φ_2 by $\dot{\Phi}_1$ and $\ddot{\Phi}_2$, respectively. Although $\ddot{\Phi}_2$ is not consistent under weak many IVs, the proof of Lemma A.8 in the Supplemental Appendix reveals that $\ddot{\Phi}_2^{-1/2} = O_P(1)$ and $\dot{\Phi}_1^{1/2} = o_P(1)$, implying $\sqrt{\dot{\Phi}_1/\ddot{\Phi}_2} \xrightarrow{p} 0 = a_2/a_1$. Hence, the plug-in estimator of a_2/a_1 remains consistent even under weak identification of the many-IV specification.

equivalent to a combination of JIVE and the AR statistic.

As a final remark, note from (3.9) that the preceding theory is developed under the premise that the GMM estimator $\hat{\beta}_1$ and the component Wald statistic $T(\beta_0)$ share the same weighting matrix $\hat{A}_n$. This, in turn, implies that the lower bound in (4.13) is even more conservative when the benchmarked Wald confidence interval is based on the TSLS estimator, while the combination test is constructed using the optimal weighting matrix. Hence, the proposed rule of thumb continues to be valid.

5 Practical Implementation

In this section, we first synthesize the preceding discussions to describe how to practically implement our combination test in an empirically relevant model. We then apply it in simulations to assess its finite-sample power performance, and also in the return-to-education application of Angrist and Krueger (1991) as a further empirical illustration.

We consider the following model with clustered data,

$$\breve{Y}_{i,g} = \breve{X}_{i,g}\beta + \breve{W}_{i,g}^{\top}\gamma + \alpha_g + \breve{e}_{i,g}, \tag{5.1}$$

$$\breve{X}_{i,g} = \breve{Z}_{i,g}^{\top}\pi + \breve{W}_{i,g}^{\top}\tau + \xi_g + \breve{V}_{i,g}, \tag{5.2}$$

where α_g and ξ_g denote cluster-specific fixed effects, and $\breve{Y}_{i,g}$, $\breve{X}_{i,g}$, $\breve{W}_{i,g}$, and $\breve{Z}_{i,g}$ represent, respectively, the outcome variable, the endogenous regressor, exogenous regressors, and (potentially many) base instruments. By demeaning at the cluster level, we partial out the fixed effects and obtain $\tilde{Y}_{i,g}$, $\tilde{X}_{i,g}$, $W_{i,g}$, $\tilde{Z}_{i,g}$, $\tilde{V}_{i,g}$, and $\tilde{e}_{i,g}$, echoing the notation in our initial model setup in (3.1) and (3.2). The setting we consider in this paper is one in which practitioners are often interested in testing $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$ using an alternative set of low-dimensional IVs $\tilde{z}_{i,g}$. For example, these instruments can be a subset of the many IVs $\tilde{Z}_{i,g}$, or a weighted or simply an unweighted average of $\tilde{Z}_{i,g}$.

5.1 Practical Guideline

Our implementation procedure starts by partialling out the covariates W from $\tilde{Y}$, $\tilde{X}$, $\tilde{Z}$, and $\tilde{z}$. This produces Y, X, Z, and z, which are consistent with the notation

used in Section 3.1. These transformed variables serve as the effective observations in all subsequent estimation and inference procedures. Note that, although the above theory is derived for a generic weighting matrix $\hat{A}_n$ in the preliminary GMM estimator $\hat{\beta}_1$ in (3.6), in practice one may prefer to employ an efficient weighting matrix. We account for this by employing a two-step GMM procedure for the low-dimensional IV estimator throughout all implementations in this paper. Specifically, in the first step we use the identity matrix as the weighting matrix to obtain consistent estimates of residuals, and in the second step we utilize the resulting efficient weighting matrix estimated from the first step. The complete procedure is described below.

1. *Preliminary consistent estimates of residuals e in (3.3).*

   (a) Use Y, X, the low-dimensional IVs z, and the TSLS weighting matrix $\hat{A}_n^{(0)} = (z^\top z)^{-1}$ to obtain a preliminary GMM estimate $\hat{\beta}_1^{(0)}$ by (3.6).

   (b) Use Y, X, the many IVs Z, and the projection matrix of Z to obtain a preliminary leave-cluster-out jackknife estimate $\hat{\beta}_2$ by (3.7).

   (c) Using the residuals from the structural equation in (3.3), computed with $\hat{\beta}_1^{(0)}$, obtain a first-step preliminary estimate $\dot{\Phi}_1^{(0)}$ (cf. (4.1)) for the asymptotic variance Φ_1 of $\hat{\beta}_1$.

   (d) Using the residuals from the structural equation in (3.3), computed with $\hat{\beta}_2$, obtain a preliminary estimate $\ddot{\Phi}_2$ (cf. (4.2)) for the asymptotic variance Φ_2 of $\hat{\beta}_2$. Pay special attention to the adjustments made in (4.3) to account for the fact that W has already been partialled out.

   (e) Use $\hat{\beta}_1^{(0)}$, $\hat{\beta}_2$, $\dot{\Phi}_1^{(0)}$, and $\ddot{\Phi}_2$ to obtain a first-step robust estimate $\hat{\beta}^{(0)}$ of β by (3.8).

   (f) Form the residuals $\hat{e}_{[g]}^{(0)} = Y_{[g]} - X_{[g]}\hat{\beta}^{(0)}$ for each g-th cluster.

2. *Estimate the optimal weighting matrix.*

   (a) Compute the optimal weighting matrix given by $\hat{A}_n^* = \left(\hat{\Omega}^{(0)}\right)^{-1}$, where

$$\hat{\Omega}^{(0)} = \sum_{g\in[G]} \left(z_{[g]}^\top \hat{e}_{[g]}^{(0)}\right)\left(z_{[g]}^\top \hat{e}_{[g]}^{(0)}\right)^\top,$$

   and fix this weighting matrix throughout the remaining steps.

3. *A robust estimate of β and the resulting robust variance estimates.* (See Section 4.1 for explicit formulas of $\hat{\Psi}$, $\hat{\Omega}$, $\widehat{\Phi}_1$, $\hat{\Sigma}$, $\widehat{\Phi}_2$ and $\hat{\Upsilon}$.)

   (a) Recompute the low-dimensional efficient GMM estimate $\hat{\beta}_1$ from (3.6) using the optimal weighting matrix $\hat{A}_n^*$.

   (b) Using the residuals from the structural equation in (3.3), computed with $\hat{\beta}_1$, obtain the second-step preliminary estimate $\dot{\Phi}_1$ for the asymptotic variance Φ_1 of $\hat{\beta}_1$.

   (c) Use $\hat{\beta}_1$ and $\dot{\Phi}_1$, together with $\hat{\beta}_2$ (obtained in Step 1(b)) and $\ddot{\Phi}_2$ (obtained in Step 1(d)) to obtain a robust estimate $\hat{\beta}$ of β by (3.8).

   (d) Using the residuals from (3.3), computed with $\hat{\beta}$, obtain the robust variance estimates $\hat{\Psi}$, $\hat{\Omega}$ and thus $\widehat{\Phi}_1$ (cf. (4.1)), $\hat{\Sigma}$ (cf. (4.3)) and thus $\widehat{\Phi}_2$ (cf. (4.2)), and finally $\hat{\Upsilon}$ (cf. (4.4)).

4. *Component statistics and their estimated weights.*

   (a) Using the hypothesized β_0 and the outputs from Step 3 above, compute the three component statistics—$T(\beta_0)$ (cf. (3.9)), $LM(\beta_0)$ (cf. (3.10)), and AR (cf. (3.11)).

   (b) Using the outputs from Step 3 above, compute $\hat{\alpha}_1$ and $\hat{\alpha}_2$ (cf. (4.5)), $\hat{\rho}_1$ and $\hat{\rho}_2$ (cf. (4.6) and (4.7)), and then the weights $(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)$ (cf. (3.16)).

5. Using outputs from Step 4, obtain the feasible combination test ϕ_n^* by (3.15).

5.2 Simulation Study

All of our simulations are based on (5.1) and (5.2). The fixed effects are generated by $\alpha_g = u_{1g} + g/G$ and $\xi_g = u_{2g} + g/G$, $g = 1, \ldots, G$, where u_{1g} and u_{2g} are independent standard normal random variables. The control variables in $\breve{W}_{i,g}$ are generated by the standard normal distribution, and the dimension of $\breve{W}$ is fixed at $d_w = 10$. The instruments in $\breve{Z}_{i,g}$ are normally distributed with mean 0 and cluster-level dependence:

within each cluster g, the covariance matrix is given by

$$\Omega_{1g} = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_1 \\ \theta_1 & 1 & \cdots & \theta_1 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1 & \theta_1 & \cdots & 1 \end{bmatrix}_{n_g \times n_g}, \quad g = 1, \ldots, G,$$

and between clusters these instruments are independent of each other; we set $\theta_1 = 0.5$ in our simulations. Finally, to obtain an arguably complex error structure, we first generate $\acute{e}_{i,g} = \rho\varepsilon_{i,g} + \sqrt{1-\rho^2}\sigma_{i,g}v_g$ and $\acute{V}_{i,g} = \rho\eta_{i,g} + \sqrt{1-\rho^2}\sigma_{i,g}v_g$, where $\sigma_{i,g} = \sqrt{\left(0.2 + (\breve{W}_{i,g}^{\top}\tau)^2\right)/2.4}$. Here, $\varepsilon_{i,g}$, $\eta_{i,g}$, and v_g are mutually independent standard normal random variables, ρ governs the degree of endogeneity, τ is specified below, and we fix $\rho = 0.5$ in all simulations. Next, within each cluster, we premultiply the vectors $\acute{e}_{i,g}$ and $\acute{V}_{i,g}$ by

$$\Omega_{2g} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \theta_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_2^{n_g-1} & \theta_2^{n_g-2} & \cdots & 1 \end{bmatrix}_{n_g \times n_g}, \quad g = 1, \ldots, G,$$

thereby generating $\breve{e}_{i,g}$ and $\breve{V}_{i,g}$. In our simulations, we set $\theta_2 = 0.7$.

For the parameters, we set $\beta = 0.3$, $\gamma = \tau = (1/\sqrt{d_w}) \times \iota_{d_w}$, where ι_{d_w} is a $d_w \times 1$ vector of ones. We specify geometrically decaying coefficients of IVs as $\pi = \left(\phi^0, \phi^1, \ldots, \phi^{K-1}\right)$, where K denotes the number of many IVs and ϕ controls the relative weight assigned to each instrument. It is noted that $\phi = 0$ represents the case in which only the first instrument has identification strength, while $\phi = 1$ corresponds to the case in which each instrument has the same identification strength. We further normalize π to have $\|\pi\|_2 = \sqrt{\psi\sqrt{K}/n}$, where ψ controls the identification strength of the many IVs. The one-dimensional IV is constructed by taking the simple average of the many IVs;[16] as ϕ approaches one, the identification strength of the low-dimensional IV becomes stronger since it is closer to the optimal instrument. We set the sample size at $n = 2{,}000$ and the number of clusters at

[16]The normalizing factor $1/K$ is innocent here, as it will be canceled out in the numerator and denominator of $\hat{\beta}_1$.

$G = 500$, and then generate heterogeneous cluster sizes following a procedure similar to that in Djogbenou et al. (2019). Specifically, for $g = 1, \cdots, G-1$, we set $n_g = \max\left\{1, n\exp(2g/G)/(1+\sum_{g=1}^{G-1}\exp(2g/G))\right\}$ and then the size of the last cluster as $n_G = \max\left\{1, n - \sum_{g=1}^{G-1} n_g\right\}$. For the dimension of $\breve{Z}$, we consider $K = 100$ and $K = 500$, respectively. All the results below are based on $5,000$ simulations.

Figure 3 displays the power curves for our combination test ϕ_n^* along with those for the component Wald and jackknife LM tests, at different values of K (the dimension of the many IVs), ψ (which governs the identification strength of the many IVs), and ϕ (which controls the identification strength of the one-dimensional IV relative to the many IVs). We identified three main observations, each of which aligns with our large-sample theory. First, in every scenario, the combination test ϕ_n^* attains the correct size and is more powerful than each of the other two tests. In particular, as shown in Panel C, the power curve ϕ_n^* is never dominated by that of the Wald test, regardless of the strength of the many IVs, thereby underscoring the "free lunch" efficiency gains delivered by our combination test. Second, within Panels A and B of Figure 3, we observe that for given values of K and ψ, the power improvement of ϕ_n^* over the Wald test becomes more substantial as the identification strength of the one-dimensional IV weakens relative to that of the many IVs (i.e., as ϕ decreases). This is reflected in the widening gap between the power curves of ϕ_n^* and the Wald test. It emphasizes how many-IV-based LM and AR statistics contribute to the power enhancement. Third, in contrast to the cases in which the one-dimensional IV dominates many IVs in strength (first figure in Panel A or B) and the power curve of ϕ_n^* coincides with that of the Wald test, noticeable gaps persist between the power curves of ϕ_n^* and the LM test in the flipped cases (second and third figures in Panel A or B). These gaps highlight how the AR component contributes to power enhancement through its correlation with the LM statistic.

5.3 Empirical Application

We consider the return-to-education application, using the dataset of Angrist and Krueger (1991). In this application, the outcome variable is the log weekly wages and the endogenous variable is the years of schooling. We follow Mikusheva and Sun (2022) and Lim et al. (2024a) to consider two specifications with 180 and $1,530$ instruments. The set of 180 instruments consists of 30 quarter and year of birth

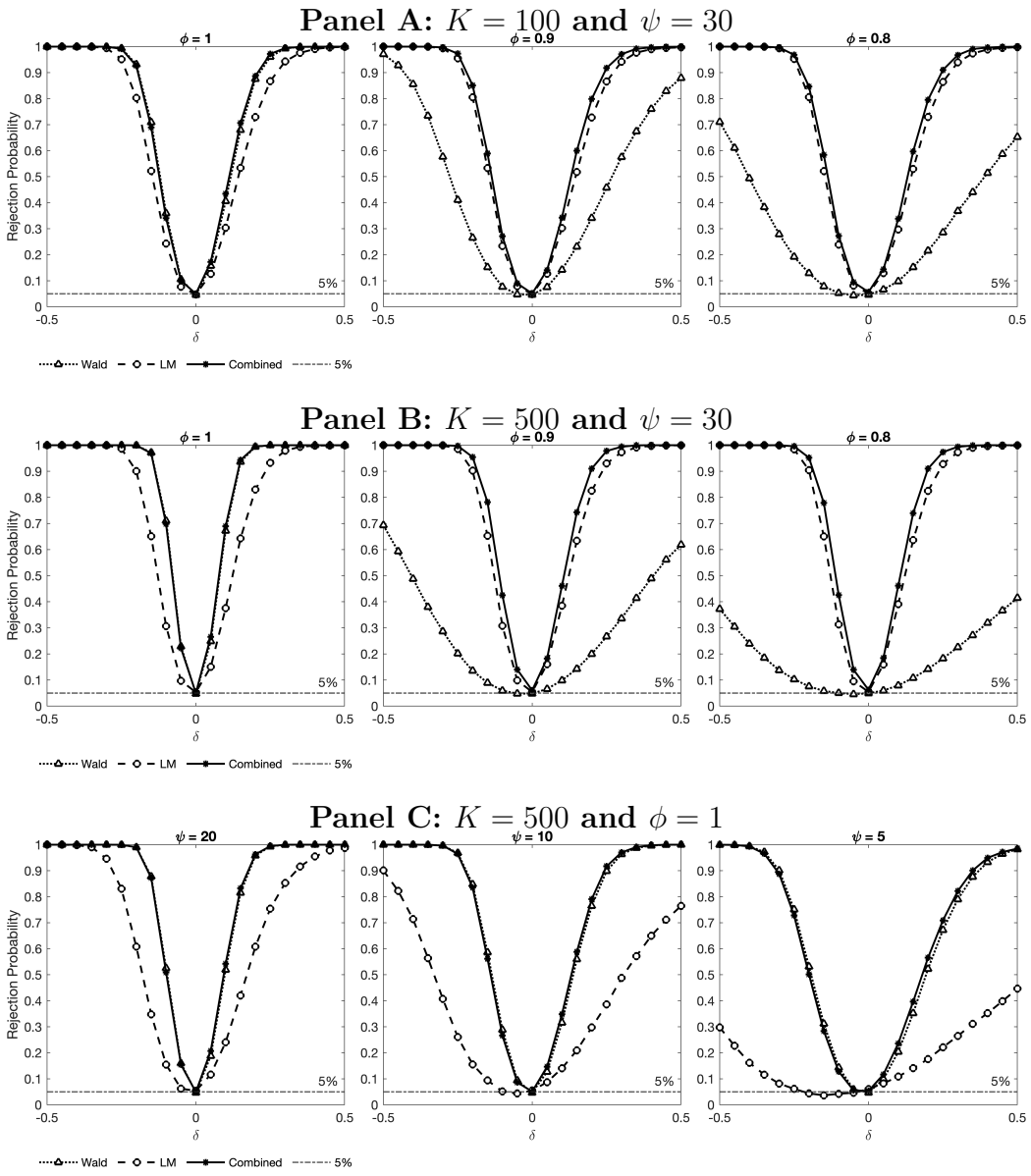


Figure 3: Power curves of the combination, Wald, and LM tests.

Notes: This figure displays the power curves for our combination test ϕ_n^* along with those for the component Wald and LM tests, at different values of K (the dimension of the many IVs), ψ (which governs the identification strength of the many IVs), and ϕ (which controls the identification strength of the one-dimensional IV relative to the many IVs). The horizontal axis represents the deviations in the parameter of interest from the maintained hypothesis, that is, we test $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$, where $\delta = \beta - \beta_0$. See Section 5 for details of the simulation setup. All results are based on $5,000$ simulations.

interactions (QoB–YoB) and 150 quarter and place of birth interactions (QoB-PoB). The set of $1,530$ instruments includes all interactions among QoB-YoB-PoB. The quantitative implications obtained from Table 2 and Figure 4 are in line with the discussion in Section 2 of the main text. Our confidence intervals are roughly 23% and 15% shorter, respectively, than the Wald CIs with the low-dimensional IVs in the specifications with $K = 180$ and $K = 1,530$. Here, we note that the actual reduction in the length of the confidence interval can be greater than 10% even when the standard error ratio is much lower than 1.05 (in the specification with $K = 1,530$), again indicating the conservativeness of our rule of thumb. However, we reiterate that, although the low-dimensional IVs (the three QoB IVs, and therefore $\hat{\beta}_1$ and the Wald confidence interval) are identical across the two specifications ($K = 180$ and $1,530$) by construction, we are not attempting to obtain improved inference based on pooling statistics across specifications, and our theoretical results do not justify such an approach.

6 Conclusion

This paper proposes an inference approach that improves conventional estimation and inference in instrumental variables regressions using low-dimensional instruments (e.g., aggregated shift-share IVs) and their underlying high-dimensional instruments. The procedure requires strong identification only for the low-dimensional IV regression while allowing the many-IV specification to be weakly identified. We also provide a practical rule of thumb for when inference improves by at least 10%, based solely on the standard error ratio of the low- and high-dimensional IV estimators commonly reported in empirical applications. Extensions to settings with many controls or diverging cluster sizes, and combinations with tests such as the sup-score test are left for future work.

	$K = 180$	$K = 1,530$
$\hat{\rho}_1$	0.491	0.236
$\hat{\rho}_2$	-0.171	-0.200
$\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$	1.195	0.817
$\hat{\beta}_1$	0.098	0.098
Wald CI	(0.059, 0.138)	(0.059, 0.138)
$\hat{\beta}_2$	0.099	0.084
LM CI	(0.066, 0.132)	(0.036, 0.133)
$\hat{\beta}^*$	0.097	0.093
Comb. CI	(0.066, 0.127)	(0.059, 0.126)

Table 2: Point estimates and confidence intervals: Return to education.

Notes: This table reports the estimation and inference results for the return-to-education example using the Angrist and Krueger (1991) dataset, shown separately for specifications with $K = 180$ instruments and $K = 1,530$ instruments. The IV set in the column labeled "$K = 180$" consists of 30 quarter and year of birth interactions (QoB–YoB) and 150 quarter and place of birth interactions (QoB–PoB), while the IV set in the column with "$K = 1,530$" includes the full set of interactions among QoB–YoB–PoB. See Appendix D in Lim et al. (2024a) for a more detailed description of data and this empirical application. The point estimates are obtained from the efficient GMM estimator with the three-dimensional QoB instruments, $\hat{\beta}_1$, and, in addition, from the leave-cluster-out estimator, $\hat{\beta}_2$, which makes use of many IVs. Wald CI and LM CI denote the confidence intervals based on $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. The estimator $\hat{\beta}^*$ is the combined estimator for β, defined in Section 4.3 of the main text. It is essentially the midpoint of the confidence interval in (4.11), which is obtained from our combination test and labeled as "Comb. CI" in the table. In addition, $\hat{\rho}_1$ and $\hat{\rho}_2$ denote estimates of the asymptotic correlation between the Wald and LM statistics, and between the LM and AR statistics, respectively. Finally, $\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$ denotes the ratio of standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$. All displayed numbers are rounded to three decimal places.

Appendix A Combination Test with Weak Low-Dimensional IVs

In this section, we study the property of our combination test when Assumption 2 is violated. Specifically, we assume that the identification strength provided by the low-dimensional IVs is weak, rendering the Wald test invalid. On the other hand, we assume that the identification strength provided by the many IVs is strong. To characterize the limiting behavior of the low-dimensional IVs under weak identification, we introduce the following assumptions.

Assumption 2′. *The following conditions hold:*

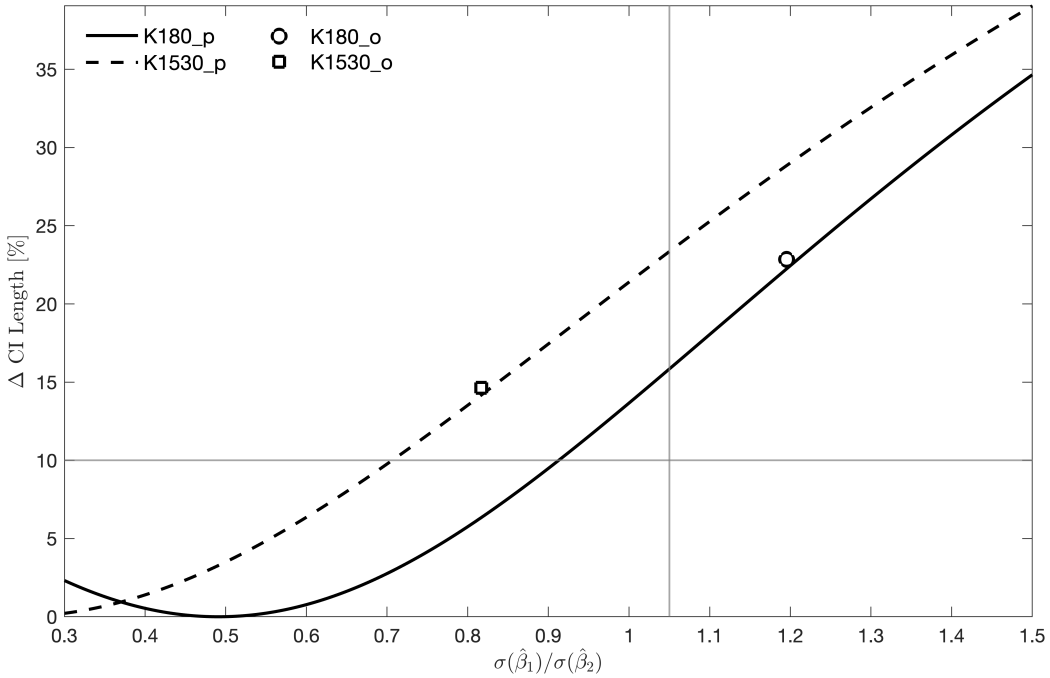


Figure 4: Realized percentage reduction in confidence interval length: Return to education

Notes: This figure shows, for each specification in the return-to-education example, the observed percentage reduction in confidence interval length (Combined CI versus Wald CI, as in Table 2, and indicated by "o" in figure legends) plotted as a point against the standard error ratio ($\hat{\sigma}(\hat{\beta}_1)/\hat{\sigma}(\hat{\beta}_2)$ in Table 2). Also shown is the theoretical lower bound for the reduction (indicated by "p" in figure legends), analogous to Figure 1 in the main text, but now computed using the specification-specific estimate $\hat{\rho}_1$, as reported in Table 2. Here, "K180" refers to the specifications with $K = 180$ instruments, and "K1530" refers to the specifications with $K = 1,530$ instruments. The horizontal axis is the ratio of standard deviations (errors) of $\hat{\beta}_1$ and $\hat{\beta}_2$. The vertical axis is the reduction in the length of confidence interval in percentage points. As a final remark, note that the actual numerical values of the relevant quantities in Table 2, rather than the rounded values shown there, are used to produce Figure 4.

1. *For* $r_n = \left\| z^\top \Pi \right\|_2$, $r_n/\sqrt{n} = O(1)$;

2. *For all* n *large enough,*

$$\frac{1}{n}\begin{pmatrix} \sum_{g\in[G]} \mathbb{E}\left(z_{[g]}^\top \tilde{e}_{[g]}\right)\left(z_{[g]}^\top \tilde{e}_{[g]}\right)^\top & \sum_{g\in[G]} \mathbb{E}\left(z_{[g]}^\top \tilde{e}_{[g]}\right)\left(z_{[g]}^\top \tilde{V}_{[g]}\right)^\top \\ \sum_{g\in[G]} \mathbb{E}\left(z_{[g]}^\top \tilde{V}_{[g]}\right)\left(z_{[g]}^\top \tilde{e}_{[g]}\right)^\top & \sum_{g\in[G]} \mathbb{E}\left(z_{[g]}^\top \tilde{V}_{[g]}\right)\left(z_{[g]}^\top \tilde{V}_{[g]}\right)^\top \end{pmatrix} > 0,$$

in the matrix sense.

Remark A.1. Compared to Assumption 2, the main difference here is that we have $r_n/\sqrt{n} = O(1)$. This corresponds to weak identification of the parameter of interest β under the low-dimensional IVs, since in this case the deterministic part and the random part of $z^\top X$ are of the same order. This setting is similar to the weak-IV asymptotics considered in Staiger and Stock (1997), where $z^\top X/\sqrt{n}$ converges to a random limit instead of diverging to infinity (the latter would happen under a standard asymptotics where $z^\top X/n$ is assumed to converge to a non-zero fixed limit). In addition, perfect correlations between the two linear forms are excluded.

In the following, we first present the results for the case with $d_z = 1$. Note that it is the most important case for empirical applications of IV regressions. For instance, 101 out of 230 specifications in Andrews et al. (2019)'s sample and 1,087 out of 1,359 in Young (2022)'s sample feature one endogenous regressor and one IV. Similarly, Lee et al. (2022) find that 61 out of 123 IV papers published in *AER* between 2013 and 2019 use one endogenous regressor and one IV. For these applications, empirical researchers can generate many IVs by using polynomials or interactions based on their one-dimensional base IV and control variables. Then, it is possible to achieve efficiency improvement using our combination procedure. Furthermore, the widely used shift-share or Bartik instrument is also one-dimensional. It turns out that in the case with $d_z = 1$, the local asymptotic power function of our combination test is equal to that of the asymptotically optimal test based on $z^\top e(\beta_0)/\sqrt{\Omega}$, $LM(\beta_0)$, and AR, where the first statistic corresponds to the conventional cluster-robust AR test using z as an instrument.

Theorem A.1. *Suppose that $d_z = 1$. Assume that the following limits exist:*

$$\bar{\rho}_1 = \lim_{n\to\infty} \frac{1}{\sqrt{\Omega\Sigma}} \sum_{g\in[G]} \mathbb{E}\left[\left(z_{[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right],$$
$$\bar{\rho}_2 = \lim_{n\to\infty} \frac{2}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\right],$$

and $\bar{\rho}_1^2 + \bar{\rho}_2^2 < 1$. Under Assumptions 1, 2′ and 3, and assuming that $\Pi^\top\Pi/\sqrt{K} \to \infty$, we have:

1. *Suppose that there exists a deterministic sequence $d_n \downarrow 0$ such that*

$$d_n \Phi_2^{-1/2} \to a > 0, \quad \text{and} \quad \beta - \beta_0 = \delta d_n,$$

for some fixed δ, then

$$\lim_{n\to\infty}\mathbb{E}\left[\phi_n^*\right]=\mathbb{P}\left(\left(-\frac{\bar{\rho}_1}{\sqrt{1-\bar{\rho}_1^2-\bar{\rho}_2^2}}\mathcal{N}_1+\frac{1}{\sqrt{1-\bar{\rho}_1^2-\bar{\rho}_2^2}}\mathcal{N}_2-\frac{\bar{\rho}_2}{\sqrt{1-\bar{\rho}_1^2-\bar{\rho}_2^2}}\mathcal{N}_3\right)^2\geqslant\mathbb{C}_\alpha\right)$$
$$=\mathbb{P}\left(\chi_1^2\left(\delta^2\frac{a^2}{1-\bar{\rho}_1^2-\bar{\rho}_2^2}\right)\geq\mathbb{C}_\alpha\right),$$

where $\begin{pmatrix}\mathcal{N}_1\\ \mathcal{N}_2\\ \mathcal{N}_3\end{pmatrix}\overset{d}{=}\mathcal{N}\left(\begin{pmatrix}0\\ a\delta\\ 0\end{pmatrix},\begin{pmatrix}1&\bar{\rho}_1&0\\ \bar{\rho}_1&1&\bar{\rho}_2\\ 0&\bar{\rho}_2&1\end{pmatrix}\right).$

2. *Suppose that $\beta-\beta_0=\delta$ for some fixed $\delta\neq 0$, then $\lim_{n\to\infty}\mathbb{E}\left[\phi_n^*\right]=1$.*

Theorem A.1 implies that, in the scalar case ($d_z=1$), our approach is equivalent to an optimal combination of the low-dimensional AR, jackknife LM, and jackknife AR. A key scalar-specific feature is that the discrepancy between the low-dimensional AR and low-dimensional Wald (i.e., $T(\beta_0)$) effectively reduces to a random sign that cannot be consistently estimated when the IV z is weak; correspondingly, $\hat{\rho}_1$ is inconsistent, and the low-dimensional Wald is non-normal. However, by appropriately choosing combination weights, this unidentifiable sign component cancels out, preserving the validity of our combination test in the important scalar setting.

When $d_z>1$, it is no longer possible to cancel the random sign. Nevertheless, if the correlation between $T(\beta_0)$ and $LM(\beta_0)$ is asymptotically negligible, our combined test remains robust to weak low-dimensional IVs, as shown in the next result.

Theorem A.2. *Assume that the following limit exists:*

$$\rho=\lim_{n\to\infty}\frac{2}{\sqrt{\Sigma\Upsilon}}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\right],$$

with $\rho^2<1$. Under Assumptions 1, 2′ and 3, and assuming that $\Pi^\top\Pi/\sqrt{K}\to\infty$ with $\Pi^\top\Pi/K\to 0$, we have:

1. *Suppose that there exists a deterministic sequence $d_n\downarrow 0$ such that*

$$d_n\Phi_2^{-1/2}\to a>0,\quad \textit{and}\quad \beta-\beta_0=\delta d_n,$$

for some fixed δ, then

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_n^*\right] = \mathbb{P}\left(\left(\frac{1}{\sqrt{1-\rho^2}}\mathcal{N}_1 - \frac{\rho}{\sqrt{1-\rho^2}}\mathcal{N}_2\right)^2 \geqslant \mathbb{C}_\alpha\right) = \mathbb{P}\left(\chi_1^2\left(\delta^2\frac{a^2}{1-\rho^2}\right) \geq \mathbb{C}_\alpha\right),$$

where $\begin{pmatrix}\mathcal{N}_1\\ \mathcal{N}_2\end{pmatrix} \overset{d}{=} \mathcal{N}\left(\begin{pmatrix}a\delta\\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$.

2. *Suppose that $\beta - \beta_0 = \delta$ for some fixed $\delta \neq 0$, then* $\lim_{n\to\infty} \mathbb{E}\left[\phi_n^*\right] = 1$.

Theorem A.2 shows that under the local alternative $\beta - \beta_0 = \delta d_n$, $\lim_{n\to\infty} \mathbb{E}\left[\phi_n^*\right]$ coincides exactly with the local asymptotic power function of the asymptotically optimal test based on $LM(\beta_0)$ and AR in Lim et al. (2024a). The assumption that $\Pi^\top\Pi/K \to 0$ is similar to that imposed in Mikusheva and Sun (2022, Theorem 4). As pointed out by Mikusheva and Sun (2022), this condition is quite weak as it still covers both weakly and strongly identified cases with many IVs. In addition, we note that under fixed alternatives, the combination test remains consistent even when the identification strength of the low-dimensional IVs is weak.

Note also that instead of using the low-dimensional Wald statistic, it is possible to combine the low-dimensional AR, jackknife LM, and jackknife AR directly, provided that the many IVs are strong. In this way, the combination test is robust against weak low-dimensional IVs when $d_z > 1$ without additional assumptions (e.g., $\Pi^\top\Pi/K \to 0$). In this paper, however, we emphasize Wald-based inference for the low-dimensional IV setting because it is the workhorse in empirical applications, and when low-dimensional IVs are strong, Wald is more powerful than AR in the overidentified case.

Finally, if both low-dimensional IVs and many IVs may be weak, then using null-imposed variance estimators for all component statistics, the combination test that combines the low-dimensional AR, jackknife LM, and jackknife AR is robust to arbitrary weak identification. It is also possible to consider a broader combination that includes low-dimensional AR and Kleibergen (2002, 2005)'s LM tests, along with jackknife LM and AR tests. This is beyond the scope of the current paper.

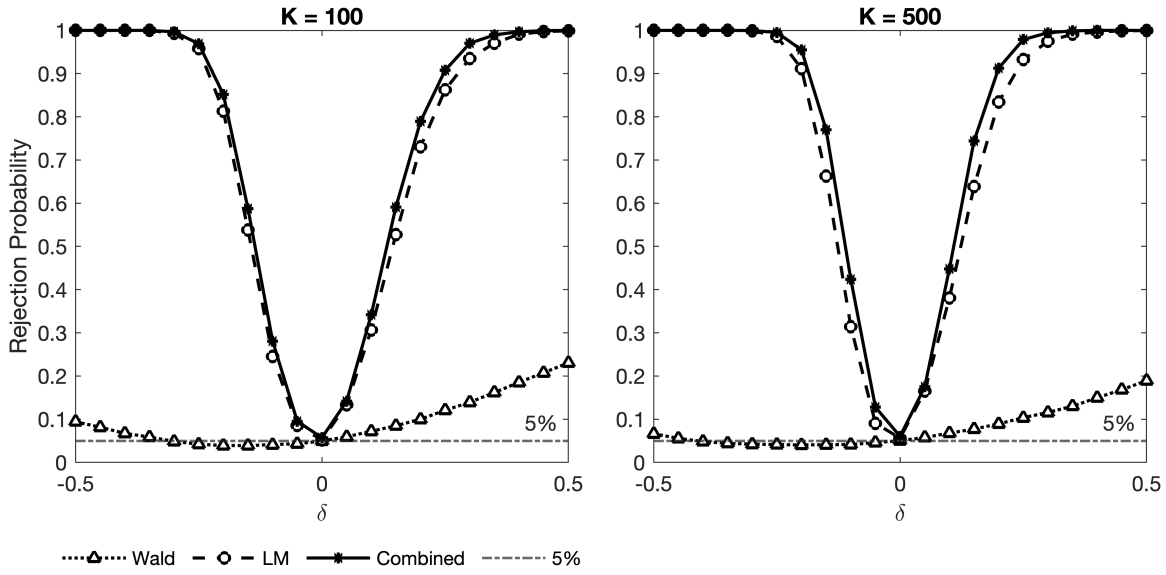


Figure 5: Power Curve of the combination, Wald, and LM tests.

Notes: This figure displays the power curves for our combination test ϕ_n^* along with those for the component Wald and LM tests, at different values of K (the dimension of the many IVs), $\psi = 30$ (the identification strength of the many IVs is relatively strong), and $\phi = 0$ (the identification strength of the one-dimensional IV is weak). The horizontal axis represents the deviations in the parameter of interest from the maintained hypothesis, that is, we are interested in testing $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$, and $\delta = \beta - \beta_0$. See Section 5 in the main text for a detailed description of the simulation setup. All results are based on $5,000$ simulations.

Appendix B Additional Simulations

In this section, we present some additional simulation results to illustrate the effect of weak low-dimensional IVs. We set $\psi = 30$ so that the identification strength of many IVs remains relatively strong, and $\phi = 0$ so that the identification strength of the one-dimensional IV is rather weak. Figure 5 displays the power curves for $K = 100$ and $K = 500$, respectively, which can be regarded as extensions of Panels A and B of Figure 3 in the main text. Overall, the dominant performance of our combination test remains robust to different strengths of one-dimensional IV. In addition, we note that although the one-dimensional IV is relatively weak, the Wald test still provides a nontrivial gain in power for our combination test (as seen from the noticeable gaps between the power curves of ϕ_n^* and the jackknife LM test). This gain arises from its correlation with the jackknife LM statistic, in line with the theoretical result stated in Theorem A.1.

References

Adao, R., M. Kolesár, and E. Morales (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010.

Anatolyev, S. and N. Gospodinov (2011): "Specification testing in models with many instruments," *Econometric Theory*, 27, 427–441.

Andrews, D. W. (2017): "Identification-robust subvector inference," Tech. rep., Cowles Foundation Discussion Paper 2105.

Andrews, I. (2016): "Conditional linear combination tests for weakly identified models," *Econometrica*, 84, 2155–2182.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 11, 727–753.

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999): "Jackknife instrumental variables estimation," *Journal of Applied Econometrics*, 14, 57–67.

Angrist, J. D. and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earning?" *Quarterly Journal of Economics*, 106, 979–1014.

Bartik, T. J. (1991): *Who benefits from state and local economic development policies?*, Kalamazoo, MI: WE Upjohn Institute for Employment Research.

Bekker, P. (1994): "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica*, 62, 657–681.

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, 2369–2429.

Blanchard, O. J., L. F. Katz, R. E. Hall, and B. Eichengreen (1992): "Regional evolutions," *Brookings Papers on Economic Activity*, 1992, 1–75.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2026): "When is TSLS actually late?" *The Review of Economic Studies*.

Boot, T. and D. Nibbering (2024): "Inference on LATEs with Covariates," *arXiv preprint arXiv:2402.12607*.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-experimental shift-share research designs," *The Review of economic studies*, 89, 181–213.

——— (2025): "A practical guide to shift-share instruments," *Journal of Economic Perspectives*, 39, 181–204.

Card, D. (2009): "Immigration and Inequality," *American Economic Review: Papers & Proceedings*, 99, 1–21.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170, 383–398.

Carrasco, M. and G. Tchuente (2015): "Regularized LIML for many instruments," *Journal of Econometrics*, 186, 427–442.

Chao, J. C. and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): "Asymptotic Distribution Of JIVE In A Heteroskedastic IV Regression With Many Instruments," *Econometric Theory*, 28, 42–86.

Chao, J. C., N. R. Swanson, and T. Woutersen (2023): "Jackknife estimation of a cluster-sample IV regression model with many weak instruments," *Journal of Econometrics*.

Djogbenou, A. A., J. G. MacKinnon, and M. O. Nielsen (2019): "Asymptotic theory and wild bootstrap inference with clustered errors," *Journal of Econometrics*, 212, 393–412.

Donald, S. G. and W. K. Newey (2001): "Choosing the number of instruments," *Econometrica*, 69, 1161–1191.

Dovì, M.-S., A. B. Kock, and S. Mavroeidis (2024): "A Ridge-Regularized Jackknifed Anderson-Rubin Test," *Journal of Business & Economic Statistics*, 42, 1083–1094.

Dufour, J.-M. and M. Taamouti (2005): "Projection-based statistical inference in linear structural models with possibly weak instruments," *Econometrica*, 73, 1351–1365.

Evdokimov, K. S. and M. Kolesár (2018): "Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects," *Working Paper*.

Frandsen, B., E. Leslie, and S. McIntyre (2025): "Cluster jackknife instrumental variables estimation," *The Review of Economics and Statistics*, forthcoming.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019): "A more powerful subvector Anderson Rubin test in linear instrumental variables regression," *Quantitative Economics*, 10, 487–526.

——— (2021): "A Powerful Subvector Anderson Rubin Test in Linear Instrumental Variables Regression with Conditional Heteroskedasticity," *arXiv preprint arXiv:2103.11371*.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression," *Econometrica*, 80, 2649–2666.

Hansen, C., J. Hausman, and W. Newey (2008): "Estimation with Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26, 398–422.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3, 211–255.

Jiang, L., L. Li, K. Miao, and Y. Zhang (2025): "Adjustments with many regressors under covariate-adaptive randomizations," *Journal of Econometrics*, 249, 105991.

Kleibergen, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70, 1781–1803.

——— (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73, 1103–1123.

Kolesár, M. (2013): "Estimation in instrumental variables models with heterogeneous treatment effects," *Working Paper*.

——— (2018): "Minimum distance approach to inference with many instruments," *Journal of Econometrics*, 204, 86–100.

Kolesár, M., P. Min, W. Wang, and Y. Zhang (2026): "Cluster-Robust Inference for Quadratic Forms," *arXiv preprint arXiv:2602.13537*.

Kuersteiner, G. and R. Okui (2010): "Constructing optimal instruments by first-stage prediction averaging," *Econometrica*, 78, 697–718.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter (2022): "Valid t-ratio Inference for IV," *American Economic Review*, 112, 3260–3290.

Lehmann, E. L. and J. P. Romano (2006): *Testing Statistical Hypotheses*, Springer Science & Business Media.

Ligtenberg, J. W. (2025): "Inference in IV models with clustered dependence, many instruments and weak identification," *arXiv preprint arXiv:2306.08559*.

Lim, D., W. Wang, and Y. Zhang (2024a): "A conditional linear combination test with many weak instruments," *Journal of Econometrics*, 238, 105602.

——— (2024b): "A Dimension-Agnostic Bootstrap Anderson-Rubin Test For Instrumental Variable Regressions," *arXiv preprint arXiv:2412.01603*.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023): "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, 232, 272–299.

Matsushita, Y. and T. Otsu (2024): "A jackknife Lagrange multiplier test with many weak instruments," *Econometric Theory*, 40, 447–470.

Mikusheva, A. (2021): "Many weak instruments in time series econometrics," in *World Congress of Econometric Society. MIT*.

Mikusheva, A. and L. Sun (2022): "Inference with many weak instruments," *Review of Economic Studies*, 89, 2663–2686.

——— (2024): "Weak identification with many instruments," *The Econometrics Journal*, 27, C1–C28.

Morimune, K. (1983): "Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size," *Econometrica*, 51, 821–841.

Müller, U. K. (2011): "Efficient tests under a weak convergence assumption," *Econometrica*, 79, 395–435.

Navjeevan, M. (2024): "An Identification and Dimensionality Robust Test for Instrumental Variables Models," *arXiv preprint arXiv:2311.14892*.

Qian, H. and P. Schmidt (1999): "Improved instrumental variables and generalized method of moments estimators," *Journal of Econometrics*, 91, 145–169.

Sølvsten, M. (2020): "Robust estimation with many instruments," *Journal of Econometrics*, 214, 495–512.

Staiger, D. and J. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

Wang, W. and F. Doko Tchatoka (2018): "On Bootstrap inconsistency and Bonferroni-based size-correction for the subset Anderson–Rubin test under conditional homoskedasticity," *Journal of Econometrics*, 207, 188–211.

Wang, W. and M. Kaffo (2016): "Bootstrap inference for instrumental variable models with many weak instruments," *Journal of Econometrics*, 192, 231–268.

Yap, L. (2025): "Inference with Many Weak Instruments and Heterogeneity," *arXiv preprint arXiv:2408.11193*.

Young, A. (2022): "Consistency without inference: Instrumental variables in practical application," *European Economic Review*, 147, 104112.

Supplemental Appendix to "An Improved Inference for IV Regressions"

Liyu Dou[a], Pengjin Min[a], Wenjie Wang[b], and Yichong Zhang[a]

[a] *Singapore Management University, Singapore*
[b] *Nanyang Technological University, Singapore*

October 26, 2026

A Technical Lemmas

We use the following notation throughout Sections A–C. Recall the definition of $\tilde{Y}$, $\tilde{X}$, $\tilde{Z}$, $\tilde{z}$, W, $\tilde{\Pi}$, $\tilde{V}$ and $\tilde{e}$. Denote P_W as the projection matrix of W, $M_W = I_n - P_W$, and let $Y = M_W\tilde{Y}$, $X = M_W\tilde{X}$, $Z = M_W\tilde{Z}$, $z = M_W\tilde{z}$, $\Pi = M_W\tilde{\Pi}$, $V = M_W\tilde{V}$ and $e = M_W\tilde{e}$. Denote P as the projection matrix of Z, $M = I_n - P$, and $Q = M_W(P - \bar{P})M_W$, where $\bar{P}$ is the block diagonal matrix corresponding to P such that the g-th block on its diagonal is $P_{[g,g]}$; also denote $\bar{Q}$ as the block diagonal matrix corresponding to Q. Let $\acute{\Pi} = zA_nz^\top X$, $\hat{\Pi} = M_W(P - \bar{P})\Pi = Q\tilde{\Pi}$, $\bar{\Pi} = (Q - \bar{Q})\tilde{\Pi}$, $\acute{X} = z\hat{A}_nz^\top X$ and $\hat{X} = M_W(P - \bar{P})X$. Finally, we use $\Omega_{\tilde{e}}$ to denote the block diagonal matrix with g-th block $\Omega_g^{\tilde{e},\tilde{e}}$, and $\Omega_{\tilde{V}}$, $\Omega_{\tilde{e},\tilde{V}}$ and $\Omega_{\tilde{V},\tilde{e}}$ are defined similarly.

Lemma A.1. *Under Assumption 1, we have*

$$\max_{g\in[G]}\left(\tilde{\Pi}_{[g]}^\top\tilde{\Pi}_{[g]}\right)^2 + \max_{g\in[G]}\mathbb{E}\left(\tilde{V}_{[g]}^\top\tilde{V}_{[g]}\right)^2 + \max_{g\in[G]}\mathbb{E}\left(\tilde{e}_{[g]}^\top\tilde{e}_{[g]}\right)^2 \leqslant C,$$
$$\max_{g\in[G]}\left(\Pi_{[g]}^\top\Pi_{[g]}\right)^2 + \max_{g\in[G]}\mathbb{E}\left(V_{[g]}^\top V_{[g]}\right)^2 + \max_{g\in[G]}\mathbb{E}\left(e_{[g]}^\top e_{[g]}\right)^2 \leqslant C,$$

for some constant $C < \infty$. In addition, let $\hat{\gamma}_{\tilde{e}} = (W^\top W)^{-1}W^\top\tilde{e}$ and $\hat{\gamma}_{\tilde{V}} = (W^\top W)^{-1}W^\top\tilde{V}$,

we have

$$\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2 = o_P(1), \quad \max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{V}}\right\|_2 = o_P(1).$$

Lemma A.2. *If Assumptions 1 and 3 hold, then both P and Q are symmetric, and satisfy*

$$\begin{aligned}\|P\|_{op} = O(1), \quad \left\|\bar{P}\right\|_{op} = O(1), \quad \|P\|_F = O(\sqrt{K}),\\ \|Q\|_{op} = O(1), \quad \left\|\bar{Q}\right\|_{op} = o(1), \quad \|Q\|_F = O(\sqrt{K}).\end{aligned}$$

In addition, let $\tilde{P}$ be the block lower triangular matrix corresponding to $P-\bar{P}$ (i.e. $\tilde{P}_{[g,h]} = P_{[g,h]}$ for $g>h$ and $\tilde{P}_{[g,h]} = 0_{n_g\times n_h}$ otherwise), then

$$\left\|\tilde{P}\tilde{P}^\top\right\|_F = O(\sqrt{K}).$$

Lemma A.3. *Under Assumptions 1 and 3, we have*

$$\tilde{u}^\top\bar{P}P_W\tilde{v} = O_P(1), \quad \tilde{u}^\top P_W\bar{P}\tilde{v} = O_P(1), \quad \tilde{u}^\top\bar{P}P_W\bar{P}\tilde{v} = O_P(1),$$

for $(\tilde{u},\tilde{v})\in\{\tilde{V},\tilde{e}\}\times\{\tilde{V},\tilde{e}\}$.

Lemma A.4. *Under Assumptions 1 and 3, we have*

$$\begin{aligned}\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}u_{[h]}^\top P_{[h,g]}v_{[g]}\right)^2 &= \frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)^2 + o_P(1)\\ &= \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)^2 + o_P(1),\\ \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(u_{[h]}^\top P_{[h,g]}v_{[g]}\right)\left(u_{[g]}^\top P_{[g,h]}v_{[h]}\right) &= \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)\left(\tilde{u}_{[g]}^\top P_{[g,h]}\tilde{v}_{[h]}\right) + o_P(1)\\ &= \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)\left(\tilde{u}_{[g]}^\top P_{[g,h]}\tilde{v}_{[h]}\right) + o_P(1),\end{aligned}$$

for $(u,v)\in\{V,e\}\times\{V,e\}$, and the same results hold if we replace P with Q. In

addition, we have

$$\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)^2=\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{u}_{[h]}^\top Q_{[h,g]}\tilde{v}_{[g]}\right)^2+o(1),$$
$$\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{u}_{[h]}^\top P_{[h,g]}\tilde{v}_{[g]}\right)\left(\tilde{u}_{[g]}^\top P_{[g,h]}\tilde{v}_{[h]}\right)$$
$$=\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{u}_{[h]}^\top Q_{[h,g]}\tilde{v}_{[g]}\right)\left(\tilde{u}_{[g]}^\top Q_{[g,h]}\tilde{v}_{[h]}\right)+o(1),$$

for $(\tilde{u},\tilde{v})\in\{\tilde{V},\tilde{e}\}\times\{\tilde{V},\tilde{e}\}$.

Lemma A.5. *Under Assumptions 1 and 3, we have* $\Sigma\geq C(\Pi^\top\Pi+K)$ *for some constant* $C>0$*, and*

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2$$
$$=\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2+o_P(1),$$
$$=\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^\top\tilde{e}_{[g]}\right)^2+\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2+o_P(1),$$

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)^2$$
$$=\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2+o_P(1),$$

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}X_{[g]}\right)^2$$
$$=\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^\top\Pi_{[g]}\right)^2+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^\top\tilde{V}_{[g]}\right)^2$$
$$+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)^2$$

$$+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)^{2}+o_{P}(1),$$

and

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}e_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\\
&+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)+o_{P}(1).
\end{aligned}$$

Lemma A.6. *Under Assumptions 1 and 3, we have*

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)+o_{P}(1),
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)+o_{P}(1),
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\\
&+\frac{1}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\left(\tilde{\Pi}_{[g]}^{\top}Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\\
&+\frac{2}{\Sigma}\sum_{g,h\in[G]^{2},g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)+o_{P}(1),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)\\
&+\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)+o_P(1).
\end{aligned}$$

Lemma A.7. *Let $\acute{\beta}$ be a generic estimator of β. Further define*

$$\begin{aligned}
\acute{\Psi}&=X^{\top}z\hat{A}_n\acute{\Omega}\hat{A}_nz^{\top}X,\\
\acute{\Sigma}&=\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\acute{e}_{[g]}\right)^2+\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\acute{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\acute{e}_{[g]}\right),\\
\acute{\Upsilon}&=2\sum_{g,h\in[G]^2,g\neq h}\left(\acute{e}_{[g]}^{\top}P_{[g,h]}\acute{e}_{[h]}\right)^2,
\end{aligned}$$

where $\acute{e}=Y-X\acute{\beta}$ and

$$\acute{\Omega}=\sum_{g\in[G]}\left(z_{[g]}^{\top}\acute{e}_{[g]}\right)\left(z_{[g]}^{\top}\acute{e}_{[g]}\right)^{\top}.$$

Suppose that $\acute{\beta}\xrightarrow{p}\beta$, then the following holds.

1. *If Assumption 1 holds, then*

$$\hat{\Omega}^{-1/2}\Omega^{1/2}=I_{d_z}+o_P(1).$$

2. *If Assumptions 1 and 2 hold, then*

$$\frac{\acute{\Psi}}{\Psi}=1+o_P(1).$$

3. *If Assumptions 1 and 3 hold, then*

$$\frac{\acute{\Sigma}}{\Sigma}=1+o_P(1),$$

$$\frac{\acute{\Upsilon}}{\Upsilon} = 1 + o_P(1).$$

Lemma A.8. *Under Assumptions 1-3, we have* $\hat{\beta} \xrightarrow{p} \beta$ *and* $(\hat{\beta} - \beta)^2 \Pi^\top \Pi / \sqrt{K} = o_P(1)$. *Alternatively, if Assumptions 1, 2′ and 3 hold and* $\Pi^\top \Pi / \sqrt{K} \to \infty$, *we have* $\hat{\beta} \xrightarrow{p} \beta$ *and* $(\hat{\beta} - \beta)^2 \Pi^\top \Pi / \sqrt{K} = o_P(1)$.

Lemma A.9. *Under Assumptions 1-4, if the assumptions for* a_1 *and* a_2 *in Theorem 4.1 hold, then*

$$\begin{pmatrix} T(\beta_0) \\ LM(\beta_0) \\ AR \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\acute{\Psi}}} \sum_{g=1}^G \acute{\Pi}_{[g]}^\top \tilde{e}_{[g]} \\ \frac{1}{\sqrt{\Sigma}} \left(\sum_{g \in [G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h \in [G]^2, g \neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \\ \frac{1}{\sqrt{\Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \end{pmatrix} + \begin{pmatrix} a_1 \delta \\ a_2 \delta \\ 0 \end{pmatrix} + o_P(1).$$

Lemma A.10. *If Assumptions 1-4 hold, then*

$$\begin{pmatrix} \frac{1}{\sqrt{\acute{\Psi}}} \sum_{g=1}^G \acute{\Pi}_{[g]}^\top \tilde{e}_{[g]} \\ \frac{1}{\sqrt{\Sigma}} \left(\sum_{g \in [G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h \in [G]^2, g \neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \\ \frac{1}{\sqrt{\Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix} \right).$$

Lemma A.11. *Under Assumptions 1 and 3, we have*

$$\frac{1}{\sqrt{\Sigma}} \Omega^{-1/2} \sum_{g \in [G]} \left[\left(z_{[g]}^\top \hat{e}_{[g]} \right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]} \right) \right] = \frac{1}{\sqrt{\Sigma}} \Omega^{-1/2} \sum_{g \in [G]} \mathbb{E} \left[\left(z_{[g]}^\top \tilde{e}_{[g]} \right) \left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \right] + o_P(1),$$

and

$$\begin{aligned} &\frac{1}{\sqrt{\Sigma \Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \left(X_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \\ = &\frac{1}{\sqrt{\Sigma \Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + o_P(1). \end{aligned}$$

Lemma A.12. *Under Assumptions 1–4, we have*

$$\hat{\rho}_1 \xrightarrow{p} \rho_1, \quad \hat{\rho}_2 \xrightarrow{p} \rho_2.$$

If in addition the assumptions for a_1 *and* a_2 *in Theorem 4.1 hold, then*

$$\hat{\alpha}_1 \xrightarrow{p} \alpha_1, \quad \hat{\alpha}_2 \xrightarrow{p} \alpha_2.$$

B Proofs of Main Results

B.1 Proof of Proposition 3.1

The proof follows exactly the same lines as in the proof of Lemma 2.2. in Lim, Wang, and Zhang (2024) and is thus omitted. □

B.2 Proof of Theorem 4.1

The result follows from Lemma A.9, Lemma A.10 and the Slutsky theorem. □

B.3 Proof of Theorem 4.2

Consider the set $\mathcal{M}'$ of data generating processes m that satisfy the weak convergence result (3.12) pointwise for all $\delta \in \mathbb{R}$. It is straightforward to see that $\mathcal{M} \subset \mathcal{M}'$. As a result, the test class $\mathfrak{C}$ under consideration is a subset of an augmented class $\mathfrak{C}'$ of ϕ_n satisfying that

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_n\right] \leqslant \alpha \quad \text{for all } m \in \mathcal{M}', \delta = 0, \tag{B.1}$$

$$\liminf_{n\to\infty} \mathbb{E}\left[\phi_n\right] \geqslant \alpha \quad \text{for all } m \in \mathcal{M}', \delta \neq 0. \tag{B.2}$$

We also note that the oracle version of the combination test, ϕ_n^o in (3.13), can be understood as taking the weak convergence result (3.12) as the starting point and is simply the UMPU test in the limiting problem (under known $a_1(\alpha_1), a_2(\alpha_2), \rho_1, \rho_2$), evaluated at the sample analogues (Wald, LM, and AR statistics). By construction, ϕ_n^o satisfies (B.1) and (B.2), so $\phi_n^o \in \mathfrak{C}'$. Furthermore, by a direct application of Theorem 1 in Müller (2011), it follows that for any $\delta_1 \neq 0$ and any $\phi_n \in \mathfrak{C}'$,

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_n\right] \leqslant \lim_{n\to\infty} \mathbb{E}\left[\phi_n^o\right] \quad \text{for all } m \in \mathcal{M}', \delta = \delta_1,$$

which subsequently implies that for any $\delta_1 \neq 0$ and any $\phi_n \in \mathfrak{C}$,

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_n\right] \leqslant \lim_{n\to\infty} \mathbb{E}\left[\phi_n^o\right] \quad \text{for all } m \in \mathcal{M}, \delta = \delta_1.$$

Lastly, note that $\mathbb{E}\left[\phi_n^*\right]$ is continuous in $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\rho}_1, \hat{\rho}_2)$, which are consistent for $(\alpha_1, \alpha_2, \rho_1, \rho_2)$ under the assumptions of Theorem 4.1 by Lemma A.12. By the con-

tinuous mapping theorem, we thus have that, for any δ, $\mathbb{E}[\phi_n^*] = \mathbb{E}[\phi_n^o] + o_P(1)$. Therefore, we have ϕ_n^* satisfies (4.8) and (4.9), and thus $\phi_n^* \in \mathfrak{C}$. In addition, for any $\delta_1 \neq 0$, and any $\phi_n \in \mathfrak{C}$,

$$\lim_{n\to\infty} \mathbb{E}[\phi_n] \leqslant \lim_{n\to\infty} \mathbb{E}[\phi_n^o] = \lim_{n\to\infty} \mathbb{E}[\phi_n^*] \quad \text{for all } m \in \mathcal{M}, \delta = \delta_1.$$

For the second part, it is straightforward to check that $\tilde{\phi}_n \in \mathfrak{C}$. Also, note that the comparison of local asymptotic power of ϕ_n^* and $\tilde{\phi}_n$ can be reduced to the comparison of the non-centrality parameters for ϕ_n^o and $\tilde{\phi}_n$ in their limiting distributions, which are given by $(a^\top V^{-1} a)\delta^2$ and $a_1^2\delta^2$, respectively, where

$$a = \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix}.$$

Direct calculation yields

$$(a^\top V^{-1} a)\delta^2 - a_1^2\delta^2 = \frac{\delta^2}{1-\rho_1^2-\rho_2^2}(a_2 - \rho_1 a_1)^2 \geqslant 0,$$

and since $\rho_1^2 + \rho_2^2 < 1$ and $\delta \neq 0$, we obtain the desired result. □

B.4 Proof of Theorem 4.3

We shall distinguish between the two cases: (i) $\Pi^\top\Pi/\sqrt{K}$ is diverging and (ii) $\Pi^\top\Pi/\sqrt{K}$ is bounded, and argue along the appropriate subsequence as in Step 2 and 3 in the proof of Lemma A.8.

For the first case, we have

$$\begin{aligned} T(\beta_0) &= \frac{(X^\top z \hat{A}_n z^\top X)^{-1}(X^\top z \hat{A}_n z^\top e)}{\sqrt{\hat{\Phi}_1}} + \frac{\delta}{\sqrt{\hat{\Phi}_1}}, \\ LM(\beta_0) &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + \frac{\delta(X^\top(P-\bar{P})X)}{\sqrt{\hat{\Sigma}}}, \\ AR &= \frac{\hat{e}^\top(P-\bar{P})\hat{e}}{\sqrt{\hat{\Upsilon}}} \end{aligned}$$

under the fixed alternative. On the event $X^\top(P-\bar{P})X > 0$, we can write

$$\frac{\delta(X^\top(P-\bar{P})X)}{\sqrt{\hat{\Sigma}}} = \frac{\delta}{\sqrt{(X^\top(P-\bar{P})X)^{-1}\hat{\Sigma}(X^\top(P-\bar{P})X)^{-1}}} = \frac{\delta}{\sqrt{\widehat{\Phi}_2}}.$$

By repeating the proof of Lemma A.9 for the other terms in the above expression beside $\delta/\sqrt{\widehat{\Phi}_1}$ and $\delta/\sqrt{\widehat{\Phi}_2}$ (which do not depend on a_1 and a_2), we can write

$$\begin{pmatrix} T(\beta_0) \\ LM(\beta_0) \\ AR \end{pmatrix} = \begin{pmatrix} \delta/\sqrt{\widehat{\Phi}_1} \\ \delta/\sqrt{\widehat{\Phi}_2} \\ 0 \end{pmatrix} + R_n,$$

where $R_n = O_p(1)$. Recall the definition $(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)'$:

$$\begin{pmatrix} \hat{\omega}_1 \\ \hat{\omega}_2 \\ \hat{\omega}_3 \end{pmatrix} = \frac{1}{\sqrt{\hat{b}_1^2 + \hat{b}_2^2 + \hat{b}_3^2}} \times \begin{pmatrix} 1 & \hat{\rho}_1 & 0 \\ \hat{\rho}_1 & 1 & \hat{\rho}_2 \\ 0 & \hat{\rho}_2 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix}, \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & 0 \\ \hat{\rho}_1 & 1 & \hat{\rho}_2 \\ 0 & \hat{\rho}_2 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ 0 \end{pmatrix},$$

and note that $(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)' = O_p(1)$. We have

$$\begin{aligned}
0 &\leq 1 - \mathbb{E}\left[\phi_n^*\right] \\
&= \mathbb{P}\left((\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR)^2 < \mathbb{C}_\alpha\right) \\
&\leq \mathbb{P}\left((\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR)^2 < \mathbb{C}_\alpha, X^\top(P-\bar{P})X > 0\right) + \mathbb{P}\left(X^\top(P-\bar{P})X \leqslant 0\right) \\
&\leq \mathbb{P}\left(\left(\delta\sqrt{\frac{1}{\widehat{\Phi}_1} + \frac{1}{\widehat{\Phi}_2}} \times \sqrt{\hat{\alpha}^\top \hat{V}^{-1}\hat{\alpha}} + O_p(1)\right)^2 < \mathbb{C}_\alpha\right) + \mathbb{P}\left(X^\top(P-\bar{P})X \leqslant 0\right),
\end{aligned}$$

where

$$\hat{\alpha} \equiv \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ 0 \end{pmatrix}, \quad \hat{V} \equiv \begin{pmatrix} 1 & \hat{\rho}_1 & 0 \\ \hat{\rho}_1 & 1 & \hat{\rho}_2 \\ 0 & \hat{\rho}_2 & 1 \end{pmatrix}, \quad \begin{pmatrix} \delta/\sqrt{\widehat{\Phi}_1} \\ \delta/\sqrt{\widehat{\Phi}_2} \\ 0 \end{pmatrix} = \delta\sqrt{\frac{1}{\widehat{\Phi}_1} + \frac{1}{\widehat{\Phi}_2}} \times \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ 0 \end{pmatrix},$$

and we used the fact that $\hat{\omega}_1 \times \hat{\alpha}_1 + \hat{\omega}_2 \times \hat{\alpha}_2 + \hat{\omega}_3 \times 0 = \sqrt{\hat{\alpha}^\top \hat{V}^{-1}\hat{\alpha}}$. In addition, we have $\widehat{\Phi}_1/\Phi_1 \xrightarrow{p} 1$ and $\widehat{\Phi}_2/\Phi_2 \xrightarrow{p} 1$ from the intermediate steps in the proof of Lemma A.8 (see (C.20) and (C.26)) and Lemma A.7. Hence, $\widehat{\Phi}_1 = o_p(1)$ and $\widehat{\Phi}_2 = o_p(1)$,

since it is also established in the those steps of the proof of Lemma A.8 (cf. (C.21) and (C.25)) that $\Phi_1 = o(1)$ and $\Phi_2 = o(1)$. Therefore, we have

$$\delta\sqrt{\frac{1}{\hat{\Phi}_1}+\frac{1}{\hat{\Phi}_2}} \times \sqrt{\hat{\alpha}^\top \hat{V}^{-1}\hat{\alpha}} \xrightarrow{p} \infty(-\infty), \quad \text{for } \delta > 0 \ (\delta < 0),$$

where we use the fact that $\|\hat{\alpha}\|_2 = 1$ by construction, $\hat{V}^{-1} \xrightarrow{p} V^{-1}$ where

$$V \equiv \begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{pmatrix},$$

by Lemma A.12, whence $\hat{\alpha}^\top \hat{V}^{-1}\hat{\alpha}$ can be bounded away from zero with probability approaching one. Together with the fact that $X^\top(P - \bar{P})X > 0$ with probability approaching one, this implies that $\lim_{n\to\infty} \mathbb{E}[\phi_n^*] = 1$.

For the second case, by Lemma A.12, we have $\hat{\rho}_1 \xrightarrow{p} \rho_1$ and $\hat{\rho}_2 \xrightarrow{p} \rho_2$, and note that $\rho_1 = 0$ in this case since $\Pi^\top \Pi/K \to 0$. In addition, similar to the proof of Step 3 of Lemma A.8, it can be shown that $\hat{\Phi}_2 \rightsquigarrow \bar{\Phi}_2$ for some random variable $\bar{\Phi}_2$ such that $\bar{\Phi}_2 > 0$ with probability one. Also, since, as shown above, $\hat{\Phi}_1 \xrightarrow{p} 0$, and by Lemma A.7, it follows that $\hat{\alpha}_1 \xrightarrow{p} 1$ and $\hat{\alpha}_2 \xrightarrow{p} 0$, by the continuous mapping theorem. Therefore, we have $\hat{\omega}_1 \xrightarrow{p} 1$, $\hat{\omega}_2 \xrightarrow{p} 0$ and $\hat{\omega}_3 \xrightarrow{p} 0$. Finally, as in the proof for the first case above, we have $T(\beta_0) = \delta/\sqrt{\hat{\Phi}_1} + O_p(1)$ where $\hat{\Phi}_1 = o_p(1)$, $LM(\beta_0) = O_p(1)$ and $AR = O_p(1)$. It follows that $\lim_{n\to\infty} \mathbb{E}[\phi_n^*] = 1$. This concludes the proof. □

B.5 Proof of Theorem A.1

We shall argue along the appropriate subsequence as in Step 4 of the proof of Lemma A.8. Suppose we are under the local alternative that $\beta - \beta_0 = \delta d_n$. We have

$$\begin{aligned}
\hat{\rho}_1 &= \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}} \sum_{g\in[G]} \left[\left(\acute{X}_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right)\right] \\
&= \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{1}{\sqrt{X^\top z \hat{A}_n \hat{\Omega} \hat{A}_n z^\top X}} X^\top z \hat{A}_n \sqrt{\hat{\Omega}} \\
&\times \sqrt{\frac{\Omega}{\hat{\Omega}}} \times \frac{1}{\sqrt{\Omega\Sigma}} \sum_{g\in[G]} \left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right)\right],
\end{aligned}$$

and note the important fact that

$$\left(\frac{1}{\sqrt{X^\top z\hat{A}_n\hat{\Omega}\hat{A}_n z^\top X}}X^\top z\hat{A}_n\sqrt{\hat{\Omega}}\right)^2 = 1$$

when $d_z = 1$, and thus

$$\hat{\rho}_1^2 = \frac{\Sigma}{\hat{\Sigma}} \times \frac{\Omega}{\hat{\Omega}} \times \left(\frac{1}{\sqrt{\Omega\Sigma}}\sum_{g\in[G]}\left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right)\right]\right)^2.$$

By Lemmas A.7 and A.11, we have $\hat{\rho}_1^2 \xrightarrow{p} \bar{\rho}_1^2$. Together with the fact that $\hat{\rho}_2^2 \xrightarrow{p} \bar{\rho}_2^2$ by Lemmas A.7 and A.11, this implies that $\hat{\rho}_1^2 + \hat{\rho}_2^2 < 1$ with probability approaching one. On that event, direct calculation yields

$$\begin{aligned}
&\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR \\
&= \frac{T(\beta_0)\hat{\alpha}_1(1-\hat{\rho}_2^2) - T(\beta_0)\hat{\rho}_1\hat{\alpha}_2 + LM(\beta_0)(\hat{\alpha}_2 - \hat{\rho}_1\hat{\alpha}_1) + AR\hat{\rho}_2(\hat{\rho}_1\hat{\alpha}_1 - \hat{\alpha}_2)}{\sqrt{1-\hat{\rho}_1^2-\hat{\rho}_2^2}\sqrt{\hat{\alpha}_1^2(1-\hat{\rho}_2^2) - 2\hat{\rho}_1\hat{\alpha}_1\hat{\alpha}_2 + \hat{\alpha}_2^2}}.
\end{aligned}$$

We shall analyze each term in turn. To begin with, we note that, similar to the proof of Step 4 of Lemma A.8, it can be shown that $\hat{\Phi}_1 \rightsquigarrow \bar{\Phi}_1$ for some random variable $\bar{\Phi}_1$ such that $\bar{\Phi}_1 > 0$ with probability one, and since $\hat{\Phi}_2 \xrightarrow{p} 0$ by (C.25), (C.26) and Lemma A.7, it follows that $\hat{\alpha}_1 \xrightarrow{p} 0$ and $\hat{\alpha}_2 \xrightarrow{p} 1$, by the continuous mapping theorem. Note also that

$$\begin{aligned}
T(\beta_0) &= \frac{(X^\top z\hat{A}_n z^\top X)^{-1}}{\sqrt{(X^\top z\hat{A}_n z^\top X)^{-2}}} \times \frac{1}{\sqrt{X^\top z\hat{A}_n\hat{\Omega}\hat{A}_n z^\top X}}X^\top z\hat{A}_n\sqrt{\hat{\Omega}} \\
&\quad \times \sqrt{\frac{\Omega}{\hat{\Omega}}} \times \frac{1}{\sqrt{\Omega}}\sum_{g\in[G]} z_{[g]}^\top \tilde{e}_{[g]} + \frac{d_n}{\sqrt{\hat{\Phi}_1}}\delta \\
&= O_p(1),
\end{aligned}$$

and thus

$$T(\beta_0)\hat{\alpha}_1(1-\hat{\rho}_2^2) = o_P(1).$$

Next, we note that

$$T(\beta_0)\hat{\rho}_1 = \frac{(X^\top z\hat{A}_n z^\top X)^{-1}}{\sqrt{(X^\top z\hat{A}_n z^\top X)^{-2}}} \times \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{\Omega}{\hat{\Omega}} \\ \times \frac{1}{\sqrt{\Omega}} \sum_{g\in[G]} z_{[g]}^\top \tilde{e}_{[g]} \times \frac{1}{\sqrt{\Omega\Sigma}} \sum_{g\in[G]} \left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right)\right] + o_P(1).$$

In addition, we can show that

$$\frac{1}{\sqrt{n}} z^\top X \rightsquigarrow \mathcal{N}\left(\pi, \Omega_z^{\tilde{V},\tilde{V}}\right),$$

which, combining with the fact that $\hat{A}_n/\lambda_n \xrightarrow{p} A$, implies that

$$\frac{(X^\top z\hat{A}_n z^\top X)^{-1}}{\sqrt{(X^\top z\hat{A}_n z^\top X)^{-2}}} = 1 + o_P(1).$$

If follows that, by Lemmas A.7 and A.11,

$$T(\beta_0)\hat{\rho}_1\hat{\alpha}_2 = \bar{\rho}_1 \times \frac{1}{\sqrt{\Omega}} \sum_{g\in[G]} z_{[g]}^\top \tilde{e}_{[g]} + o_P(1).$$

Next, we note that

$$LM(\beta_0) = \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + a\delta + o_P(1) \\ = \frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) + a\delta + o_P(1),$$

where the first equality is by (C.26) and Lemma A.7, and the second equality is by Lemmas A.3 and A.7. It follows that

$$LM(\beta_0)(\hat{\alpha}_2 - \hat{\rho}_1\hat{\alpha}_1) = \frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) + a\delta + o_P(1).$$

Next, we note that

$$AR = \frac{1}{\sqrt{\hat{\Upsilon}}}\left(\hat{e}^\top(P-\bar{P})\hat{e} - e^\top(P-\bar{P})e\right) + \frac{1}{\sqrt{\hat{\Upsilon}}}\left(e^\top(P-\bar{P})e - \tilde{e}^\top(P-\bar{P})\tilde{e}\right) + \frac{1}{\sqrt{\hat{\Upsilon}}}\tilde{e}^\top(P-\bar{P})\tilde{e}$$

$$= \frac{1}{\sqrt{\Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} + o_P(1).$$

by (C.22), (C.24), and Lemmas A.3 and A.7. It follows that

$$AR\hat{\rho}_2(\hat{\rho}_1\hat{\alpha}_1 - \hat{\alpha}_2) = -\bar{\rho}_2 \times \frac{1}{\sqrt{\Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} + o_P(1)$$

by Lemma A.11. Finally, we have

$$\begin{aligned} &1 - \hat{\rho}_1^2 - \hat{\rho}_2^2 \xrightarrow{p} 1 - \bar{\rho}_1^2 - \bar{\rho}_2^2, \\ &\hat{\alpha}_1^2(1 - \hat{\rho}_2^2) - 2\hat{\rho}_1\hat{\alpha}_1\hat{\alpha}_2 + \hat{\alpha}_2^2 \xrightarrow{p} 1. \end{aligned}$$

Combining all the results, we have

$$\begin{aligned} &\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR \\ &= \frac{-\bar{\rho}_1}{\sqrt{1 - \bar{\rho}_1^2 - \bar{\rho}_2^2}} \times \frac{1}{\sqrt{\Omega}} \sum_{g \in [G]} z_{[g]}^\top \tilde{e}_{[g]} \\ &+ \frac{1}{\sqrt{1 - \bar{\rho}_1^2 - \bar{\rho}_2^2}} \times \left(\frac{1}{\sqrt{\Sigma}} \left(\sum_{g \in [G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h \in [G]^2, g \neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + a\delta \right) \\ &+ \frac{-\bar{\rho}_2}{\sqrt{1 - \bar{\rho}_1^2 - \bar{\rho}_2^2}} \times \frac{1}{\sqrt{\Upsilon}} \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} + o_P(1). \end{aligned}$$

Furthermore, we note that in the proof of Lemma A.10, we only require $\acute{\Pi}$ to satisfy

$$\frac{1}{\Psi} \sum_{g \in [G]} \acute{\Pi}_{[g]}^\top \acute{\Pi}_{[g]} = O(1), \quad \frac{1}{\Psi} \max_{g \in [G]} \acute{\Pi}_{[g]}^\top \acute{\Pi}_{[g]} = o(1),$$

which is guaranteed by Assumptions 1 and 2. We also have

$$\frac{1}{\Omega} \sum_{g \in [G]} z_{[g]}^\top z_{[g]} = O(1), \quad \frac{1}{\Omega} \max_{g \in [G]} z_{[g]}^\top z_{[g]} = o(1),$$

by Assumption 1 alone. Therefore, we can replace $\frac{1}{\sqrt{\Psi}} \sum_{g=1}^G \acute{\Pi}_{[g]}^\top \tilde{e}_{[g]}$ with $\frac{1}{\sqrt{\Omega}} \sum_{g \in [G]} z_{[g]}^\top \tilde{e}_{[g]}$ in the proof of Lemma A.10. Note that in this way, Assumption 2 is no longer needed.

We thus can follow the same argument in the proof of Lemma A.10 and obtain that

$$\begin{pmatrix} \frac{1}{\sqrt{\Omega}} \sum_{g\in[G]} z_{[g]}^\top \tilde{e}_{[g]} \\ \frac{1}{\sqrt{\Sigma}} \left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + a\delta \\ \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ a\delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \bar{\rho}_1 & 0 \\ \bar{\rho}_1 & 1 & \bar{\rho}_2 \\ 0 & \bar{\rho}_2 & 1 \end{pmatrix} \right).$$

The desired result then follows.

Next, suppose we are under the fixed alternative. From the proof above, we have

$$\begin{aligned} &\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR \\ &= \frac{(1+o_P(1))}{\sqrt{1-\bar{\rho}_1^2-\bar{\rho}_2^2}} LM(\beta_0) + O_P(1) \\ &= \frac{(1+o_P(1))}{\sqrt{1-\bar{\rho}_1^2-\bar{\rho}_2^2}} \times \left(\frac{\delta}{\sqrt{\widehat{\Phi}_2}} \times (1+o_P(1)) + O_P(1) \right) + O_P(1) \end{aligned}$$

by (C.26) and Lemma A.7, and note that $\widehat{\Phi}_2 = o_P(1)$ by (C.25) and Lemma A.7. It follows that

$$(\hat{\omega}_1 T(\beta_0) + \hat{\omega}_2 LM(\beta_0) + \hat{\omega}_3 AR)^2 \xrightarrow{p} \infty,$$

and the desired result follows. This concludes the proof. □

B.6 Proof of Theorem A.2

Under the local alternative, similar to the proof of Theorem A.1, we have

$$\begin{aligned} T(\beta_0) &= \frac{(X^\top z \hat{A}_n z^\top X)^{-1}}{\sqrt{(X^\top z \hat{A}_n z^\top X)^{-2}}} \times \frac{1}{\sqrt{X^\top z \hat{A}_n \hat{\Omega} \hat{A}_n z^\top X}} X^\top z \hat{A}_n \hat{\Omega}^{1/2} \\ &\quad \times \hat{\Omega}^{-1/2} \Omega^{1/2} \times \Omega^{-1/2} \sum_{g\in[G]} z_{[g]}^\top \tilde{e}_{[g]} + \frac{d_n}{\sqrt{\widehat{\Phi}_1}} \delta = O_p(1), \\ LM(\beta_0) &= \frac{1}{\sqrt{\Sigma}} \left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + a\delta + o_p(1), \\ AR &= \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} + o_p(1). \end{aligned}$$

By repeating the proof of Lemma A.10 and ignoring $\frac{1}{\sqrt{\Psi}}\sum_{g=1}^{G}\acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}$ (and thus Assumption 2 is not needed), we have

$$\begin{pmatrix} LM(\beta_0) \\ AR \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} a\delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

by the Slutsky theorem. For $\hat{\rho}_1$, we have

$$\begin{aligned}
\hat{\rho}_1 &= \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}}\sum_{g\in[G]}\left[\left(\acute{X}_{[g]}^{\top}\hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^{\top}\hat{e}_{[g]}\right)\right] \\
&= \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{1}{\sqrt{X^{\top}z\hat{A}_n\hat{\Omega}\hat{A}_nz^{\top}X}}X^{\top}z\hat{A}_n\hat{\Omega}^{1/2} \\
&\times \hat{\Omega}^{-1/2}\Omega^{1/2} \times \frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\left[\left(z_{[g]}^{\top}\hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^{\top}\hat{e}_{[g]}\right)\right],
\end{aligned}$$

where

$$\frac{1}{\sqrt{X^{\top}z\hat{A}_n\hat{\Omega}\hat{A}_nz^{\top}X}}X^{\top}z\hat{A}_n\hat{\Omega}^{1/2} = O_P(1),$$

and by Lemma A.11,

$$\frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\left[\left(z_{[g]}^{\top}\hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^{\top}\hat{e}_{[g]}\right)\right] = \frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\mathbb{E}\left[\left(z_{[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\right] + o_P(1).$$

Further note that the first term on the right-hand side of the above display is $o(1)$ since $\Pi^{\top}\Pi/K \to 0$, so that $\hat{\rho}_1 \xrightarrow{p} 0$ by Lemma A.7. In addition, we have $\hat{\rho}_2 \xrightarrow{p} \rho$ by Lemmas A.7 and A.11. Finally, similar to the proof of Step 4 of Lemma A.8, it can be shown that $\hat{\Phi}_1 \rightsquigarrow \bar{\Phi}_1$ for some random variable $\bar{\Phi}_1$ such that $\bar{\Phi}_1 > 0$ with probability one, and since $\hat{\Phi}_2 \xrightarrow{p} 0$ by (C.25), (C.26) and Lemma A.7, it follows that $\hat{\alpha}_1 \xrightarrow{p} 0$ and $\hat{\alpha}_2 \xrightarrow{p} 1$, by the continuous mapping theorem. This implies that $\hat{\omega}_1 \xrightarrow{p} 0$, $\hat{\omega}_2 \xrightarrow{p} 1/\sqrt{1-\rho^2}$ and $\hat{\omega}_3 \xrightarrow{p} -\rho/\sqrt{1-\rho^2}$, whence

$$\begin{aligned}
&\hat{\omega}_1T(\beta_0) + \hat{\omega}_2LM(\beta_0) + \hat{\omega}_3AR \\
&= \frac{1}{\sqrt{1-\rho^2}}LM(\beta_0) - \frac{\rho}{\sqrt{1-\rho^2}}AR + o_p(1) \\
&\rightsquigarrow \frac{1}{\sqrt{1-\rho^2}}\mathcal{N}_1 - \frac{\rho}{\sqrt{1-\rho^2}}\mathcal{N}_2,
\end{aligned}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are defined in Theorem A.2.

Under the fixed alternative, similar to the proof of Theorem A.1, we have $\lim_{n\to\infty} \mathbb{E}\left[\phi_n^*\right] = 1$. This concludes the proof. □

C Proofs of Technical Lemmas

C.1 Proof of Lemma A.1

The first result follows readily from Assumption 1. For the second result, note that for any $g \in [G]$, we have

$$\left(\Pi_{[g]}^\top \Pi_{[g]}\right)^2 = \left(\sum_{i\in I_g} \Pi_{i,g}^2\right)^2 \leqslant C \sum_{i\in I_g} \Pi_{i,g}^4 \leqslant C.$$

In addition, by an abuse of notation, for any $i \in [n]$ and $g \in [G]$, we denote $M_W^{(i)}$ as the i-th column of M_W, and $M_{W,[g]}^{(i)} \in \Re^{n_g}$ is vector that collects all elements in the n-dimensional vector $M_W^{(i)}$ that belong to the g-th cluster. Then, we have

$$\begin{aligned}
\mathbb{E}V_{i,g}^4 &= \mathbb{E}\left(\sum_{g\in[G]} M_{W,[g]}^{(i),\top} \tilde{V}_{[g]}\right)^4 \\
&\leqslant C\left(\sum_{g\in[G]} \left(M_{W,[g]}^{(i),\top} M_{W,[g]}^{(i)}\right)^2 + \sum_{g,h\in[G]^2, h\neq g} \left(M_{W,[g]}^{(i),\top} M_{W,[g]}\right)\left(M_{W,[h]}^{(i),\top} M_{W,[h]}\right)\right) \\
&\leqslant C,
\end{aligned}$$

by the first result and the fact that

$$\sum_{g\in[G]} \left(M_{W,[g]}^{(i),\top} M_{W,[g]}^{(i)}\right) = M_W^{(i),\top} M_W^{(i)} = M_{W,ii} \leqslant C.$$

It follows that

$$\mathbb{E}\left(V_{[g]}^\top V_{[g]}\right)^2 = \mathbb{E}\left(\sum_{i\in I_g} V_{i,g}^2\right)^2 \leqslant C \sum_{i\in I_g} \mathbb{E}V_{i,g}^4 \leqslant C.$$

Using the same argument, we also have

$$\mathbb{E}\left(e_{[g]}^{\top} e_{[g]}\right)^2 \leqslant C,$$

and the desired result follows. Finally, for the last result, by Assumption 1, we have $\hat{\gamma}_{\tilde{e}} = O_P(1/\sqrt{n})$ and $\hat{\gamma}_{\tilde{V}} = O_P(1/\sqrt{n})$, and thus

$$\max_{g\in[G]} \left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2 \leqslant \max_{g\in[G]} n_g \times \max_{i\in I_g, g\in[G]} \|W_{i,g}\|_2^2 \times \|\hat{\gamma}_{\tilde{e}}\|_2^2 = o_P(1),$$
$$\max_{g\in[G]} \left\|W_{[g]}\hat{\gamma}_{\tilde{V}}\right\|_2^2 \leqslant \max_{g\in[G]} n_g \times \max_{i\in I_g, g\in[G]} \|W_{i,g}\|_2^2 \times \|\hat{\gamma}_{\tilde{V}}\|_2^2 = o_P(1).$$

□

C.2 Proof of Lemma A.2

For the first part of Lemma A.2, the results for P are standard for projection matrix, so we focus on the results for Q. We have

$$\|Q\|_{op} = \left\|M_W(P-\bar{P})M_W\right\|_{op} \leqslant \|M_W\|_{op}^2 \left\|P-\bar{P}\right\|_{op} = O(1),$$

and

$$\|Q\|_F = \sqrt{\text{trace}(M_W(P-\bar{P})M_W(P-\bar{P})M_W)} \leqslant C\left\|P-\bar{P}\right\|_F = O(\sqrt{K}).$$

In addition, we note that

$$\bar{Q} = \bar{P}_W\bar{P} + \bar{P}\bar{P}_W - \hat{P},$$

where $\bar{P}_W$ is the block diagonal matrix corresponding to P_W and $\hat{P}$ is a block diagonal matrix such that the g-th block on its diagonal is $\sum_{h=1}^G P_{W,[g,h]}P_{[h,h]}P_{W,[h,g]}$ (corresponding to the block diagonals of $P_W\bar{P}P_W$). By Assumption 1, we have

$$\max_{1\leq i\leq n} P_{W,ii} \leq \max_{i\in I_g, g\in[G]} \|W_{i,g}\|_2^2 \times \lambda_{\max}\left((W^{\top}W)^{-1}\right) = o(1),$$

and thus

$$\max_{g\in[G]} \lambda_{\max}\left(P_{W,[g,g]}\right) \leq \max_{g\in[G]} n_g \times \max_{1\leq i\leq n} P_{W,ii} = o(1),$$

which implies that $\lambda_{\max}\left(\bar{P}_W\right) = o(1)$. It follows that

$$\begin{aligned}
\left\|\bar{Q}\right\|_{op} &\leqslant \left\|\bar{P}_W\bar{P}\right\|_{op} + \left\|\bar{P}\bar{P}_W\right\|_{op} + \left\|\hat{P}\right\|_{op} \\
&\leqslant 2\left\|\bar{P}_W\right\|_{op}\left\|\bar{P}\right\|_{op} + \max_{g\in[G]}\left\|\sum_{h\in[G]} P_{W,[g,h]}P_{[h,h]}P_{W,[h,g]}\right\|_{op} \\
&\leq C \max_{1\leq g,h\leq G}\left\|P_{W,[g,h]}P_{W,[h,g]}\right\|_{op} + o(1) \\
&\leqslant C \max_{g\in[G]}\left\|P_{W,[g,g]}\right\|_{op} + o(1) \\
&= o(1).
\end{aligned}$$

For the second part of Lemma A.2, we shall use an argument similar to Chao, Swanson, Hausman, Newey, and Woutersen (2012). A closer inspection of their proof suggests that, all the equalities in the proof of their Lemma B.2. remain unchanged if we replace P_{ij} with $P_{[g,h]}$ and keep the trace operator; note also that we have $\text{trace}\left((P-\bar{P})^4\right) = O(K)$ so that (i) of Lemma B.2. still holds. To obtain (iii) of Lemma B.2. we establish results similar to their Lemma B.1.: for any subset $\mathcal{I}_2$ of the set $\{g,h\}_{g,h=1}^G$, we have

$$\begin{aligned}
\text{trace}\left(\sum_{\mathcal{I}_2} P_{[g,h]}P_{[h,g]}P_{[g,h]}P_{[h,g]}\right) &\leqslant C\sum_{\mathcal{I}_2}\text{trace}\left(P_{[g,h]}P_{[h,g]}\right) \\
&\leqslant C\sum_{g,h\in[G]^2}\text{trace}\left(P_{[g,h]}P_{[h,g]}\right) \\
&= O(K),
\end{aligned}$$

and similarly for any subset $\mathcal{I}_3$ of the set $\{g,h,k\}_{g,h,k=1}^G$, we have

$$\begin{aligned}
\text{trace}\left(\sum_{\mathcal{I}_3} P_{[g,h]}P_{[h,k]}P_{[k,h]}P_{[h,g]}\right) &= O(K), \\
\text{trace}\left(\sum_{\mathcal{I}_3} P_{[g,h]}P_{[h,g]}P_{[g,k]}P_{[k,g]}\right) &= O(K),
\end{aligned}$$

and then it is easy to see that (iii) of Lemma B.2. holds. To obtain (ii) of Lemma

B.2. we define, as in their paper, the following random variables

$$\Delta_1 = \sum_{g<h<k} \left(\xi_{[h]}^\top P_{[h,g]} P_{[g,k]} \xi_{[k]} + \xi_{[g]}^\top P_{[g,h]} P_{[h,k]} \xi_{[k]} + \xi_{[g]}^\top P_{[g,k]} P_{[k,h]} \xi_{[h]} \right)$$
$$\Delta_2 = \sum_{g<h<k} \left(\xi_{[h]}^\top P_{[h,g]} P_{[g,k]} \xi_{[k]} + \xi_{[g]}^\top P_{[g,h]} P_{[h,k]} \xi_{[k]} \right)$$
$$\Delta_3 = \sum_{g<h<k} \left(\xi_{[g]}^\top P_{[g,k]} P_{[k,h]} \xi_{[h]} \right)$$

where $\{\xi_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with mean 0 and variance 1, and independent of $\tilde{Z}, W$ (note that ξ_i are not only independent across clusters but also within clusters), and then it is straightforward to verify that (ii) of Lemma B.2. also holds. These results, together with a similar argument as in the proof of their Lemma B.3., allow us to conclude that $\left\| \tilde{P} \tilde{P}^\top \right\|_F = O(\sqrt{K})$. This concludes the proof. □

C.3 Proof of Lemma A.3

We focus on the case when $\tilde{u} = \tilde{V}$ and $\tilde{v} = \tilde{e}$. We have

$$\left| \tilde{V}^\top P_W \bar{P} \tilde{e} \right| \leqslant \sqrt{\tilde{V}^\top P_W \tilde{V}} \times \sqrt{\tilde{e}^\top \bar{P} P_W \bar{P} \tilde{e}} = O_P(1),$$

because

$$\mathbb{E}\left(\tilde{V}^\top P_W \tilde{V} \right) = \text{trace}\left(P_W \Omega_{\tilde{V}} \right) \leq \lambda_{\max}(\Omega_{\tilde{V}}) \text{trace}\left(P_W \right) \leqslant C,$$
$$\mathbb{E}\left(\tilde{e}^\top \bar{P} P_W \bar{P} \tilde{e} \right) = \text{trace}\left(\bar{P} P_W \bar{P} \Omega_{\tilde{e}} \right) \leq \lambda_{\max}(\Omega_{\tilde{e}}) \lambda_{\max}(\bar{P}^2) \text{trace}\left(P_W \right) \leqslant C,$$

since d_w is fixed. The other two terms can be handled similarly. □

C.4 Proof of Lemma A.4

For the first result, we focus on the case when $u = V$ and $v = e$. To show

$$\frac{1}{K} \sum_{g \in [G]} \left(\sum_{h \in [G], h \neq g} V_{[h]}^\top P_{[h,g]} e_{[g]} \right)^2 - \frac{1}{K} \sum_{g \in [G]} \left(\sum_{h \in [G], h \neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 = o_P(1), \quad \text{(C.1)}$$

we note that

$$\begin{aligned}
&\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} V_{[h]}^{\top}P_{[h,g]}e_{[g]}\right)^2 \\
&= \frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} (\tilde{V}_{[h]} - W_{[h]}\hat{\gamma}_{\tilde{V}})^{\top}P_{[h,g]}(\tilde{e}_{[g]} - W_{[g]}\hat{\gamma}_{\tilde{e}})\right)^2 \\
&= \frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\begin{pmatrix} \tilde{V}_{[h]}P_{[h,g]}\tilde{e}_{[g]} - (W_{[h]}\hat{\gamma}_{\tilde{V}})^{\top}P_{[h,g]}\tilde{e}_{[g]} \\ -\tilde{V}_{[h]}^{\top}P_{[h,g]}(W_{[g]}\hat{\gamma}_{\tilde{e}}) + (W_{[h]}\hat{\gamma}_{\tilde{V}})^{\top}P_{[h,g]}(W_{[g]}\hat{\gamma}_{\tilde{e}})\end{pmatrix}\right)^2.
\end{aligned}$$

We have

$$\begin{aligned}
\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}(W_{[h]}\hat{\gamma}_{\tilde{V}})^{\top}P_{[h,g]}\tilde{e}_{[g]}\right)^2 &= \frac{1}{K}\sum_{g\in[G]}\left((W_{[g]}\hat{\gamma}_{\tilde{V}})^{\top}P_{[g,g]}\tilde{e}_{[g]}\right)^2 \\
&\leq \frac{C\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{V}}\right\|_2^2}{K}\sum_{g\in[G]}\left\|P_{[g,g]}\tilde{e}_{[g]}\right\|_2^2 \\
&= o_P(1),
\end{aligned}$$

by Lemma A.1, where we use the fact that $W^{\top}P = 0_{n\times n}$ and

$$\mathbb{E}\frac{1}{K}\sum_{g\in[G]}\left\|P_{[g,g]}\tilde{e}_{[g]}\right\|_2^2 \leq \frac{C}{K}\sum_{g\in[G]}\operatorname{trace}(P_{[g,g]}) = O(1).$$

Similarly, we have

$$\begin{aligned}
\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}P_{[h,g]}(W_{[g]}\hat{\gamma}_{\tilde{e}})\right)^2 &\leq \frac{\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2}{K}\sum_{g\in[G]}\left\|\sum_{h\in[G],h\neq g}P_{[g,h]}\tilde{V}_{[h]}\right\|_2^2 \\
&= o_P(1),
\end{aligned}$$

since

$$\begin{aligned}
\mathbb{E}\frac{1}{K}\sum_{g\in[G]}\left\|\sum_{h\in[G],h\neq g}P_{[g,h]}\tilde{V}_{[h]}\right\|_2^2 &= \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left\|P_{[g,h]}\tilde{V}_{[h]}\right\|_2^2 \\
&\leq \frac{C}{K}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}(P_{[g,h]}P_{[h,g]})
\end{aligned}$$

$$= O(1).$$

Finally, we have

$$\begin{aligned}
&\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}(W_{[h]}\hat{\gamma}_{\tilde{V}})^\top P_{[h,g]}(W_{[g]}\hat{\gamma}_{\tilde{e}})\right)^2 \\
&= \frac{1}{K}\sum_{g\in[G]}\left((W_{[g]}\hat{\gamma}_{\tilde{V}})^\top P_{[g,g]}(W_{[g]}\hat{\gamma}_{\tilde{e}})\right)^2 \\
&\leq \frac{C\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{V}}\right\|_2^2\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2}{K}\sum_{g\in[G]}\mathrm{trace}(P_{[g,g]}) \\
&= o_P(1).
\end{aligned}$$

Combining the above results with the triangle inequality, we have

$$\left(\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}V_{[h]}^\top P_{[h,g]}e_{[g]}\right)^2\right)^{1/2} = \left(\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2\right)^{1/2} + o_P(1).$$

In addition, we have

$$\mathbb{E}\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}P_{[h,g]}\tilde{e}_{[g]}\right)^2 = \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}P_{[h,g]}\tilde{e}_{[g]}\right)^2 \leq C||P-\bar{P}||_F^2/K = O(1),$$

which implies

$$\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}P_{[h,g]}\tilde{e}_{[g]}\right)^2 = O_P(1).$$

Therefore, we have

$$\begin{aligned}
&\left|\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}V_{[h]}^\top P_{[h,g]}e_{[g]}\right)^2 - \frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2\right| \\
&= \left|\left(\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}V_{[h]}^\top P_{[h,g]}e_{[g]}\right)^2\right)^{1/2} - \left(\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2\right)^{1/2}\right|
\end{aligned}$$

$$\times \left| \left(\frac{1}{K} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} V_{[h]}^\top P_{[h,g]} e_{[g]} \right)^2 \right)^{1/2} + \left(\frac{1}{K} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 \right)^{1/2} \right| = o_P(1).$$

Next, we show that

$$\frac{1}{K} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 - \frac{1}{K} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 = o_P(1). \quad \text{(C.2)}$$

By Markov inequality, it suffices to show that the RHS of the following display is $o(1)$

$$\begin{aligned}
&\mathbb{E} \left(\frac{1}{K} \left(\sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 - \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 \right) \right)^2 \\
&= \frac{1}{K^2} \mathbb{V} \left(\sum_{g,h,k\in[G]^3,h\neq g,k\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \tilde{V}_{[k]}^\top P_{[k,g]} \tilde{e}_{[g]} \right) \\
&\leqslant \frac{C}{K^2} \mathbb{V} \left(\sum_{g,h\in[G]^2,g\neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 \right) + \frac{C}{K^2} \mathbb{V} \left(\sum_{g,h,k\in[G]^3,g\neq h\neq k} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \tilde{V}_{[k]}^\top P_{[k,g]} \tilde{e}_{[g]} \right).
\end{aligned} \quad \text{(C.3)}$$

For the first term on the RHS of (C.3), we have

$$\begin{aligned}
&\frac{1}{K^2} \mathbb{V} \left(\sum_{g,h\in[G]^2,g\neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 \right) \\
&= \frac{1}{K^2} \mathbb{E} \left(\sum_{g,h\in[G]^2,g\neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 - \mathbb{E} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 \right)^2 \\
&\leqslant \frac{C}{K^2} \mathbb{E} \left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace} \left(P_{[g,h]} \left(\tilde{V}_{[h]} \tilde{V}_{[h]}^\top - \Omega_h^{\tilde{V},\tilde{V}} \right) P_{[h,g]} \left(\tilde{e}_{[g]} \tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}} \right) \right) \right)^2 \\
&+ \frac{C}{K^2} \mathbb{E} \left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace} \left(P_{[g,h]} \Omega_h^{\tilde{V},\tilde{V}} P_{[h,g]} \left(\tilde{e}_{[g]} \tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}} \right) \right) \right)^2 \\
&+ \frac{C}{K^2} \mathbb{E} \left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace} \left(\left(\tilde{V}_{[h]} \tilde{V}_{[h]}^\top - \Omega_h^{\tilde{V},\tilde{V}} \right) P_{[h,g]} \Omega_g^{\tilde{e},\tilde{e}} P_{[g,h]} \right) \right)^2.
\end{aligned}$$

In addition, we have

$$
\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}\left(\tilde{V}_{[h]}\tilde{V}_{[h]}^\top-\Omega_h^{\tilde{V},\tilde{V}}\right)P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)\right)\right)^2\\
\leqslant&\frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\operatorname{trace}\left(P_{[g,h]}\left(\tilde{V}_{[h]}\tilde{V}_{[h]}^\top-\Omega_h^{\tilde{V},\tilde{V}}\right)P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)\right)\right)^2\\
\leqslant&\frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}\mathbb{E}\left(\tilde{V}_{[h]}\tilde{V}_{[h]}^\top-\Omega_h^{\tilde{V},\tilde{V}}\right)^2P_{[h,g]}\right)\\
&\qquad\qquad\times\operatorname{trace}\left(P_{[h,g]}\mathbb{E}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)^2P_{[g,h]}\right)\\
\leqslant&\frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right)\\
\leqslant&\frac{C}{K^2}\operatorname{trace}\left((P-\bar{P})^2\right)\\
=&\,o(1),
\end{aligned}
$$

where the second inequality is by the trace Cauchy-Schwartz inequality (e.g., Magnus and Neudecker (2019)). Similarly, we have

$$
\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)\right)\right)^2\\
=&\frac{1}{K^2}\mathbb{E}\left(\sum_{g\in[G]}\operatorname{trace}\left(\left(\sum_{h\neq g}P_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\right)\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)\right)\right)^2\\
=&\frac{1}{K^2}\sum_{g\in[G]}\mathbb{E}\left(\operatorname{trace}\left(\left(\sum_{h\neq g}P_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\right)\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)\right)\right)^2\\
\leqslant&\frac{1}{K^2}\sum_{g\in[G]}\operatorname{trace}\left(\left(\sum_{h\neq g}P_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\right)^2\right)\operatorname{trace}\left(\mathbb{E}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)^2\right)\\
\leqslant&\frac{C}{K^2}\sum_{g\in[G]}\operatorname{trace}\left(\sum_{h\neq g}P_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\right)\\
\leqslant&\frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right)\\
=&\,o(1),
\end{aligned}
$$

where the first inequality is by the trace Cauchy-Schwartz inequality, and, following the same argument,

$$\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(\left(\tilde{V}_{[h]}\tilde{V}_{[h]}^\top-\Omega_h^{\tilde{V},\tilde{V}}\right)P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right)^2=o(1).$$

Combining these bounds with (C.3), we have

$$\frac{1}{K^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2\right)=o(1). \tag{C.4}$$

Now consider the second term on the RHS of (C.3). We have

$$\begin{aligned}
&\frac{1}{K^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\tilde{V}_{[k]}^\top P_{[k,g]}\tilde{e}_{[g]}\right)\\
&=\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\tilde{V}_{[k]}^\top P_{[k,g]}\tilde{e}_{[g]}\right)^2\\
&\leqslant\frac{C}{K^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)P_{[g,k]}\tilde{V}_{[k]}\right)^2\\
&+\frac{C}{K^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}\right)^2,
\end{aligned}$$

where

$$\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)P_{[g,k]}\tilde{V}_{[k]}\right)^2\\
&\leqslant\frac{C}{K^2}\sum_{g,h,k\in[G]^3,g\neq h\neq k}\mathbb{E}\left(\tilde{V}_{[h]}^\top P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top-\Omega_g^{\tilde{e},\tilde{e}}\right)P_{[g,k]}\tilde{V}_{[k]}\right)^2\\
&\leqslant\frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right)\\
&=o(1),
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}\right)^2\\
&=\frac{1}{K^2}\mathbb{E}\left(\sum_{h,k\in[G]^2,h\neq k}\tilde{V}_{[h]}^\top\left(\sum_{g\neq h\neq k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right)^2\\
&\leqslant\frac{C}{K^2}\sum_{h,k\in[G]^2,h\neq k}\mathbb{E}\left(\tilde{V}_{[h]}^\top\left(\sum_{g\neq h\neq k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right)^2\\
&\leqslant\frac{C}{K^2}\sum_{h,k\in[G]^2,h\neq k}\operatorname{trace}\left(\left((P-\bar{P})\Omega_{\tilde{e}}(P-\bar{P})\right)_{[h,k]}\left((P-\bar{P})\Omega_{\tilde{e}}(P-\bar{P})\right)_{[k,h]}\right)\\
&\leqslant\frac{C}{K^2}\operatorname{trace}\left((P-\bar{P})\Omega_{\tilde{e}}(P-\bar{P})(P-\bar{P})\Omega_{\tilde{e}}(P-\bar{P})\right)\\
&=o(1).
\end{aligned}
$$

This implies that

$$
\frac{1}{K^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\tilde{V}_{[k]}^\top P_{[k,g]}\tilde{e}_{[g]}\right)=o(1). \tag{C.5}
$$

Combining (C.3)–(C.5), we have established (C.2), which further implies the desired result that

$$
\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}V_{[h]}^\top P_{[h,g]}e_{[g]}\right)^2=\frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2+o_P(1).
$$

Note that, by Lemma A.2 and the fact that $W^\top Q=0$, we can show

$$
\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}V_{[h]}^\top Q_{[h,g]}e_{[g]}\right)^2=\frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2+o_P(1)
$$

in the same manner by replacing P by Q.

Next, we note that

$$
\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(V_{[h]}^\top P_{[h,g]}e_{[g]}\right)\left(V_{[g]}^\top P_{[g,h]}e_{[h]}\right)
$$

$$= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\underbrace{\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]}}_{U_{gh}^{(1)}} - \underbrace{(W_{[h]} \hat{\gamma}_{\tilde{V}})^\top P_{[h,g]} \tilde{e}_{[g]}}_{U_{gh}^{(2)}} - \underbrace{\tilde{V}_{[h]}^\top P_{[h,g]} (W_{[g]} \hat{\gamma}_{\tilde{e}})}_{U_{gh}^{(3)}} + \underbrace{(W_{[h]} \hat{\gamma}_{\tilde{V}})^\top P_{[h,g]} (W_{[g]} \hat{\gamma}_{\tilde{e}})}_{U_{gh}^{(4)}} \right)$$
$$\times \left(\underbrace{\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}}_{U_{hg}^{(1)}} - \underbrace{(W_{[g]} \hat{\gamma}_{\tilde{V}})^\top P_{[g,h]} \tilde{e}_{[h]}}_{U_{hg}^{(2)}} - \underbrace{\tilde{V}_{[g]}^\top P_{[g,h]} (W_{[h]} \hat{\gamma}_{\tilde{e}})}_{U_{hg}^{(3)}} + \underbrace{(W_{[g]} \hat{\gamma}_{\tilde{V}})^\top P_{[g,h]} (W_{[h]} \hat{\gamma}_{\tilde{e}})}_{U_{hg}^{(4)}} \right),$$

and by using a similar argument as in the proof for (C.1), we have

$$\begin{aligned}
\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(1),2} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{hg}^{(1),2} = O_P(1), \\
\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(2),2} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{hg}^{(2),2} = o_P(1), \\
\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(3),2} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{hg}^{(3),2} = o_P(1), \\
\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(4),2} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{hg}^{(4),2} = o_P(1).
\end{aligned}$$

Then by repeatedly applying Cauchy-Schwarz inequality, we have

$$\left| \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(s_1)} U_{hg}^{(s_2)} \right| \leqslant \left(\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(s_1),2} \right)^{1/2} \left(\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} U_{gh}^{(s_2),2} \right)^{1/2} = o_P(1)$$

for $s_1 \neq 1$ or $s_2 \neq 1$, whence

$$\begin{aligned}
&\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(V_{[h]}^\top P_{[h,g]} e_{[g]} \right) \left(V_{[g]}^\top P_{[g,h]} e_{[h]} \right) \\
&= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + o_P(1).
\end{aligned}$$

Next, we note that

$$\begin{aligned}
&\mathbb{V} \left(\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \right) \\
&= \frac{1}{K^2} \mathbb{E} \left(\sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\Bigg)^2 \\
&\leqslant \frac{C}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&+ \frac{C}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&+ \frac{C}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)P_{[h,g]}\Omega_g^{\tilde{e},\tilde{V}}P_{[g,h]}\right)\right)^2 \\
&= o(1),
\end{aligned}
$$

where the last equality holds because

$$
\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&\leqslant \frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\operatorname{trace}\left(P_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&\leqslant \frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}\mathbb{E}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)\left(\tilde{V}_{[h]}\tilde{e}_{[h]}^\top - \Omega_h^{\tilde{V},\tilde{e}}\right)P_{[h,g]}\right) \\
&\qquad\qquad \times \operatorname{trace}\left(P_{[h,g]}\mathbb{E}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\left(\tilde{V}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{V},\tilde{e}}\right)P_{[g,h]}\right) \\
&\leqslant \frac{C}{K^2}\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right) \\
&= o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&= \frac{1}{K^2}\mathbb{E}\left(\sum_{g\in[G]} \operatorname{trace}\left(\left(\sum_{h\neq g} P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\right)\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2 \\
&= \frac{1}{K^2}\sum_{g\in[G]} \mathbb{E}\left(\operatorname{trace}\left(\left(\sum_{h\neq g} P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\right)\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\right)\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \frac{1}{K^2} \sum_{g\in[G]} \operatorname{trace}\left(\left(\sum_{h\neq g} P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}} P_{[h,g]}\right)\left(\sum_{k\neq g} P_{[g,k]}\Omega_k^{\tilde{V},\tilde{e}} P_{[k,g]}\right)\right) \\
&\qquad\qquad \times \operatorname{trace}\left(\mathbb{E}\left(\tilde{e}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{V}}\right)\left(\tilde{V}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{V},\tilde{e}}\right)\right) \\
&\leqslant \frac{C}{K^2} \sum_{g,h,k\in[G]^2, h\neq g, k\neq g} \operatorname{trace}\left(P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}} P_{[h,g]}P_{[g,k]}\Omega_k^{\tilde{V},\tilde{e}} P_{[k,g]}\right) \\
&\leqslant \frac{C}{K^2} \sum_{g,h,k\in[G]^2, h\neq g, k\neq g} \left(\operatorname{trace}\left(P_{[k,g]}P_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}\Omega_h^{\tilde{V},\tilde{e}} P_{[h,g]}P_{[g,k]}\right)\right)^{1/2} \\
&\qquad\qquad \times \left(\operatorname{trace}\left(P_{[h,g]}P_{[g,k]}\Omega_k^{\tilde{V},\tilde{e}}\Omega_k^{\tilde{e},\tilde{V}} P_{[k,g]}P_{[g,h]}\right)\right)^{1/2} \\
&\leqslant \frac{C}{K^2} \sum_{g,h,k\in[G]^2, h\neq g, k\neq g} \operatorname{trace}\left(P_{[k,g]}P_{[g,h]}P_{[h,g]}P_{[g,k]}\right) \\
&\leqslant \frac{C}{K^2} \sum_{g,h\in[G]^2, g\neq h} \operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right) \\
&= o(1),
\end{aligned}
$$

and, following the same argument,

$$
\frac{1}{K^2}\mathbb{E}\left(\sum_{g,h\in[G]^2, g\neq h} \operatorname{trace}\left(\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right) P_{[h,g]}\Omega_g^{\tilde{e},\tilde{V}} P_{[g,h]}\right)\right)^2 = o(1).
$$

This implies that

$$
\begin{aligned}
&\frac{1}{K} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right) \\
&= \frac{1}{K} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right) + o_P(1).
\end{aligned}
$$

Note also that we can replace P by Q, as in the proof for (C.1) and (C.2), and this concludes the proof for the first result.

For the second result, we focus on the case when $\tilde{u} = \tilde{V}$ and $\tilde{v} = \tilde{e}$. Recall that

$$
Q = M_W(P - \bar{P})M_W = P - \bar{P} + P_W\bar{P} + \bar{P}P_W - P_W\bar{P}P_W,
$$

which implies that

$$Q_{[h,g]} = P_{[h,g]} + P_{W,[h,g]}P_{[g,g]} + P_{[h,h]}P_{W,[h,g]} - \sum_{k\in[G]} P_{W,[h,k]}P_{[k,k]}P_{W,[k,g]}, \quad g \neq h.$$

Therefore, we have

$$\begin{aligned}
&\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&= \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top \left(P_{[h,g]} + P_{W,[h,g]}P_{[g,g]} + P_{[h,h]}P_{W,[h,g]} - \sum_{k\in[G]} P_{W,[h,k]}P_{[k,k]}P_{W,[k,g]}\right)\tilde{e}_{[g]}\right)^2,
\end{aligned}$$

where

$$\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top P_{W,[h,g]}P_{[g,g]}\tilde{e}_{[g]}\right)^2 \leqslant \frac{C}{K}\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(P_{W,[g,h]}P_{W,[h,g]}\right) = o(1),$$

since d_w is fixed, and

$$\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,g]}P_{W,[g,h]}\tilde{e}_{[g]}\right)^2 = o(1)$$

by the same argument as above, and

$$\begin{aligned}
&\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{V}_{[h]}^\top \left(\sum_{k\in[G]} P_{W,[h,k]}P_{[k,k]}P_{W,[k,g]}\right)\tilde{e}_{[g]}\right)^2 \\
&\leqslant \frac{C}{K}\sum_{g,h\in[G]^2,g\neq h} \operatorname{trace}\left(\left(P_W\bar{P}P_W\right)_{[g,h]}\left(P_W\bar{P}P_W\right)_{[h,g]}\right) \\
&\leqslant \frac{C}{K}\operatorname{trace}\left(P_W\bar{P}P_WP_W\bar{P}P_W\right) \\
&= o(1).
\end{aligned}$$

It follows that

$$\frac{1}{K}\sum_{g\in[G]} \mathbb{E}\left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 = \frac{1}{K}\sum_{g\in[G]} \mathbb{E}\left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)^2 + o(1).$$

Similarly, we can show that

$$\begin{aligned}&\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\\=&\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)+o(1).\end{aligned}$$

This concludes the proof. □

C.5 Proof of Lemma A.5

We prove each result in turn. To begin with, we note that

$$\Sigma=\mathbb{V}\left(\sum_{g\in[G]}\hat{\Pi}_{[g]}^\top\tilde{e}_{[g]}\right)+\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right),$$

where

$$\mathbb{V}\left(\sum_{g\in[G]}\hat{\Pi}_{[g]}^\top\tilde{e}_{[g]}\right)=\mathbb{E}\left(\hat{\Pi}^\top\tilde{e}\right)^2\geqslant\frac{1}{C}\hat{\Pi}^\top\hat{\Pi}\geqslant\frac{1}{C}\Pi^\top\Pi,$$

and

$$\begin{aligned}&\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2\\=&\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2+\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\\=&\frac{1}{2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left[\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)+\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\right]^2\\=&\frac{1}{2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left[\begin{pmatrix}\tilde{e}_{[g]} & \tilde{V}_{[g]}\end{pmatrix}\begin{pmatrix}0_{n_g\times n_h} & P_{[g,h]}\\ P_{[g,h]} & 0_{n_g\times n_h}\end{pmatrix}\begin{pmatrix}\tilde{e}_{[h]}\\ \tilde{V}_{[h]}\end{pmatrix}\right]^2\\\geqslant&\frac{1}{2C}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left[\begin{pmatrix}0_{n_g\times n_h} & P_{[g,h]}\\ P_{[g,h]} & 0_{n_g\times n_h}\end{pmatrix}\begin{pmatrix}0_{n_h\times n_g} & P_{[h,g]}\\ P_{[h,g]} & 0_{n_h\times n_g}\end{pmatrix}\right]\\=&\frac{1}{C}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left[P_{[g,h]}P_{[h,g]}\right]\\=&\frac{1}{C}\sum_{g\in[G]}\operatorname{trace}\left[P_{[g,g]}-P_{[g,g]}^2\right]\end{aligned}$$

$$\geqslant \frac{1}{C}\sum_{g\in[G]}\left(1-\lambda_{\max}(P_{[g,g]})\right)\operatorname{trace}\left[P_{[g,g]}\right]$$
$$\geqslant \frac{1}{C}K.$$

These two lower bounds imply the desired result that

$$\Sigma \geq (\Pi^\top \Pi + K)/C.$$

Next, we note that

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&+ \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&+ \frac{2}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{\Pi}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right). \qquad \text{(C.6)}
\end{aligned}$$

For the first term on the RHS of (C.6), we have

$$\mathbb{E}\frac{1}{\Sigma}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 - \sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2\right) = 0,$$

and by Assumption 3

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2\right) \\
&= \frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^\top\tilde{e}_{[g]}\right)^2\right) \\
&\leqslant \frac{1}{\Sigma^2}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^\top\tilde{e}_{[g]}\right)^4 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^\top\bar{\Pi}_{[g]}\right)^2
\end{aligned}$$

$$\leqslant \frac{C \max_{g\in[G]} \left\|\bar{\Pi}_{[g]}\right\|_2^2 \bar{\Pi}^\top \bar{\Pi}}{\left(\Pi^\top \Pi + K\right)^2}$$
$$= o(1).$$

Therefore, we have

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 = \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + o_P(1). \quad \text{(C.7)}$$

For the second term on the RHS of (C.6), by Lemma A.4 and the fact that $K/\Sigma = O(1)$, we have

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 = \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + o_P(1). \quad \text{(C.8)}$$

For the last term on the RHS of (C.6), we have

$$\mathbb{E}\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{k\in[G],k\neq g}\tilde{\Pi}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right)\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right) = 0,$$

and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{k\in[G],k\neq g}\tilde{\Pi}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right)\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right)\\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top \tilde{e}_{[g]}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2\\
&\leqslant \frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}}\right)Q_{[g,h]}\tilde{V}_{[h]}\right)^2\\
&+ \frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}Q_{[g,h]}\tilde{V}_{[h]}\right)^2,
\end{aligned}$$

where

$$\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}}\right)Q_{[g,h]}\tilde{V}_{[h]}\right)^2$$

$$
\begin{aligned}
&\leqslant \frac{C}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\bar{\Pi}_{[g]}^\top \left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}}\right) Q_{[g,h]}\tilde{V}_{[h]}\right)^2 \\
&\leqslant \frac{C \max_{g\in[G]} \left\|\bar{\Pi}_{[g]}\right\|_2^2}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \operatorname{trace}\left(Q_{[g,h]}Q_{[h,g]}\right) \\
&= o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2, g\neq h} \bar{\Pi}_{[g]}^\top \Omega_g^{\tilde{e},\tilde{e}} Q_{[g,h]}\tilde{V}_{[h]}\right)^2 \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\bar{\Pi}^\top \Omega_{\tilde{e}}(Q-\bar{Q})\tilde{V}\right)^2 \\
&\leqslant \frac{C\bar{\Pi}^\top\bar{\Pi}}{\left(\Pi^\top\Pi + K\right)^2} \\
&= o(1).
\end{aligned}
$$

Therefore, we have

$$
\frac{2}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{\Pi}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right) = o_P(1). \tag{C.9}
$$

Combining (C.6)–(C.9), we have the desired result that

$$
\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + o_P(1) \\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + o_P(1) \\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + o_P(1).
\end{aligned}
$$

Next, we note that

$$\frac{1}{\Sigma}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)^{2}-\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2}\right)$$

$$=\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)^{2} \tag{C.10}$$

$$-\frac{2}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right). \tag{C.11}$$

For the first term on the RHS of (C.10), we have

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)^{2}\\
&\leqslant\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_{2}^{2}\times\frac{1}{\Sigma}\sum_{g\in[G]}\left\|\sum_{h\in[G],h\neq g}Q_{[g,h]}\tilde{X}_{[h]}\right\|_{2}^{2}\\
&\leqslant\max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_{2}^{2}\times\frac{C}{\Sigma}\sum_{g\in[G]}\left(\left\|\bar{\Pi}_{[g]}\right\|_{2}^{2}+\left\|\sum_{h\in[G],h\neq g}Q_{[g,h]}\tilde{V}_{[h]}\right\|_{2}^{2}\right)\\
&=o_{P}(1),
\end{aligned}$$

by Lemma A.1. For the second term on the RHS of (C.10), we have

$$\begin{aligned}
&\left|\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)\right|\\
&\leqslant\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)^{2}\right)^{1/2}\times\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)^{2}\right)^{1/2}\\
&=o_{P}(1).
\end{aligned}$$

Combining the two bounds above, we have the desired result that

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)^{2}=\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2}+o_{P}(1).$$

Next, we note that

$$
\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)^{2}\\
&=\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}X_{[g]}\right)^{2}}_{R_1}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)^{2}}_{R_2}\\
&+2\times\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^{\top}X_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}X_{[g]}}_{R_3}. \qquad \text{(C.12)}
\end{aligned}
$$

For R_1, we have

$$
\begin{aligned}
R_1 &= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^{2}+\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}V_{[g]}\right)^{2}+\frac{2}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}V_{[g]}\right)\\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^{2}+\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)^{2}+\frac{2}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)+o_P(1)\\
&= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^{2}+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)^{2}+o_P(1),
\end{aligned}
$$

where the second equality holds by using $V_{[g]}=\tilde{V}_{[g]}-W_{[g]}\hat{\gamma}_{\tilde{V}}$ and Lemma A.1, and the last equality holds because

$$
\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)^{2}\right)\\
&\leq\frac{1}{\Sigma^2}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)^{4}\\
&\leq\frac{C}{\Sigma^2}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\bar{\Pi}_{[g]}\right)^{2}\\
&\leq\frac{C\max_{g\in[G]}\left\|\bar{\Pi}_{[g]}\right\|_2^2\bar{\Pi}^{\top}\bar{\Pi}}{(\Pi^{\top}\Pi+K)^2}\\
&=o(1),
\end{aligned}
$$

and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\right)\\
&=\frac{1}{\Sigma^2}\sum_{g\in[G]}\mathbb{E}\left(\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\right)^2\\
&\leq\frac{C}{\Sigma^2}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^2\left(\bar{\Pi}_{[g]}^{\top}\bar{\Pi}_{[g]}\right)\\
&\leq\frac{C\max_{g\in[G]}\left\|\bar{\Pi}_{[g]}\right\|_2^2\bar{\Pi}^{\top}\bar{\Pi}}{(\Pi^{\top}\Pi+K)^2}\\
&=o(1).
\end{aligned}$$

For R_2, we have

$$\begin{aligned}
R_2=&\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)^2}_{R_{2,1}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}V_{[g]}\right)^2}_{R_{2,2}}\\
&+2\times\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}V_{[g]}\right)}_{R_{2,3}}.
\end{aligned}$$

For $R_{2,1}$, we have

$$\begin{aligned}
\mathbb{V}\left(R_{2,1}\right)\leqslant&\frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)^2\right)\\
&+\frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\tilde{V}_{[k]}^{\top}Q_{[k,g]}\Pi_{[g]}\right)\right),
\end{aligned}$$

where

$$\frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)^2\right)$$

$$
\begin{aligned}
&= \frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{h\in[G]}\tilde{V}_{[h]}^\top\left(\sum_{g\in[G],g\neq h}Q_{[h,g]}\Pi_{[g]}\Pi_{[g]}^\top Q_{[g,h]}\right)\tilde{V}_{[h]}\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{h\in[G]}\left\|\sum_{g\in[G],g\neq h}Q_{[h,g]}\Pi_{[g]}\Pi_{[g]}^\top Q_{[g,h]}\right\|_{op}^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{h\in[G]}\operatorname{trace}\left(\sum_{g\in[G],g\neq h}Q_{[h,g]}\Pi_{[g]}\Pi_{[g]}^\top Q_{[g,h]}\right) \\
&\leq \frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right) \\
&= o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)\left(\tilde{V}_{[k]}^\top Q_{[k,g]}\Pi_{[g]}\right)\right) \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)\left(\tilde{V}_{[k]}^\top Q_{[k,g]}\Pi_{[g]}\right)\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g,h,k\in[G]^3,g\neq h\neq k}\mathbb{E}\left(\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)\left(\tilde{V}_{[k]}^\top Q_{[k,g]}\Pi_{[g]}\right)\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g,h,k\in[G]^3,g\neq h\neq k}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right)\operatorname{trace}\left(Q_{[k,g]}Q_{[g,k]}\right) \\
&= o(1).
\end{aligned}
$$

This implies that

$$
R_{2,1} = \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)^2 + o_P(1).
$$

For $R_{2,2}$, since $K/\Sigma = O(1)$, by Lemma A.6, we have

$$
R_{2,2} = \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)^2 + o_P(1).
$$

For $R_{2,3}$, we have

$$\begin{aligned} R_{2,3} &= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{V}_{[g]}\right)+o_P(1) \\ &= o_P(1), \end{aligned}$$

where the second equality holds because

$$\mathbb{E}\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{V}_{[g]}\right)=0,$$

$$\begin{aligned} &\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{V}_{[g]}\right)\right) \\ &\leq \frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right) \\ &+\frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{V}_{[g]}\right), \end{aligned}$$

where

$$\begin{aligned} &\frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right) \\ &\leq \frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\Pi_{[g]}^{\top}Q_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}Q_{[h,g]}\tilde{V}_{[g]}\right)^2 \\ &+\frac{C}{\Sigma^2}\mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\Pi_{[g]}^{\top}Q_{[g,h]}(\tilde{V}_{[h]}\tilde{V}_{[h]}^{\top}-\Omega_h^{\tilde{V},\tilde{V}})Q_{[h,g]}\tilde{V}_{[g]}\right)^2 \\ &\leq \frac{C}{\Sigma^2}\sum_{g\in[G]}\left\|\left(\sum_{h\in[G],h\neq g}Q_{[g,h]}\Omega_h^{\tilde{V},\tilde{V}}Q_{[h,g]}\right)\Pi_{[g]}\right\|_2^2 \\ &+\frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)^2 \\ &\leq \frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right) \end{aligned}$$

$$= o(1),$$

and

$$\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{V}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[k]}^\top Q_{[k,g]}\tilde{V}_{[g]}\right)\\
&=\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h,k\in[G]^3,g\neq h\neq k}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[k]}^\top Q_{[k,g]}\tilde{V}_{[g]}\right)^2\\
&\leq\frac{C}{\Sigma^2}\sum_{g,h,k\in[G]^3,g\neq h\neq k}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\tilde{V}_{[k]}^\top Q_{[k,g]}\tilde{V}_{[g]}\right)^2\\
&\leq\frac{C}{\Sigma^2}\sum_{g,h,k\in[G]^3,g\neq h\neq k}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right)\operatorname{trace}\left(Q_{[k,g]}Q_{[g,k]}\right)\\
&=o(1).
\end{aligned}$$

Combining the results above, we have

$$R_2=\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)^2+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)^2+o_P(1).$$

For R_3, we have

$$\begin{aligned}
R_3=&\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top\Pi_{[g]}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}}_{R_{3,1}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top\Pi_{[g]}\tilde{V}_{[h]}^\top Q_{[h,g]}V_{[g]}}_{R_{3,2}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top V_{[g]}\tilde{V}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}}_{R_{3,3}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^\top V_{[g]}\tilde{V}_{[h]}^\top Q_{[h,g]}V_{[g]}}_{R_{3,4}}.
\end{aligned}$$

For $R_{3,1}$, it has mean zero and

$$\begin{aligned}
\mathbb{V}\left(R_{3,1}\right) &= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{h\in[G]}\left\|\sum_{g\in[G],g\neq h}Q_{[h,g]}\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right\|_2^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{h\in[G]}\left\|\sum_{g\in[G]}Q_{[h,g]}\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right\|_2^2 + \frac{C}{\Sigma^2}\sum_{h\in[G]}\left\|Q_{[h,h]}\Pi_{[h]}\bar{\Pi}_{[h]}^{\top}\Pi_{[h]}\right\|_2^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{h\in[G]}\left\|\sum_{g\in[G]}Q_{[h,g]}\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right\|_2^2 + o(1) \\
&= o(1),
\end{aligned}$$

where the last equality holds because

$$\begin{aligned}
&\frac{1}{\Sigma^2}\sum_{h\in[G]}\left\|\sum_{g\in[G]}Q_{[h,g]}\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right\|_2^2 \\
&= \frac{1}{\Sigma^2}\sum_{g,g'\in[G]^2}\left(\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^{\top}\left(\sum_{h\in[G]}Q_{[g,h]}Q_{[h,g']}\right)\Pi_{[g']}\bar{\Pi}_{[g']}^{\top}\Pi_{[g']} \\
&\leq \frac{1}{\Sigma^2}\sum_{g\in[G]}\left\|\Pi_{[g]}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right\|_2^2 \\
&= o(1).
\end{aligned}$$

For $R_{3,2}$, we have

$$R_{3,2} = \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]} + o_P(1),$$

where the first term has mean zero and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right) \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{C}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\bar{\Pi}_{[g]}^\top \Pi_{[g]} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]}\right)^2 \\
&\leq \frac{C \max_{g\in[G]} \left\|\bar{\Pi}_{[g]}\right\|_2^2}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \operatorname{trace}\left(Q_{[h,g]} Q_{[g,h]}\right) \\
&= o(1).
\end{aligned}
$$

By using the same argument, we also have $R_{3,3} = o_P(1)$. For $R_{3,4}$, we have

$$
R_{3,4} = \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} + o_P(1),
$$

where the first term has mean zero and

$$
\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]}\right) \\
&\leq \frac{C}{\Sigma^2} \mathbb{V}\left(\sum_{g,h\in[G]^2, g\neq h} \bar{\Pi}_{[g]}^\top \Omega_g^{\tilde{V},\tilde{V}} Q_{[g,h]} \tilde{V}_{[h]}\right) \\
&+ \frac{C}{\Sigma^2} \mathbb{V}\left(\sum_{g,h\in[G]^2, g\neq h} \bar{\Pi}_{[g]}^\top (\tilde{V}_{[g]} \tilde{V}_{[g]}^\top - \Omega_g^{\tilde{V},\tilde{V}}) Q_{[g,h]} \tilde{V}_{[h]}\right) \\
&\leq \frac{C}{\Sigma^2} \mathbb{E}\left(\bar{\Pi}^\top \Omega_{\tilde{V}} (Q - \bar{Q}) \tilde{V}\right)^2 \\
&+ \frac{C}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]}\right)^2 \\
&\leq \frac{C \bar{\Pi}^\top \bar{\Pi}}{\Sigma^2} + \frac{C \max_{g\in[G]} \left\|\bar{\Pi}_{[g]}\right\|_2^2}{\Sigma^2} \sum_{g,h\in[G]^2, g\neq h} \operatorname{trace}\left(Q_{[h,g]} Q_{[g,h]}\right) \\
&= o(1).
\end{aligned}
$$

Combining the results above, we have $R_3 = o_P(1)$. Therefore, by combining (C.12) with the calculations about for terms R_1 to R_3, we have

$$
\begin{aligned}
\frac{1}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G], h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]}\right)^2 &= \frac{1}{\Sigma} \sum_{g\in[G]} \left(\bar{\Pi}_{[g]}^\top \Pi_{[g]}\right)^2 + \frac{1}{\Sigma} \sum_{g\in[G]} \mathbb{E}\left(\bar{\Pi}_{[g]}^\top \tilde{V}_{[g]}\right)^2 \\
&+ \frac{1}{\Sigma} \sum_{g\in[G]} \mathbb{E}\left(\sum_{h\in[G], h\neq g} \tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]}\right)^2
\end{aligned}
$$

$$+\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)^{2}+o_P(1).$$

Finally, we note that

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}e_{[g]}\right)\\
&=\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}e_{[g]}\right)}_{R_4}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}V_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}e_{[g]}\right)}_{R_5}. \qquad \text{(C.13)}
\end{aligned}$$

For R_4, we have

$$\begin{aligned}
R_4&=\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)+o_P(1)\\
&=\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{4,1}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)}_{R_{4,2}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{4,3}}\\
&+\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)}_{R_{4,4}}+o_P(1).
\end{aligned}$$

By using a similar argument as in the proof for R_1, we have $R_{4,1}=o_P(1)$. For $R_{4,2}$,

it has mean zero and

$$\begin{aligned}
\mathbb{V}\left(R_{4,2}\right) &= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\left(\bar{\Pi}_{[g]}^{\top}\Pi_{[g]}\right)^2\mathbb{E}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&\leq \frac{C\max_{g\in[G]}\left\|\bar{\Pi}_{[g]}\right\|_2^2}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right) \\
&= o(1).
\end{aligned}$$

Therefore, we have $R_{4,2} = o_P(1)$. Using the same argument, we also have $R_{4,3} = o_P(1)$. In addition, by using a similar argument as in the proof for $R_{2,3}$, we have $R_{4,4} = o_P(1)$, which implies $R_4 = o_P(1)$.

For R_5, we have

$$\begin{aligned}
R_5 &= \frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)+o_P(1) \\
&= \underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{5,1}} \\
&\quad+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)}_{R_{5,2}} \\
&\quad+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{5,3}} \\
&\quad+\underbrace{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)}_{R_{5,4}}+o_P(1).
\end{aligned}$$

For $R_{5,1}$, by using a similar argument as in the proof for R_1, we have

$$R_{5,1} = \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\bar{\Pi}_{[g]}^{\top}\tilde{V}_{[g]}\right)\left(\bar{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)+o_P(1).$$

In addition, by using a similar argument as in the proof for $R_{3,4}$, we have $R_{5,2} = o_P(1)$ and $R_{5,3} = o_P(1)$. Lastly, by using a similar argument as in the proof for $R_{2,2}$, we have

$$R_{5,4} = \frac{1}{\Sigma} \sum_{g \in [G]} \mathbb{E} \left(\sum_{h \in [G], h \neq g} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) \left(\sum_{k \in [G], k \neq g} \tilde{V}_{[k]}^\top Q_{[k,g]} \tilde{e}_{[g]} \right) + o_P(1).$$

Combining the results with (C.13), we have

$$\begin{aligned}
&\frac{1}{\Sigma} \sum_{g \in [G]} \left(\sum_{h \in [G], h \neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right) \left(\sum_{k \in [G], k \neq g} \tilde{X}_{[k]}^\top Q_{[k,g]} e_{[g]} \right) \\
&= \frac{1}{\Sigma} \sum_{g \in [G]} \mathbb{E} \left(\bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \right) \left(\bar{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \\
&+ \frac{1}{\Sigma} \sum_{g \in [G]} \mathbb{E} \left(\sum_{h \in [G], h \neq g} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) \left(\sum_{k \in [G], k \neq g} \tilde{V}_{[k]}^\top Q_{[k,g]} \tilde{e}_{[g]} \right) + o_P(1).
\end{aligned}$$

This concludes the proof. □

C.6 Proof of Lemma A.6

For the first result in Lemma A.6, we note that

$$\begin{aligned}
&\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right) \\
&= \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right)}_{R_6} \\
&+ \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right)}_{R_7} \\
&+ 2 \times \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right)}_{R_8}.
\end{aligned}$$

For R_6, it has mean zero and

$$\begin{aligned}
\mathbb{V}(R_6) &= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[g,h]}Q_{[h,g]}\right) \\
&= o(1),
\end{aligned}$$

whence $R_6 = o_P(1)$. For R_7, since $K/\Sigma = O(1)$, by an application of Lemma A.4, we have

$$R_7 = \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right) + o_P(1).$$

For R_8, we have

$$\begin{aligned}
\mathbb{V}(R_8) &= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&+ \frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&= o(1),
\end{aligned}$$

where the last equality holds because

$$\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
\leqslant\ &\frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\left(\tilde{e}_{[h]}\tilde{V}_{[h]}^\top - \Omega_h^{\tilde{e},\tilde{V}}\right)Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
\leqslant\ &\frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[g,h]}Q_{[h,g]}\right)
\end{aligned}$$

$$= o(1),$$

and

$$\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{\Pi}_{[g]}^{\top}Q_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g\in[G]}\tilde{\Pi}_{[g]}^{\top}\left(\sum_{h\neq g}Q_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}Q_{[h,g]}\right)\tilde{e}_{[g]}\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{g\in[G]}\text{trace}\left(\left(\sum_{h\neq g}Q_{[g,h]}\Omega_h^{\tilde{e},\tilde{V}}Q_{[h,g]}\right)\left(\sum_{k\neq g}Q_{[g,k]}\Omega_k^{\tilde{V},\tilde{e}}Q_{[k,g]}\right)\right) \\
&\leq \frac{C}{\Sigma^2}\text{trace}\left((Q-\bar{Q})\Omega^{\tilde{e},\tilde{V}}(Q-\bar{Q})^2\Omega^{\tilde{V},\tilde{e}}(Q-\bar{Q})\right) \\
&= o(1).
\end{aligned}$$

It follows that $R_8 = o_P(1)$. Combining the results above, we have

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right) \\
&= \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right) + o_P(1).
\end{aligned}$$

Next, we turn to the second result in Lemma A.6. We note that

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right) \\
&= \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right) \\
&+ \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}W_{[h]}\hat{\gamma}_{\tilde{e}}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right) \\
&- \frac{2}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}W_{[h]}\hat{\gamma}_{\tilde{e}}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right),
\end{aligned}$$

where

$$\left|\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}W_{[h]}\hat{\gamma}_{\tilde{e}}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\right|$$

$$
\begin{aligned}
&\leqslant \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} W_{[h]} \hat{\gamma}_{\tilde{e}}\right)^2 \\
&\leqslant \max_{g\in[G]} \left\| W_{[g]} \hat{\gamma}_{\tilde{e}} \right\|_2^2 \times \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left\| Q_{[h,g]} \tilde{X}_{[g]} \right\|_2^2 \\
&= o_P(1),
\end{aligned}
$$

by Lemma A.1, and

$$
\begin{aligned}
&\left| \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} W_{[h]} \hat{\gamma}_{\tilde{e}}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) \right| \\
&\leqslant \left(\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} W_{[h]} \hat{\gamma}_{\tilde{e}}\right)^2 \right)^{1/2} \\
&\times \left(\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2 \right)^{1/2} \\
&= o_P(1).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} e_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} e_{[g]}\right) \\
&= \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) + o_P(1) \\
&= \frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) + o_P(1).
\end{aligned}
$$

Next, we turn to the third result in Lemma A.6. Note that

$$
\begin{aligned}
&\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} X_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]}\right) \\
&= \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} V_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} V_{[g]}\right)}_{R_9} \\
&+ \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \Pi_{[h]}\right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \Pi_{[g]}\right)}_{R_{10}}
\end{aligned}
$$

$$+ 2 \times \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} V_{[g]} \right)}_{R_{11}}.$$

By using the same argument as in the proof above, we have

$$R_9 = \frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) + o_P(1).$$

For R_{10}, we have

$$\begin{aligned} R_{10} = & \frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right) \\ & + \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right)}_{R_{10,1}} \\ & + 2 \times \underbrace{\frac{1}{\Sigma} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right)}_{R_{10,2}}. \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{V}\left(R_{10,1}\right) &= \frac{1}{\Sigma^2} \mathbb{E} \left(\sum_{g,h \in [G]^2, g \neq h} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right) \right)^2 \\ &\leq \frac{C}{\Sigma^2} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\left(\tilde{V}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right) \right)^2 \\ &\leq \frac{C}{\Sigma^2} \sum_{g,h \in [G]^2, g \neq h} \operatorname{trace} \left(Q_{[h,g]} Q_{[g,h]} \right) \\ &= o(1), \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}\left(R_{10,2}\right) &= \frac{1}{\Sigma^2} \mathbb{E} \left(\sum_{g,h \in [G]^2, g \neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right) \right)^2 \\ &\leq \frac{C}{\Sigma^2} \sum_{h \in [G]} \left\| \sum_{g \in [G], g \neq h} Q_{[h,g]} \Pi_{[g]} \tilde{\Pi}_{[g]}^\top Q_{[g,h]} \Pi_{[h]} \right\|_2^2 \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{C}{\Sigma^2} \sum_{h\in[G]} \left\| \sum_{g\in[G],g\neq h} Q_{[h,g]}\Pi_{[g]}\tilde{\Pi}_{[g]}^\top Q_{[g,h]} \right\|_F^2 \\
&\leq \frac{C}{\Sigma^2} \sum_{h\in[G]} \sum_{g,g'\in[G]^2,g,g'\neq h} \operatorname{trace}\left(Q_{[h,g]}\Pi_{[g]}\tilde{\Pi}_{[g]}^\top Q_{[g,h]} Q_{[h,g']}\tilde{\Pi}_{[g']}\Pi_{[g']}^\top Q_{[g',h]} \right) \\
&\leq \frac{C}{\Sigma^2} \sum_{h\in[G]} \sum_{g,g'\in[G]^2,g,g'\neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]} Q_{[h,g']}\tilde{\Pi}_{[g']} \right)^2 \\
&+ \frac{C}{\Sigma^2} \sum_{h\in[G]} \sum_{g,g'\in[G]^2,g,g'\neq h} \left(\Pi_{[g]}^\top Q_{[g,h]} Q_{[h,g']}\Pi_{[g']} \right)^2 \\
&\leq \frac{C}{\Sigma^2} \sum_{h\in[G]} \sum_{g,g'\in[G]^2,g,g'\neq h} \operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]} \right) \operatorname{trace}\left(Q_{[h,g']}Q_{[g',h]} \right) \\
&= o(1).
\end{aligned}
$$

Therefore, we have

$$
R_{10} = \frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\Pi_{[g]} \right) + o_P(1).
$$

For R_{11}, we have

$$
\begin{aligned}
R_{11} &= \frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]} \right) + o_P(1) \\
&= \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]} \right)}_{R_{11,1}} \\
&+ \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]} \right)}_{R_{11,2}} \\
&+ \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]} \right)}_{R_{11,3}} \\
&+ \underbrace{\frac{1}{\Sigma} \sum_{g,h\in[G]^2,g\neq h} \left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]} \right)}_{R_{11,4}}.
\end{aligned}
$$

By using a similar argument as in the proof for $R_{10,1}$ and $R_{10,2}$, we have $R_{11,1} = o_P(1)$

and $R_{11,2} = o_P(1)$. For $R_{11,3}$, we have

$$\begin{aligned}
\mathbb{V}\left(R_{11,3}\right) &= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g\in[G]}\left\|\sum_{h\in[G],h\neq g} Q_{[g,h]}\Pi_{[h]}\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\right\|_F^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g\in[G]}\sum_{h,h'\in[G]^2,h,h'\neq g}\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}Q_{[g,h']}\tilde{\Pi}_{[h']}\right)^2 \\
&+ \frac{C}{\Sigma^2}\sum_{g\in[G]}\sum_{h,h'\in[G]^2,h,h'\neq g}\left(\Pi_{[h]}^\top Q_{[h,g]}Q_{[g,h']}\Pi_{[h']}\right)^2 \\
&\leq \frac{C}{\Sigma^2}\sum_{g\in[G]}\sum_{h,h'\in[G]^2,h,h'\neq g}\operatorname{trace}\left(Q_{[g,h]}Q_{[h,g]}\right)\operatorname{trace}\left(Q_{[g,h']}Q_{[h',g]}\right) \\
&= o(1),
\end{aligned}$$

and thus

$$R_{11,3} = \frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right) + o_P(1).$$

For $R_{11,4}$, we have

$$\begin{aligned}
\mathbb{V}\left(R_{11,4}\right) \leq\ & \mathbb{V}\left(\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^\top Q_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}Q_{[g,h]}\Pi_{[h]}\right) \\
&+ \mathbb{V}\left(\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^\top Q_{[h,g]}(\tilde{V}_{[g]}\tilde{V}_{[g]}^\top - \Omega_g^{\tilde{V},\tilde{V}})Q_{[g,h]}\Pi_{[h]}\right),
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^\top Q_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}Q_{[g,h]}\Pi_{[h]}\right) \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{h\in[G]}\tilde{V}_{[h]}^\top\left(\sum_{g\in[G],g\neq h}Q_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}Q_{[g,h]}\right)\Pi_{[h]}\right)^2 \\
&\leq \frac{C\Pi^\top\Pi}{(\Pi^\top\Pi+K)^2} \\
&= o(1),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^\top Q_{[h,g]}(\tilde{V}_{[g]}\tilde{V}_{[g]}^\top-\Omega_g^{\tilde{V},\tilde{V}})Q_{[g,h]}\Pi_{[h]}\right)\\
&=\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{V}_{[h]}^\top Q_{[h,g]}(\tilde{V}_{[g]}\tilde{V}_{[g]}^\top-\Omega_g^{\tilde{V},\tilde{V}})Q_{[g,h]}\Pi_{[h]}\right)^2\\
&\leq\frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)^2\\
&\leq\frac{C}{\Sigma^2}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right)\\
&=o(1).
\end{aligned}$$

It follows that $R_{11,4}=o_P(1)$, whence

$$R_{11}=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)+o_P(1).$$

Combining the results above, we have

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}X_{[g]}\right)\\
&=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\\
&+\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{\Pi}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\Pi_{[g]}\right)\\
&+\frac{2}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)+o_P(1).
\end{aligned}$$

Finally, for the last result of Lemma A.6, we note that

$$\begin{aligned}
&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)\\
&=\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}V_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)}_{R_{12}}
\end{aligned}$$

$$+\underbrace{\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)}_{R_{13}}.$$

By using the same argument as in the proof for R_9, we have

$$R_{12}=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)+o_P(1).$$

In addition, by using the same argument as in the proof for R_{11}, we have

$$R_{13}=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)+o_P(1).$$

It follows that

$$\begin{aligned}&\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)\\&=\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\\&+\frac{1}{\Sigma}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)+o_P(1).\end{aligned}$$

This concludes the proof. □

C.7 Proof of Lemma A.7

Throughout the proof we denote $\acute{\Delta}=\acute{\beta}-\beta$, and note that $\acute{\Delta}=o_P(1)$ by assumption. We divide the proof into four steps.

Step 1: Consistency of $\acute{\Omega}$. Note that $1/C\leq\lambda_{\min}(\Omega/n)\leq\lambda_{\min}(\Omega/n)\leq C$ by Assumption 1, and thus it suffices to show that

$$\frac{1}{n}\acute{\Omega}-\frac{1}{n}\Omega=o_P(1).$$

Let

$$\tilde{\Omega}=\sum_{g\in[G]}\left(z_{[g]}^\top\tilde{e}_{[g]}\right)\left(z_{[g]}^\top\tilde{e}_{[g]}\right)^\top\quad\text{and}\quad\bar{\Omega}=\sum_{g\in[G]}\left(z_{[g]}^\top e_{[g]}\right)\left(z_{[g]}^\top e_{[g]}\right)^\top.$$

We aim to show that

$$\frac{1}{n}\tilde{\Omega} - \frac{1}{n}\Omega = o_P(1), \tag{C.14}$$
$$\frac{1}{n}\bar{\Omega} - \frac{1}{n}\tilde{\Omega} = o_P(1), \tag{C.15}$$
$$\frac{1}{n}\acute{\Omega} - \frac{1}{n}\bar{\Omega} = o_P(1). \tag{C.16}$$

For (C.14), consider its (j,k)-th element for $1 \leq j,k \leq d_z$, given by

$$\frac{1}{n}\sum_{g\in[G]}\left[z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k} - \mathbb{E}\left(z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k}\right)\right],$$

where we use $z_{[g],j}$ ($z_{[g],k}$) to denote the j-th (k-th) column of $z_{[g]}$; note that it has mean zero and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{n}\sum_{g\in[G]}\left[z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k} - \mathbb{E}\left(z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k}\right)\right]\right)\\
&\leqslant \frac{1}{n^2}\sum_{g\in[G]}\mathbb{E}\left(z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k}\right)^2\\
&\leqslant \frac{1}{n^2}\sum_{g\in[G]} z_{[g],j}^\top z_{[g],j} z_{[g],k}^\top z_{[g],k}\mathbb{E}\left(\tilde{e}_{[g]}^\top \tilde{e}_{[g]}\right)^2\\
&\leqslant \frac{C\max_{i\in I_g, g\in[G]}\|z_{i,g}\|_2^2}{n} \times \frac{1}{n}\sum_{g\in[G]} z_{[g],j}^\top z_{[g],j}\\
&= o(1)
\end{aligned}$$

by Assumption 1, and the result follows since d_z is fixed.

For (C.15), its (j,k)-th element can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{g\in[G]}\left(z_{[g],j}^\top e_{[g]}e_{[g]}^\top z_{[g],k} - z_{[g],j}^\top \tilde{e}_{[g]}\tilde{e}_{[g]}^\top z_{[g],k}\right)\\
&= \frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^\top W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],k}^\top W_{[g]}\hat{\gamma}_{\tilde{e}})\\
&- \frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^\top W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],k}^\top \tilde{e}_{[g]})\\
&- \frac{1}{n}\sum_{g\in[G]}(z_{[g],k}^\top W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],j}^\top \tilde{e}_{[g]}).
\end{aligned}$$

We have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],k}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})\right| \\
&\leqslant \sqrt{\frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})^2}\times\sqrt{\frac{1}{n}\sum_{g\in[G]}(z_{[g],k}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})^2} \\
&\leqslant \max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2\times\sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],j}^{\top}z_{[g],j}}\times\sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],k}^{\top}z_{[g],k}} \\
&\leqslant \max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2\times\sqrt{v_j^{\top}\left(\frac{1}{n}\sum_{i\in I_g,g\in[G]}z_{i,g}z_{i,g}^{\top}\right)v_j}\times\sqrt{v_k^{\top}\left(\frac{1}{n}\sum_{i\in I_g,g\in[G]}z_{i,g}z_{i,g}^{\top}\right)v_k} \\
&= o_P(1)
\end{aligned}
$$

by Assumption 1 and Lemma A.1, where we use v_j (v_k) to denote the d_z-dimensional unit vector with j-th (k-th) element one and other elements zero; we also have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],k}^{\top}\tilde{e}_{[g]})\right| \\
&\leqslant \max_{g\in[G]}\left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2\times\sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],j}^{\top}z_{[g],j}}\times\sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],k}^{\top}\tilde{e}_{[g]}\tilde{e}_{[g]}^{\top}z_{[g],k}} \\
&= o_P(1),
\end{aligned}
$$

and by using the same argument,

$$
\left|\frac{1}{n}\sum_{g\in[G]}(z_{[g],k}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}})(z_{[g],j}^{\top}\tilde{e}_{[g]})\right| = o_P(1).
$$

For (C.16), its (j,k)-th element can be written as

$$
\begin{aligned}
&\frac{1}{n}\sum_{g\in[G]}\left(z_{[g],j}^{\top}\acute{e}_{[g]}\acute{e}_{[g]}^{\top}z_{[g],k}-z_{[g],j}^{\top}e_{[g]}e_{[g]}^{\top}z_{[g],k}\right) \\
&=\frac{\acute{\Delta}^2}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}X_{[g]})(z_{[g],k}^{\top}X_{[g]}) \\
&-\frac{\acute{\Delta}}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}X_{[g]})(z_{[g],k}^{\top}e_{[g]})
\end{aligned}
$$

$$-\frac{\acute{\Delta}}{n}\sum_{g\in[G]}(z_{[g],k}^{\top}X_{[g]})(z_{[g],j}^{\top}e_{[g]}),$$

and note that

$$\begin{aligned}\frac{1}{n}\sum_{g\in[G]}(z_{[g],j}^{\top}X_{[g]})(z_{[g],k}^{\top}X_{[g]}) &\leqslant \sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],j}^{\top}z_{[g],j}X_{[g]}^{\top}X_{[g]}}\times\sqrt{\frac{1}{n}\sum_{g\in[G]}z_{[g],k}^{\top}z_{[g],k}X_{[g]}^{\top}X_{[g]}}\\ &= O_P(1),\end{aligned}$$

since $\max_{g\in[G]}\mathbb{E}\left(X_{[g]}^{\top}X_{[g]}\right)=O(1)$ by Lemma A.1; the other two terms can be handled similarly.

Step 2: Consistency of $\acute{\Psi}$. By Assumption 2, we have

$$\frac{1}{r_n}z^{\top}X=\frac{1}{r_n}z^{\top}\Pi+\frac{1}{r_n}z^{\top}V=\frac{1}{r_n}z^{\top}\Pi+o_P(1), \tag{C.17}$$

and note that

$$\begin{aligned}\frac{1}{\lambda_n}\hat{A}_n&=\frac{1}{\lambda_n}A_n+\frac{1}{\lambda_n}(\hat{A}_n-A_n)\\ &=\frac{1}{\lambda_n}A_n+(\frac{1}{\lambda_n}A_n)^{1/2}(A_n^{-1/2}\hat{A}_nA_n^{-1/2}-I_{d_z})(\frac{1}{\lambda_n}A_n)^{1/2}\\ &=\frac{1}{\lambda_n}A_n+o_P(1),\end{aligned} \tag{C.18}$$

where $\lambda_n=\lambda_{\max}(A_n)$. In addition, we have

$$\frac{1}{(\frac{1}{r_n}\Pi^{\top}z)(\frac{1}{\lambda_n}A_n)(\frac{1}{n}\Omega)(\frac{1}{\lambda_n}A_n)(\frac{1}{r_n}z^{\top}\Pi)}\leq C.$$

Therefore, we have

$$\frac{\acute{\Psi}}{\Psi}=\frac{X^{\top}z\hat{A}_n\acute{\Omega}\hat{A}_nz^{\top}X}{\Pi^{\top}zA_n\Omega A_nz^{\top}\Pi}=\frac{(\frac{1}{r_n}X^{\top}z)(\frac{1}{\lambda_n}\hat{A}_n)(\frac{1}{n}\acute{\Omega})(\frac{1}{\lambda_n}\hat{A}_n)(\frac{1}{r_n}z^{\top}X)}{(\frac{1}{r_n}\Pi^{\top}z)(\frac{1}{\lambda_n}A_n)(\frac{1}{n}\Omega)(\frac{1}{\lambda_n}A_n)(\frac{1}{r_n}z^{\top}\Pi)}=1+o_P(1)$$

by Step 1.

Step 3: Consistency of $\acute{\Sigma}$. Define

$$\bar{\Sigma} = \mathbb{E}\left(\tilde{\Pi}^\top (Q - \bar{Q})\tilde{e}\right)^2 + \mathbb{E}\left(\tilde{V}^\top (Q - \bar{Q})\tilde{e}\right)^2,$$

and recall

$$\Sigma = \mathbb{E}\left(\hat{\Pi}^\top \tilde{e}\right)^2 + \mathbb{E}\left(\tilde{V}^\top (P - \bar{P})\tilde{e}\right)^2 = \mathbb{E}\left(\tilde{\Pi}^\top Q\tilde{e}\right)^2 + \mathbb{E}\left(\tilde{V}^\top (P - \bar{P})\tilde{e}\right)^2.$$

Using the notation $\|X\|_{\mathbb{P},2} = \left(\mathbb{E}\left(X^2\right)\right)^{1/2}$, we have

$$\begin{aligned}
\left|\frac{\bar{\Sigma} - \Sigma}{\Sigma}\right| &\leq \left|\frac{\left\|\tilde{\Pi}^\top Q\tilde{e}\right\|_{\mathbb{P},2}^2 - \left\|\tilde{\Pi}^\top (Q - \bar{Q})\tilde{e}\right\|_{\mathbb{P},2}^2}{\Sigma}\right| + \left|\frac{\left\|\tilde{V}^\top (P - \bar{P})\tilde{e}\right\|_{\mathbb{P},2}^2 - \left\|\tilde{V}^\top (Q - \bar{Q})\tilde{e}\right\|_{\mathbb{P},2}^2}{\Sigma}\right| \\
&\leq \left(\frac{2\left\|\tilde{\Pi}^\top Q\tilde{e}\right\|_{\mathbb{P},2}\left\|\tilde{\Pi}^\top \bar{Q}\tilde{e}\right\|_{\mathbb{P},2} + \left\|\tilde{\Pi}^\top \bar{Q}\tilde{e}\right\|_{\mathbb{P},2}^2}{\Sigma}\right) + \left|\frac{\left\|\tilde{V}^\top (P - \bar{P})\tilde{e}\right\|_{\mathbb{P},2}^2 - \left\|\tilde{V}^\top (Q - \bar{Q})\tilde{e}\right\|_{\mathbb{P},2}^2}{\Sigma}\right| \\
&\leq \left(\frac{2\left\|\tilde{\Pi}^\top Q\tilde{e}\right\|_{\mathbb{P},2}\left\|\tilde{\Pi}^\top \bar{Q}\tilde{e}\right\|_{\mathbb{P},2} + \left\|\tilde{\Pi}^\top \bar{Q}\tilde{e}\right\|_{\mathbb{P},2}^2}{\Pi^\top \Pi}\right) + \left|\frac{\left\|\tilde{V}^\top (P - \bar{P})\tilde{e}\right\|_{\mathbb{P},2}^2 - \left\|\tilde{V}^\top (Q - \bar{Q})\tilde{e}\right\|_{\mathbb{P},2}^2}{K}\right| \\
&= o(1),
\end{aligned}$$

where the first inequality is by triangular inequality, the second inequality is by Cauchy-Schwarz inequality, the third inequality is by Lemma A.5, and the last equality holds because

$$\mathbb{E}\left(\tilde{\Pi}^\top \bar{Q}\tilde{e}\right)^2 \leqslant C\left\|\bar{Q}\right\|_{op}^2 \tilde{\Pi}^\top \tilde{\Pi} = o(\Pi^\top \Pi)$$

by Lemma A.2, and

$$\begin{aligned}
\frac{1}{K}\mathbb{E}\left(\tilde{V}^\top (Q - \bar{Q})\tilde{e}\right)^2 &= \frac{1}{K}\sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)^2 + \left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right] \\
&= \frac{1}{K}\sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2 + \left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\right] + o(1) \\
&= \frac{1}{K}\mathbb{E}\left(\tilde{V}^\top (P - \bar{P})\tilde{e}\right)^2 + o(1)
\end{aligned}$$

by Lemma A.4. Therefore, we have

$$\frac{\bar{\Sigma}}{\Sigma} \to 1.$$

In addition, note that we can write

$$\bar{\Sigma} = \sum_{g\in[G]} \mathbb{E}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2 + \sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right),$$

as in Chao et al. (2012), and thus

$$\begin{aligned}
\frac{\acute{\Sigma} - \bar{\Sigma}}{\bar{\Sigma}} = {} & \underbrace{\frac{1}{\bar{\Sigma}}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2 - \sum_{g\in[G]} \mathbb{E}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2\right)}_{R_{14}} \\
& + \underbrace{\frac{1}{\bar{\Sigma}}\begin{pmatrix} \sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) \\ -\sum_{g,h\in[G]^2,g\neq h} \mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) \end{pmatrix}}_{R_{15}} \\
& + \underbrace{\frac{1}{\bar{\Sigma}}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \acute{e}_{[g]}\right)^2 - \sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2\right)}_{R_{16}} \\
& + \underbrace{\frac{1}{\bar{\Sigma}}\begin{pmatrix} \sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \acute{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \acute{e}_{[g]}\right) \\ -\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right) \end{pmatrix}}_{R_{17}}.
\end{aligned}$$

Note also that, by Lemmas A.5 and A.6, we have $R_{14} = o_P(1)$ and $R_{15} = o_P(1)$.

For R_{16}, we have

$$\begin{aligned}
R_{16} = {} & \underbrace{\frac{1}{\bar{\Sigma}}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \acute{e}_{[g]}\right)^2 - \sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} e_{[g]}\right)^2\right)}_{R_{16,1}} \\
& + \underbrace{\frac{1}{\bar{\Sigma}}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} e_{[g]}\right)^2 - \sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]}\right)^2\right)}_{R_{16,2}}.
\end{aligned}$$

For $R_{16,1}$, we have

$$R_{16,1} = \frac{\acute{\Delta}^2}{\bar{\Sigma}} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right)^2 - \frac{2\acute{\Delta}}{\bar{\Sigma}} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right) \left(\sum_{k\in[G],k\neq g} \tilde{X}_{[k]}^\top Q_{[k,g]} e_{[g]} \right).$$

By Lemma A.5, we have

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right)^2 = O_P(1),$$

since

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \left(\bar{\Pi}_{[g]}^\top \Pi_{[g]} \right)^2 \leqslant \frac{C\bar{\Pi}^\top \bar{\Pi}}{\bar{\Sigma}} = O(1),$$

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \mathbb{E} \left(\bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \right)^2 \leqslant \frac{C\bar{\Pi}^\top \bar{\Pi}}{\bar{\Sigma}} = O(1),$$

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top Q_{[h,g]} \Pi_{[g]} \right)^2 \leqslant \frac{C\Pi^\top \Pi}{\bar{\Sigma}} = O(1),$$

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right)^2 \leqslant \frac{C}{\bar{\Sigma}} \sum_{g,h\in[G]^2,g\neq h} \operatorname{trace} \left(Q_{[g,h]} Q_{[h,g]} \right) = O(1).$$

By the same lemma, we also have

$$\frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right) \left(\sum_{k\in[G],k\neq g} \tilde{X}_{[k]}^\top Q_{[k,g]} e_{[g]} \right) = O_P(1),$$

since

$$\left| \frac{1}{\bar{\Sigma}} \sum_{g\in[G]} \mathbb{E} \left(\bar{\Pi}_{[g]}^\top \tilde{V}_{[g]} \right) \left(\bar{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \right| \leqslant \frac{C\bar{\Pi}^\top \bar{\Pi}}{\bar{\Sigma}} = O(1)$$

and

$$\left|\frac{1}{\bar{\Sigma}}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^{\top}Q_{[k,g]}\tilde{e}_{[g]}\right)\right|$$
$$\leqslant\frac{C}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[g,h]}Q_{[h,g]}\right)=O(1).$$

It follows that $R_{16,1}=o_P(1)$, and since $R_{16,2}=o_P(1)$ by Lemma A.5, we have $R_{16}=o_P(1)$.

For R_{17}, we have

$$R_{17}=\underbrace{\frac{1}{\bar{\Sigma}}\begin{pmatrix}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\acute{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\acute{e}_{[g]}\right)\\-\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)\end{pmatrix}}_{R_{17,1}}$$
$$+\underbrace{\frac{1}{\bar{\Sigma}}\begin{pmatrix}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)\\-\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)\end{pmatrix}}_{R_{17,2}}.$$

For $R_{17,1}$, we have

$$R_{17,1}=\frac{\hat{\Delta}^2}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)$$
$$-\frac{2\hat{\Delta}}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right).$$

By Lemma A.6, we have

$$\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)=O_P(1),$$

since

$$\left|\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{X}_{[g]}^{\top}Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)\right|\leq\frac{C}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right)=O(1),$$
$$\left|\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{\Pi}_{[g]}^{\top}Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^{\top}Q_{[h,g]}\Pi_{[g]}\right)\right|\leq\frac{C\Pi^{\top}\Pi}{\bar{\Sigma}}=O(1),$$

$$\left|\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\right| \leq \frac{C\Pi^\top\Pi}{\bar{\Sigma}} = O(1).$$

By the same lemma, we also have

$$\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}X_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right) = O_P(1),$$

since

$$\left|\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right| \leq \frac{C}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\operatorname{trace}\left(Q_{[h,g]}Q_{[g,h]}\right) = O(1),$$

$$\left|\frac{1}{\bar{\Sigma}}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\Pi_{[h]}\right)\left(\tilde{\Pi}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right| \leq \frac{C\Pi^\top\Pi}{\bar{\Sigma}} = O(1).$$

It follows that $R_{17,1} = o_P(1)$, and since $R_{17,2} = o_P(1)$ by Lemma A.6, we have $R_{17} = o_P(1)$. Combining the results above, we have

$$\frac{\acute{\Sigma} - \bar{\Sigma}}{\bar{\Sigma}} = o_P(1),$$

and the desired result follows from $\Sigma/\bar{\Sigma} \to 1$.

Step 4: Consistency of $\acute{\Upsilon}$. We have

$$\begin{aligned}\frac{\acute{\Upsilon} - \Upsilon}{\Upsilon} = 2 \times \underbrace{\frac{1}{\Upsilon}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2 - \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2\right)}_{R_{18}} \\ + 2 \times \underbrace{\frac{1}{\Upsilon}\left(\sum_{g,h\in[G]^2,g\neq h}\left(e_{[g]}^\top P_{[g,h]}e_{[h]}\right)^2 - \sum_{g,h\in[G]^2,g\neq h}\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)^2\right)}_{R_{19}} \\ + 2 \times \underbrace{\frac{1}{\Upsilon}\left(\sum_{g,h\in[G]^2,g\neq h}\left(\acute{e}_{[g]}^\top P_{[g,h]}\acute{e}_{[h]}\right)^2 - \sum_{g,h\in[G]^2,g\neq h}\left(e_{[g]}^\top P_{[g,h]}e_{[h]}\right)^2\right)}_{R_{20}},\end{aligned}$$

and note that

$$\begin{aligned}
\Upsilon &= \mathbb{V}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right)\\
&= \mathbb{E}\left(\sum_{g,h\in[G]^2,g\neq h}\tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right)^2\\
&= 2\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right)^2\\
&= 2\sum_{g,h\in[G]^2,g\neq h}\text{trace}\left[\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\right]\\
&\geqslant \frac{1}{C}\sum_{g,h\in[G]^2,g\neq h}\text{trace}\left[P_{[g,h]}P_{[h,g]}\right]\\
&\geqslant \frac{1}{C}K. \qquad \text{(C.19)}
\end{aligned}$$

By Lemma A.4, we have $R_{18} = o_P(1)$ and $R_{19} = o_P(1)$. For R_{20}, it can be written as

$$\begin{aligned}
R_{20} = \frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}&\Big(\big((\acute{e}_{[g]}-e_{[g]})^{\top}P_{[g,h]}(\acute{e}_{[h]}-e_{[h]})+e_{[g]}^{\top}P_{[g,h]}(\acute{e}_{[h]}-e_{[h]})\\
&+(\acute{e}_{[g]}-e_{[g]})^{\top}P_{[g,h]}e_{[h]}+e_{[g]}^{\top}P_{[g,h]}e_{[h]}\big)^2-\big(e_{[g]}^{\top}P_{[g,h]}e_{[h]}\big)^2\Big)\\
= \acute{\Delta}^4\times\underbrace{\frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^{\top}P_{[g,h]}X_{[h]}\right)^2}_{R_{20,1}}&\\
+\acute{\Delta}^2\times\underbrace{\frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}\left(e_{[g]}^{\top}P_{[g,h]}X_{[h]}\right)^2}_{R_{20,2}}&\\
+\acute{\Delta}^2\times\underbrace{\frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^{\top}P_{[g,h]}e_{[h]}\right)^2}_{R_{20,3}}&\\
-4\times\acute{\Delta}^3\times\underbrace{\frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^{\top}P_{[g,h]}X_{[h]}\right)\left(e_{[g]}^{\top}P_{[g,h]}X_{[h]}\right)}_{R_{20,4}}&\\
+2\times\acute{\Delta}^2\times\underbrace{\frac{1}{\Upsilon}\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^{\top}P_{[g,h]}X_{[h]}\right)\left(e_{[g]}^{\top}P_{[g,h]}e_{[h]}\right)}_{R_{20,5}}&
\end{aligned}$$

$$+ 2 \times \acute{\Delta}^2 \times \underbrace{\frac{1}{\Upsilon} \sum_{g,h \in [G]^2, g \neq h} \left(e_{[g]}^\top P_{[g,h]} X_{[h]} \right) \left(X_{[g]}^\top P_{[g,h]} e_{[h]} \right)}_{R_{20,6}}$$

$$- 4 \times \acute{\Delta} \times \underbrace{\frac{1}{\Upsilon} \sum_{g,h \in [G]^2, g \neq h} \left(e_{[g]}^\top P_{[g,h]} X_{[h]} \right) \left(e_{[g]}^\top P_{[g,h]} e_{[h]} \right)}_{R_{20,7}}.$$

By using the same argument as in the proof of Lemma A.6, we can show that

$$R_{20,i} = O_P(1), \quad i = 1, \ldots, 7,$$

whence $R_{20} = o_P(1)$. Combining the results above, we have

$$\frac{\acute{\Upsilon} - \Upsilon}{\Upsilon} = o_P(1).$$

This concludes the proof. □

C.8 Proof of Lemma A.8

We first introduce some notation. Denote

$$\dot{\Phi}_1 = (X^\top z \hat{A}_n z^\top X)^{-1} (X^\top z \hat{A}_n \dot{\Omega} \hat{A}_n z^\top X) (X^\top z \hat{A}_n z^\top X)^{-1},$$
$$\dot{\Omega} = \sum_{g \in [G]} \left(z_{[g]}^\top \dot{e}_{[g]} \right) \left(z_{[g]}^\top \dot{e}_{[g]} \right)^\top,$$

where $\dot{e} = Y - X\hat{\beta}_1$, and denote $\hat{\Delta}_1 = \hat{\beta}_1 - \beta$. In addition, denote

$$\ddot{\Phi}_2 = \left(X^\top (P - \bar{P}) X \right)^{-1} \ddot{\Sigma} \left(X^\top (P - \bar{P}) X \right)^{-1},$$
$$\ddot{\Sigma} = \sum_{g \in [G]} \left(\sum_{h \in [G], h \neq g} \tilde{X}_{[h]}^\top Q_{[h,g]} \ddot{e}_{[g]} \right)^2 + \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \ddot{e}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \ddot{e}_{[g]} \right),$$

where $\ddot{e} = Y - X\hat{\beta}_2$, and denote $\hat{\Delta}_2 = \hat{\beta}_2 - \beta$. Finally, denote $\tilde{\omega}_1 = \ddot{\Phi}_2^{1/2} / \left(\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2} \right)$ and $\tilde{\omega}_2 = \dot{\Phi}_1^{1/2} / \left(\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2} \right)$.

We divide the proof into four steps. In the first step, we show that if Assumptions

1-3 hold, then

$$\hat{\beta}_1 \xrightarrow{p} \beta, \quad \dot{\Phi}_1 = o_P(1), \quad \text{and} \quad (\hat{\beta}_1 - \beta)/\dot{\Phi}_1^{1/2} = O_P(1).$$

In the second step, we show that if Assumptions 1 and 3 hold and $\Pi^\top\Pi/\sqrt{K} \to \infty$, we have

$$(\hat{\beta}_2 - \beta)^2\Pi^\top\Pi/\sqrt{K} = o_P(1) \quad \text{and} \quad \ddot{\Phi}_2\Pi^\top\Pi/\sqrt{K} = o_P(1).$$

In the third step, we show that if Assumptions 1-3 hold and $\Pi^\top\Pi/\sqrt{K} = O(1)$, then

$$\hat{\beta}_2 - \beta = O_P(1) \quad \text{and} \quad 1/\ddot{\Phi}_2^{1/2} = O_P(1).$$

In the last step, we show that if Assumptions 1, 2′ and 3 hold, then

$$\hat{\beta}_1 - \beta = O_P(1) \quad \text{and} \quad 1/\dot{\Phi}_1^{1/2} = O_P(1).$$

Consequently, if Assumptions 1-3 hold and $\Pi^\top\Pi/\sqrt{K} \to \infty$, by the results in Steps 1 and 2, we have

$$\begin{aligned}(\hat{\beta} - \beta)^2\Pi^\top\Pi/\sqrt{K} \leq & \left(\frac{\ddot{\Phi}_2^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}}\right)^2 \times (\hat{\beta}_1 - \beta)^2\Pi^\top\Pi/\sqrt{K} \\ & + \left(\frac{\dot{\Phi}_1^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}}\right)^2 \times (\hat{\beta}_2 - \beta)^2\Pi^\top\Pi/\sqrt{K} \\ & \leq \frac{(\hat{\beta}_1 - \beta)^2}{\dot{\Phi}_1}\ddot{\Phi}_2\Pi^\top\Pi/\sqrt{K} + o_P(1) = o_P(1).\end{aligned}$$

If Assumptions 1-3 hold and $\Pi^\top\Pi/\sqrt{K} = O(1)$, then by results in Steps 1 and 3, we have

$$\begin{aligned}(\hat{\beta} - \beta)^2\Pi^\top\Pi/\sqrt{K} \leq & \left(\frac{\ddot{\Phi}_2^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}}\right)^2 \times (\hat{\beta}_1 - \beta)^2\Pi^\top\Pi/\sqrt{K} \\ & + \left(\frac{\dot{\Phi}_1^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}}\right)^2 \times (\hat{\beta}_2 - \beta)^2\Pi^\top\Pi/\sqrt{K}\end{aligned}$$

$$= (\hat{\beta}_1 - \beta)^2 \times O_P(1) + \dot{\Phi}_1 \times O_P(1) = o_P(1).$$

If Assumptions 1, 2′ and 3 hold and $\Pi^\top \Pi / \sqrt{K} \to \infty$, by results in Steps 2 and 4, we have

$$\begin{aligned}(\hat{\beta} - \beta)^2 \Pi^\top \Pi / \sqrt{K} &\leq \left(\frac{\ddot{\Phi}_2^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}} \right)^2 \times (\hat{\beta}_1 - \beta)^2 \Pi^\top \Pi / \sqrt{K} \\ &+ \left(\frac{\dot{\Phi}_1^{1/2}}{\dot{\Phi}_1^{1/2} + \ddot{\Phi}_2^{1/2}} \right)^2 \times (\hat{\beta}_2 - \beta)^2 \Pi^\top \Pi / \sqrt{K} \\ &\leq \frac{(\hat{\beta}_1 - \beta)^2}{\dot{\Phi}_1} \ddot{\Phi}_2 \Pi^\top \Pi / \sqrt{K} + o_P(1) = o_P(1).\end{aligned}$$

Therefore, we have established the desired results. Next, we focus on proving results in Steps 1–4.

Step 1: Assumptions 1-3 hold

We have

$$\hat{\beta}_1 - \beta = \frac{X^\top z \hat{A}_n z^\top e}{X^\top z \hat{A}_n z^\top X}.$$

By (C.17) and (C.18), we have

$$\frac{X^\top z \hat{A}_n z^\top X}{\Pi^\top z A_n z^\top \Pi} = \frac{(\frac{1}{r_n} X^\top z)(\frac{1}{\lambda_n} \hat{A}_n)(\frac{1}{r_n} z^\top X)}{(\frac{1}{r_n} \Pi^\top z)(\frac{1}{\lambda_n} A_n)(\frac{1}{r_n} z^\top \Pi)} = 1 + o_P(1). \tag{C.20}$$

In addition, we have

$$\begin{aligned}\frac{X^\top z \hat{A}_n z^\top e}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} &= \frac{(\frac{1}{r_n} X^\top z)(\frac{1}{\lambda_n} \hat{A}_n)(\frac{1}{\sqrt{n}} z^\top e)}{\sqrt{(\frac{1}{r_n} \Pi^\top z)(\frac{1}{\lambda_n} A_n)(\frac{1}{n} \Omega)(\frac{1}{\lambda_n} A_n)(\frac{1}{r_n} z^\top \Pi)}} \\ &= \frac{(\frac{1}{r_n} \Pi^\top z)(\frac{1}{\lambda_n} A_n)(\frac{1}{\sqrt{n}} z^\top \tilde{e}) + o_P(1)}{\sqrt{(\frac{1}{r_n} \Pi^\top z)(\frac{1}{\lambda_n} A_n)(\frac{1}{n} \Omega)(\frac{1}{\lambda_n} A_n)(\frac{1}{r_n} z^\top \Pi)}} \\ &= \frac{\Pi^\top z A_n z^\top \tilde{e}}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} + o_p(1),\end{aligned}$$

where the last equality holds because by Assumptions 1 and 2,

$$\left(\frac{1}{r_n}\Pi^\top z\right)\left(\frac{1}{\lambda_n}A_n\right)\left(\frac{1}{n}\Omega\right)\left(\frac{1}{\lambda_n}A_n\right)\left(\frac{1}{r_n}z^\top\Pi\right) \geqslant \lambda_{\min}\left(\left(\frac{1}{\lambda_n}A_n\right)\left(\frac{1}{n}\Omega\right)\left(\frac{1}{\lambda_n}A_n\right)\right) \geqslant \frac{1}{C}.$$

Combining the above results and recalling

$$\Phi_1 = (\Pi^\top z A_n z^\top \Pi)^{-1}(\Pi^\top z A_n \Omega A_n z^\top \Pi)(\Pi^\top z A_n z^\top \Pi)^{-1},$$

we have

$$\frac{\hat{\beta}_1 - \beta}{\sqrt{\Phi_1}} = \frac{\Pi^\top z A_n z^\top \Pi}{X^\top z \hat{A}_n z^\top X} \times \frac{X^\top z \hat{A}_n z^\top e}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} = O_P(1),$$

and

$$\Phi_1 = \frac{\Pi^\top z A_n \Omega A_n z^\top \Pi}{(\Pi^\top z A_n z^\top \Pi)^2} = O\left(\frac{n}{r_n^2}\right) = o(1), \tag{C.21}$$

which further imply

$$\hat{\beta}_1 - \beta = O_P(\sqrt{\Phi_1}) = o_P(1).$$

Given the consistency of $\hat{\beta}_1$, we can apply Lemma A.7 to show that $\dot{\Phi}_1/\Phi_1 \xrightarrow{p} 1$, which further implies that

$$\dot{\Phi}_1 \xrightarrow{p} 0 \quad \text{and} \quad \frac{(\hat{\beta}_1 - \beta)^2}{\dot{\Phi}_1} = O_P(1).$$

Step 2: Assumptions 1 and 3 hold, and $\Pi^\top\Pi/\sqrt{K} \to \infty$

We have

$$\hat{\beta}_2 - \beta = \frac{X^\top(P - \bar{P})e}{X^\top(P - \bar{P})X},$$

and

$$X^\top(P - \bar{P})e = \hat{\Pi}^\top\tilde{e} + \tilde{V}^\top(P - \bar{P})\tilde{e} + \tilde{V}^\top P_W \bar{P}\tilde{e} + \tilde{V}^\top \bar{P} P_W \tilde{e} - \tilde{V}^\top P_W \bar{P} P_W \tilde{e}, \tag{C.22}$$

where we use the fact that $P_W P = P P_W = 0_{n\times n}$ since $Z = M_W \tilde{Z}$. For the third term on the RHS of (C.22), as $\Sigma \geq C(\Pi^\top \Pi + K)$ by Lemma A.5, we have

$$\tilde{V}^\top P_W \bar{P} \tilde{e} = O_P(1) = o_P(\sqrt{\Sigma})$$

by Lemma A.3. Following the similar argument, we can show that

$$\tilde{V}^\top \bar{P} P_W \tilde{e} = o_P(\sqrt{\Sigma}) \quad \text{and} \quad \tilde{V}^\top P_W \bar{P} P_W \tilde{e} = o_P(\sqrt{\Sigma}).$$

Then, by (C.22), we have

$$X^\top (P - \bar{P}) e = \hat{\Pi}^\top \tilde{e} + \tilde{V}^\top (P - \bar{P}) \tilde{e} + o_P(\sqrt{\Sigma}). \tag{C.23}$$

In addition, we have

$$\begin{aligned}
X^\top (P - \bar{P}) X &= \Pi^\top (P - \bar{P}) \Pi + 2\hat{\Pi}^\top \tilde{V} + \tilde{V}^\top M_W (P - \bar{P}) M_W \tilde{V} \\
&= \Pi^\top (P - \bar{P}) \Pi + 2\hat{\Pi}^\top \tilde{V} + \tilde{V}^\top (P - \bar{P}) \tilde{V} + 2\tilde{V}^\top P_W (P - \bar{P}) \tilde{V} + \tilde{V}^\top P_W (P - \bar{P}) P_W \tilde{V} \\
&= \Pi^\top (P - \bar{P}) \Pi + 2\hat{\Pi}^\top \tilde{V} + \tilde{V}^\top (P - \bar{P}) \tilde{V} - 2\tilde{V}^\top P_W \bar{P} \tilde{V} - \tilde{V}^\top P_W \bar{P} P_W \tilde{V} \\
&= \Pi^\top (P - \bar{P}) \Pi + 2\hat{\Pi}^\top \tilde{V} + \tilde{V}^\top (P - \bar{P}) \tilde{V} + o_P(\sqrt{\Sigma}) \\
&= \Pi^\top (P - \bar{P}) \Pi + O_P(\sqrt{\Sigma}),
\end{aligned} \tag{C.24}$$

where the second last equality is by

$$\tilde{V}^\top P_W \bar{P} \tilde{V} = o_P(\sqrt{\Sigma}) \quad \text{and} \quad \tilde{V}^\top P_W \bar{P} P_W \tilde{V} = o_P(\sqrt{\Sigma})$$

due to the same argument above, and the last equality holds because

$$\mathbb{V}\left(\hat{\Pi}^\top \tilde{V}\right) = O(\Pi^\top \Pi) = O(\Sigma)$$

and

$$\mathbb{V}\left(\tilde{V}^\top (P - \bar{P}) \tilde{V}\right) \leq C ||P - \bar{P}||_F^2 \leq CK = O(\Sigma).$$

Combining (C.23) and (C.24), we have

$$\hat{\beta}_2 - \beta = \frac{X^\top (P - \bar{P})e}{X^\top (P - \bar{P})X} = \frac{\hat{\Pi}^\top \tilde{e} + \tilde{V}^\top (P - \bar{P})\tilde{e} + o_P(\sqrt{\Sigma})}{\Pi^\top (P - \bar{P})\Pi + O_P(\sqrt{\Sigma})}.$$

In addition, we have

$$\Phi_2 = \frac{\Sigma}{\left(\Pi^\top (P - \bar{P})\Pi\right)^2} = O(\frac{\Pi^\top \Pi + K}{(\Pi^\top \Pi)^2}) = o(1) \tag{C.25}$$

because as $\Pi^\top \Pi / \sqrt{K} \to \infty$, we have $\Pi^\top \Pi \to \infty$; this also implies that

$$\frac{X^\top (P - \bar{P})X}{\Pi^\top (P - \bar{P})\Pi} = \frac{\Pi^\top (P - \bar{P})\Pi + O_P(\sqrt{\Sigma})}{\Pi^\top (P - \bar{P})\Pi} = 1 + o_P(1). \tag{C.26}$$

Therefore, we have

$$\frac{|\hat{\beta}_2 - \beta|}{\sqrt{\Phi_2}} = \frac{\left(\hat{\Pi}^\top \tilde{e} + \tilde{V}^\top (P - \bar{P})\tilde{e}\right)/\sqrt{\Sigma} + o_P(1)}{1 + O_P\left(\sqrt{\Sigma}/|\Pi^\top (P - \bar{P})\Pi|\right)} = O_P(1),$$

where we use the fact that

$$\hat{\Pi}^\top \tilde{e} = O_P(\sqrt{\Pi^\top \Pi}) = O_P(\sqrt{\Sigma})$$

and

$$\tilde{V}^\top (P - \bar{P})\tilde{e} = O_P(\sqrt{K}) = O_P(\sqrt{\Sigma}).$$

This also implies

$$\begin{aligned}
(\hat{\beta}_2 - \beta)^2 \Pi^\top \Pi / \sqrt{K} &= O_P(\Phi_2 \Pi^\top \Pi / \sqrt{K}) \\
&= O_P\left(\frac{\Pi^\top \Pi + K}{(\Pi^\top \Pi)^2} \times \frac{\Pi^\top \Pi}{\sqrt{K}}\right) \\
&= O_P\left(\frac{1}{\sqrt{K}} + \frac{\sqrt{K}}{\Pi^\top \Pi}\right) = o_P(1).
\end{aligned}$$

Given the consistency of $\hat{\beta}_2$, we can apply Lemma A.7 to show that $\ddot{\Phi}_2 / \Phi_2 \xrightarrow{p} 1$,

which further implies that

$$\ddot{\Phi}_2 \Pi^\top \Pi / \sqrt{K} = o_P(1).$$

Step 3: Assumptions 1 and 3 hold, and $\Pi^\top \Pi / \sqrt{K}$ is bounded

Note that $\Pi^\top (P - \bar{P}) \Pi / \sqrt{K}$ is bounded in this case. In addition, let

$$\begin{aligned}
\Gamma_{\tilde{V},\tilde{V}} &= \sum_{g,h \in [G]^2, g \neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{V}_{[h]} \right)^2 + \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]} \right) \right], \\
\Gamma_{\tilde{V},\tilde{e}} &= \sum_{g,h \in [G]^2, g \neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right)^2 + \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \right], \\
\Lambda &= \sum_{g,h \in [G]^2, g \neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{V}_{[h]} \right) + \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]} \right) \right],
\end{aligned}$$

then by Assumption 3.4 we have $1/C \leq \Gamma_{\tilde{V},\tilde{V}}/K \leq C$, $1/C \leq \Gamma_{\tilde{V},\tilde{e}}/K \leq C$ and $|\Lambda| / \sqrt{\Gamma_{\tilde{V},\tilde{V}} \Gamma_{\tilde{V},\tilde{e}}} \leq C < 1$. We shall argue along the subsequence where $\Pi^\top (P - \bar{P}) \Pi / \sqrt{K} \to \gamma \in \Re$ and

$$\frac{1}{K} \begin{pmatrix} \Gamma_{\tilde{V},\tilde{V}} & \Lambda \\ \Lambda & \Gamma_{\tilde{V},\tilde{e}} \end{pmatrix} \to \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \equiv \Gamma,$$

where $\Gamma > 0$ (in the matrix sense) by $\Gamma_{\tilde{V},\tilde{V}}/K \geq 1/C > 0$, $\Gamma_{\tilde{V},\tilde{e}}/K \geq 1/C > 0$ and $|\Lambda/K| / \sqrt{(\Gamma_{\tilde{V},\tilde{V}}/K)(\Gamma_{\tilde{V},\tilde{e}}/K)} \leq C < 1$.

By Assumption 3, we have

$$\begin{pmatrix} \frac{1}{\sqrt{K}} \tilde{V}^\top (P - \bar{P}) \tilde{V} \\ \frac{1}{\sqrt{K}} \tilde{V}^\top (P - \bar{P}) \tilde{e} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \right),$$

which can be proved by following the same steps as in the proof of Lemma A.10.

In addition, by (C.24), we have

$$\begin{aligned}
\frac{1}{\sqrt{K}} X^\top (P - \bar{P}) X = & \frac{1}{\sqrt{K}} \Pi^\top (P - \bar{P}) \Pi + \frac{2}{\sqrt{K}} \hat{\Pi}^\top \tilde{V} \\
& + \frac{1}{\sqrt{K}} \tilde{V}^\top (P - \bar{P}) \tilde{V} - \frac{2}{\sqrt{K}} \tilde{V}^\top P_W \bar{P} \tilde{V} - \frac{1}{\sqrt{K}} \tilde{V}^\top P_W \bar{P} P_W \tilde{V} \\
& \rightsquigarrow \mathcal{N}(\gamma, \Gamma_{11}),
\end{aligned}$$

where we use the facts that

$$\mathbb{V}\left(\frac{1}{\sqrt{K}}\hat{\Pi}\tilde{V}\right) = O\left(\frac{\Pi^\top\Pi}{K}\right) = o(1),$$

$$\frac{2}{\sqrt{K}}\tilde{V}^\top P_W\bar{P}\tilde{V} = O_P\left(1/\sqrt{K}\right) = o_P(1), \quad \text{and} \quad \frac{1}{\sqrt{K}}\tilde{V}^\top P_W\bar{P}P_W\tilde{V} = O_P\left(1/\sqrt{K}\right) = o_P(1).$$

This implies

$$\frac{1}{\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X} = O_P(1). \tag{C.27}$$

Similarly, by (C.22), we have

$$\begin{aligned}\frac{1}{\sqrt{K}}X^\top(P-\bar{P})e &= \frac{1}{\sqrt{K}}\hat{\Pi}^\top\tilde{e} + \frac{1}{\sqrt{K}}\tilde{V}^\top(P-\bar{P})\tilde{e} \\ &\quad + \frac{1}{\sqrt{K}}\tilde{V}^\top P_W\bar{P}\tilde{e} + \frac{1}{\sqrt{K}}\tilde{V}^\top\bar{P}P_W\tilde{e} - \frac{1}{\sqrt{K}}\tilde{V}^\top P_W\bar{P}P_W\tilde{e} \\ &= \frac{1}{\sqrt{K}}\tilde{V}^\top(P-\bar{P})\tilde{e} + o_P(1) = O_P(1),\end{aligned}$$

which further implies

$$\hat{\Delta}_2 = \frac{X^\top(P-\bar{P})e}{X^\top(P-\bar{P})X} = \frac{\frac{1}{\sqrt{K}}X^\top(P-\bar{P})e}{\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X} = O_P(1).$$

Next, we analyze $\ddot{\Phi}_2$. Note that

$$\frac{1}{K}\ddot{\Sigma} = \frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right)^2 + \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\ddot{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right), \tag{C.28}$$

where $\ddot{e} = Y - X\hat{\beta}_2$. For the first term on the RHS of (C.28), we have

$$\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right)^2$$

$$= \underbrace{\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2}}_{R_{21}}$$
$$+ \underbrace{\frac{1}{K}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)^{2} - \sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2}\right)}_{R_{22}}$$
$$+ \underbrace{\frac{1}{K}\left(\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\ddot{e}_{[g]}\right)^{2} - \sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}e_{[g]}\right)^{2}\right)}_{R_{23}}.$$

By the fact that $\Pi^{\top}\Pi/K = o(1)$, we have

$$1/C \leq \Sigma/K \leq C$$

for some constant $C \in (0,\infty)$, so that we can apply Lemma A.5 to obtain

$$R_{21} = \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2} + o_P(1),$$
$$= \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{e}_{[g]}\right)^{2} + o_P(1),$$

and $R_{22} = o_P(1)$. For R_{23}, we have

$$R_{23} = \hat{\Delta}_2^2 \times \underbrace{\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)^{2}}_{R_{23,1}}$$
$$- 2\times\hat{\Delta}_2 \times \underbrace{\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^{\top}Q_{[h,g]}X_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{X}_{[k]}^{\top}Q_{[k,g]}e_{[g]}\right)}_{R_{23,2}}.$$

By Lemma A.5 and the fact that $\Pi^{\top}\Pi/K = o(1)$, we have

$$R_{23,1} = \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}Q_{[h,g]}\tilde{V}_{[g]}\right)^{2} + o_P(1),$$

$$R_{23,2} = \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right) + o_P(1).$$

It follows that

$$\begin{aligned}
&\frac{1}{K}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right)^2 \\
&= \frac{1}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 \\
&+ \frac{\hat{\Delta}_2^2}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)^2 \\
&- \frac{2\hat{\Delta}_2}{K}\sum_{g\in[G]}\mathbb{E}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\sum_{k\in[G],k\neq g}\tilde{V}_{[k]}^\top Q_{[k,g]}\tilde{e}_{[g]}\right) + o_P(1), \qquad \text{(C.29)}
\end{aligned}$$

since $\hat{\Delta}_2 = O_P(1)$. For the second term on the RHS of (C.28), we have

$$\begin{aligned}
&\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\ddot{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right) \\
&= \underbrace{\frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)}_{R_{24}} \\
&+ \underbrace{\frac{1}{K}\begin{pmatrix}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right) \\ -\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\end{pmatrix}}_{R_{25}} \\
&+ \underbrace{\frac{1}{K}\begin{pmatrix}\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}\ddot{e}_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}\ddot{e}_{[g]}\right) \\ -\sum_{g,h\in[G]^2,g\neq h}\left(\tilde{X}_{[g]}^\top Q_{[g,h]}e_{[h]}\right)\left(\tilde{X}_{[h]}^\top Q_{[h,g]}e_{[g]}\right)\end{pmatrix}}_{R_{26}}.
\end{aligned}$$

By Lemma A.6 and the fact that $\Pi^\top\Pi/K = o(1)$, we have

$$R_{24} = \frac{1}{K}\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right) + o_P(1),$$

and $R_{25} = o_P(1)$. For R_{26}, we have

$$R_{26} = \hat{\Delta}^2 \times \underbrace{\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} X_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} X_{[g]} \right)}_{R_{26,1}}$$
$$- 2 \times \hat{\Delta} \times \underbrace{\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} X_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} e_{[g]} \right)}_{R_{26,2}}.$$

By Lemma A.6 and the fact that $\Pi^\top \Pi / K = o(1)$, we have

$$\begin{aligned} R_{26,1} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) + o_P(1) \\ &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) + o_P(1), \end{aligned}$$

and

$$\begin{aligned} R_{26,2} &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right) + o_P(1) \\ &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right) + o_P(1). \end{aligned}$$

It follows that

$$\begin{aligned} &\frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \left(\tilde{X}_{[g]}^\top Q_{[g,h]} \ddot{e}_{[h]} \right) \left(\tilde{X}_{[h]}^\top Q_{[h,g]} \ddot{e}_{[g]} \right) \\ &= \frac{1}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right) \\ &+ \frac{\hat{\Delta}_2^2}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{V}_{[g]} \right) \\ &- \frac{2\hat{\Delta}_2}{K} \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top Q_{[g,h]} \tilde{V}_{[h]} \right) \left(\tilde{V}_{[h]}^\top Q_{[h,g]} \tilde{e}_{[g]} \right) + o_P(1), \end{aligned} \tag{C.30}$$

since $\hat{\Delta}_2 = O_P(1)$. Combining (C.28)–(C.30), we obtain that

$$\begin{aligned}
\frac{1}{K}\ddot{\Sigma} &= \frac{\hat{\Delta}_2^2}{K}\left(\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)^2 + \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\right) \\
&+ \frac{1}{K}\left(\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)^2 + \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right) \\
&- \frac{2\hat{\Delta}_2}{K}\left[\sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{V}_{[g]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right. \\
&\qquad \left. + \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top Q_{[g,h]}\tilde{V}_{[h]}\right)\left(\tilde{V}_{[h]}^\top Q_{[h,g]}\tilde{e}_{[g]}\right)\right] + o_P(1) \\
&= \frac{\hat{\Delta}_2^2}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{V}\right)^2 - \frac{2\hat{\Delta}_2}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{V}\right)\left(\tilde{V}^\top(Q-\bar{Q})\tilde{e}\right) \\
&+ \frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{e}\right)^2 + o_P(1).
\end{aligned}$$

In addition, by Lemma A.4, we have

$$\begin{aligned}
\frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{V}\right)^2 &= \frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(P-\bar{P})\tilde{V}\right)^2 + o(1) = \Gamma_{11} + o(1), \\
\frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{e}\right)^2 &= \frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(P-\bar{P})\tilde{e}\right)^2 + o(1) = \Gamma_{22} + o(1), \\
\frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(Q-\bar{Q})\tilde{V}\right)\left(\tilde{V}^\top(Q-\bar{Q})\tilde{e}\right) &= \frac{1}{K}\mathbb{E}\left(\tilde{V}^\top(P-\bar{P})\tilde{V}\right)\left(\tilde{V}^\top(P-\bar{P})\tilde{e}\right) + o(1) = \Gamma_{12} + o(1).
\end{aligned}$$

Combining the above results, we have

$$\frac{1}{K}\ddot{\Sigma} = \hat{\Delta}_2^2\Gamma_{11} - 2\hat{\Delta}_2\Gamma_{12} + \Gamma_{22} + o_P(1). \tag{C.31}$$

It follows that

$$\begin{aligned}
\ddot{\Phi}_2 &= \left(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X\right)^{-2}\left(\frac{1}{K}\ddot{\Sigma}\right) \\
&= \frac{(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})e)^2}{(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X)^4}\times\Gamma_{11} - 2\frac{(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})e)}{(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X)^3}\times\Gamma_{12} \\
&+ \frac{1}{(\frac{1}{\sqrt{K}}X^\top(P-\bar{P})X)^2}\times\Gamma_{22} + o_P(1) \\
&\rightsquigarrow \bar{\Phi}_2,
\end{aligned}$$

where the second equality is by (C.27) and (C.31), and $\bar{\Phi}_2$ is defined as

$$\begin{aligned}\bar{\Phi}_2 &= \frac{\eta_{\tilde{V},\tilde{e}}^2}{(\eta_{\tilde{V},\tilde{V}}+\gamma)^4}\Gamma_{11} - \frac{2\eta_{\tilde{V},\tilde{e}}}{(\eta_{\tilde{V},\tilde{V}}+\gamma)^3}\Gamma_{12} + \frac{1}{(\eta_{\tilde{V},\tilde{V}}+\gamma)^2}\Gamma_{22} \\ &= \frac{1}{(\eta_{\tilde{V},\tilde{V}}+\gamma)^4}\left(\eta_{\tilde{V},\tilde{e}}^2\Gamma_{11} - 2\eta_{\tilde{V},\tilde{e}}(\eta_{\tilde{V},\tilde{V}}+\gamma)\Gamma_{12} + (\eta_{\tilde{V},\tilde{V}}+\gamma)^2\Gamma_{22}\right) \\ &= \frac{\begin{pmatrix}-\eta_{\tilde{V},\tilde{e}} \\ \eta_{\tilde{V},\tilde{V}}+\gamma\end{pmatrix}^\top \begin{pmatrix}\Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22}\end{pmatrix}\begin{pmatrix}-\eta_{\tilde{V},\tilde{e}} \\ \eta_{\tilde{V},\tilde{V}}+\gamma\end{pmatrix}}{(\eta_{\tilde{V},\tilde{V}}+\gamma)^4},\end{aligned}$$

with

$$\begin{pmatrix}\eta_{\tilde{V},\tilde{V}} \\ \eta_{\tilde{V},\tilde{e}}\end{pmatrix} \overset{d}{=} \mathcal{N}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22}\end{pmatrix}\right).$$

Note that

$$\begin{pmatrix}-\eta_{\tilde{V},\tilde{e}} \\ \eta_{\tilde{V},\tilde{V}}+\gamma\end{pmatrix}^\top \begin{pmatrix}\Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22}\end{pmatrix}\begin{pmatrix}-\eta_{\tilde{V},\tilde{e}} \\ \eta_{\tilde{V},\tilde{V}}+\gamma\end{pmatrix} \geqslant 0,$$

and the equality holds if and only if $\eta_{\tilde{V},\tilde{e}} = 0$ and $\eta_{\tilde{V},\tilde{V}} + \gamma = 0$, which has probability zero since $\eta_{\tilde{V},\tilde{e}}$ has a non-degenerate normal distribution and similarly for $\eta_{\tilde{V},\tilde{V}}$. In addition, the denominator of $\bar{\Phi}_2$ is positive with probability one. Therefore, we have $\bar{\Phi}_2 > 0$ with probability one, which implies that

$$1/\ddot{\Phi}_2^{1/2} = O_P(1).$$

Step 4: Assumptions 1, 2′ and 3 hold

By Assumptions 2′.1 and 2′.2, we can argue along the subsequence where $z^\top\Pi/\sqrt{n} \to \pi \in \Re^{d_z}$,

$$\frac{1}{n}\begin{pmatrix}\sum_{g\in[G]}\mathbb{E}\left(z_{[g]}^\top\tilde{e}_{[g]}\right)\left(z_{[g]}^\top\tilde{e}_{[g]}\right)^\top & \sum_{g\in[G]}\mathbb{E}\left(z_{[g]}^\top\tilde{e}_{[g]}\right)\left(z_{[g]}^\top\tilde{V}_{[g]}\right)^\top \\ \sum_{g\in[G]}\mathbb{E}\left(z_{[g]}^\top\tilde{V}_{[g]}\right)\left(z_{[g]}^\top\tilde{e}_{[g]}\right)^\top & \sum_{g\in[G]}\mathbb{E}\left(z_{[g]}^\top\tilde{V}_{[g]}\right)\left(z_{[g]}^\top\tilde{V}_{[g]}\right)^\top\end{pmatrix} \to \begin{pmatrix}\Omega_z^{\tilde{e},\tilde{e}} & \Omega_z^{\tilde{e},\tilde{V}} \\ \Omega_z^{\tilde{V},\tilde{e}} & \Omega_z^{\tilde{V},\tilde{V}}\end{pmatrix} > 0,$$

and

$$\frac{1}{n}\sum_{g\in[G]}\left(z_{[g]}^{\top}\Pi_{[g]}\right)\left(z_{[g]}^{\top}\Pi_{[g]}\right)^{\top}\to\Omega_z^{\Pi,\Pi}\geqslant 0,$$

in the matrix sense. In addition, note that since Assumption 1.5 holds, A_n/λ_n has eigenvalues bounded and bounded away from zero, where $\lambda_n=\lambda_{\max}(A_n)$. Therefore, without loss of generality, we assume that $A_n/\lambda_n\to A$ for some non-random positive definite matrix A with eigenvalues bounded and bounded away from zero (otherwise we argue along a further subsequence).

By Assumption 2′, we have

$$\begin{pmatrix}\frac{1}{\sqrt{n}}z^{\top}\tilde{e}\\ \frac{1}{\sqrt{n}}z^{\top}\tilde{V}\end{pmatrix}\rightsquigarrow\mathcal{N}\left(0,\begin{pmatrix}\Omega_z^{\tilde{e},\tilde{e}} & \Omega_z^{\tilde{e},\tilde{V}}\\ \Omega_z^{\tilde{V},\tilde{e}} & \Omega_z^{\tilde{V},\tilde{V}}\end{pmatrix}\right),$$

which can be proved by following the same steps as in the proof of Lemma A.10 (see also Hansen and Lee (2019) and Djogbenou, MacKinnon, and Nielsen (2019)). This also implies that

$$\frac{1}{\sqrt{n}}z^{\top}X\rightsquigarrow\mathcal{N}\left(\pi,\Omega_z^{\tilde{V},\tilde{V}}\right).$$

We have

$$\hat{\Delta}_1=\frac{X^{\top}z\hat{A}_nz^{\top}e}{X^{\top}z\hat{A}_nz^{\top}X}=\frac{(\frac{1}{\sqrt{n}}X^{\top}z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^{\top}\tilde{e})}{(\frac{1}{\sqrt{n}}X^{\top}z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^{\top}X)}$$

and

$$(\frac{1}{\sqrt{n}}X^{\top}z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^{\top}\tilde{e})=O_P(1).$$

We also have

$$\frac{1}{(\frac{1}{\sqrt{n}}X^{\top}z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^{\top}X)}\leqslant\frac{1}{\xi_n^2},$$

where ξ_n is the first element of $(\hat{A}_n/\tilde{\lambda}_n)^{1/2}(z^{\top}X/\sqrt{n})$, and note that $\xi_n\rightsquigarrow\xi$ where ξ

has a non-degenerate normal distribution. This implies that

$$\frac{1}{(\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\lambda_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)} = O_P(1),$$

whence $\hat{\Delta}_1 = O_P(1)$.

Next, we analyze $\dot{\Phi}_1$. Note that

$$\begin{aligned}\frac{1}{n}\dot{\Omega} &= \left(\frac{1}{n}\dot{\Omega} - \frac{1}{n}\bar{\Omega}\right) + \left(\frac{1}{n}\bar{\Omega} - \frac{1}{n}\tilde{\Omega}\right) + \left(\frac{1}{n}\tilde{\Omega} - \frac{1}{n}\Omega\right) + \frac{1}{n}\Omega \\ &= \left(\frac{1}{n}\dot{\Omega} - \frac{1}{n}\bar{\Omega}\right) + \Omega_z^{\tilde{e},\tilde{e}} + o_P(1),\end{aligned}$$

as in the proof of Lemma A.7, where $\dot{\Omega}$ is just $\acute{\Omega}$ with $\acute{\beta} = \hat{\beta}_1$. We have

$$\begin{aligned}\frac{1}{n}\tilde{\Omega} - \frac{1}{n}\bar{\Omega} = \hat{\Delta}_1^2 \times \underbrace{\frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top X_{[g]})(X_{[g]}^\top z_{[g]})}_{R_{14}} \\ - \hat{\Delta}_1 \times \underbrace{\frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top X_{[g]})e_{[g]}^\top z_{[g]}}_{R_{15}} \\ - \hat{\Delta}_1 \times \underbrace{\frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top e_{[g]})(X_{[g]}^\top z_{[g]})}_{R_{16}}.\end{aligned}$$

By using the same argument as in the proof of Lemma A.7, we have

$$\begin{aligned}R_{14} &= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \Pi_{[g]})(\Pi_{[g]}^\top z_{[g]}) + \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top V_{[g]})(V_{[g]}^\top z_{[g]}) + o_P(1) \\ &= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \Pi_{[g]})(\Pi_{[g]}^\top z_{[g]}) + \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \tilde{V}_{[g]})(\tilde{V}_{[g]}^\top z_{[g]}) + o_P(1) \\ &= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \Pi_{[g]})(\Pi_{[g]}^\top z_{[g]}) + \frac{1}{n}\sum_{g\in[G]}\mathbb{E}(z_{[g]}^\top \tilde{V}_{[g]})(\tilde{V}_{[g]}^\top z_{[g]}) + o_P(1) \\ &= \Omega_z^{\Pi,\Pi} + \Omega_z^{\tilde{V},\tilde{V}} + o_P(1),\end{aligned}$$

and

$$R_{15} = \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top V_{[g]})(e_{[g]}^\top z_{[g]}) + o_P(1)$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \tilde{V}_{[g]})(\tilde{e}_{[g]}^\top z_{[g]}) + o_P(1) \\
&= \frac{1}{n}\sum_{g\in[G]}\mathbb{E}(z_{[g]}^\top \tilde{V}_{[g]})(\tilde{e}_{[g]}^\top z_{[g]}) + o_P(1) \\
&= \Omega_z^{\tilde{V},\tilde{e}} + o_P(1),
\end{aligned}
$$

and

$$
\begin{aligned}
R_{16} &= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top e_{[g]})(V_{[g]}^\top z_{[g]}) + o_P(1) \\
&= \frac{1}{n}\sum_{g\in[G]}(z_{[g]}^\top \tilde{e}_{[g]})(\tilde{V}_{[g]}^\top z_{[g]}) + o_P(1) \\
&= \frac{1}{n}\sum_{g\in[G]}\mathbb{E}(z_{[g]}^\top \tilde{e}_{[g]})(\tilde{V}_{[g]}^\top z_{[g]}) + o_P(1) \\
&= \Omega_z^{\tilde{e},\tilde{V}} + o_P(1).
\end{aligned}
$$

We thus obtain

$$
\frac{1}{n}\tilde{\Omega} = \hat{\Delta}_1^2\Omega_z^{\Pi,\Pi} + \hat{\Delta}_1^2\Omega_z^{\tilde{V},\tilde{V}} - \hat{\Delta}_1\Omega_z^{\tilde{V},\tilde{e}} - \hat{\Delta}_1\Omega_z^{\tilde{e},\tilde{V}} + \Omega_z^{\tilde{e},\tilde{e}} + o_P(1),
$$

since $\hat{\Delta}_1 = O_P(1)$.

Combining the above results, we have

$$
\begin{aligned}
\dot{\Phi}_1 = &\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^{-2}\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{n}\tilde{\Omega})(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right) \\
= &\frac{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top \tilde{e})\right)^2}{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^4}\times\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)\Omega_z^{\Pi,\Pi}(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right) \\
&+\frac{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top \tilde{e})\right)^2}{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^4}\times\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)\Omega_z^{\tilde{V},\tilde{V}}(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right) \\
&-\frac{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top \tilde{e})\right)}{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^3}\times\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)\Omega_z^{\tilde{V},\tilde{e}}(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right) \\
&-\frac{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top \tilde{e})\right)}{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^3}\times\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)\Omega_z^{\tilde{e},\tilde{V}}(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)
\end{aligned}
$$

$$+\frac{1}{\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)^2}\times\left((\frac{1}{\sqrt{n}}X^\top z)(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)\Omega_z^{\tilde{e},\tilde{e}}(\frac{1}{\tilde{\lambda}_n}\hat{A}_n)(\frac{1}{\sqrt{n}}z^\top X)\right)+o_P(1),$$

and thus $\dot{\Phi}_1 \rightsquigarrow \bar{\Phi}_1$, by the continuous mapping theorem, where

$$\begin{aligned}
\bar{\Phi}_1 &= \frac{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^4}\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\Pi,\Pi}A(\zeta_{\tilde{V}}+\pi)\right)\\
&+\frac{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^4}\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{V},\tilde{V}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&-\frac{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^3}\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{V},\tilde{e}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&-\frac{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^3}\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{e},\tilde{V}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&+\frac{1}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^2}\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{e},\tilde{e}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&=\frac{1}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^4}\left\{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\Pi,\Pi}A(\zeta_{\tilde{V}}+\pi)\right)\right.\\
&+\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{V},\tilde{V}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&-\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)\times\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{V},\tilde{e}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&-\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)\times\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{e},\tilde{V}}A(\zeta_{\tilde{V}}+\pi)\right)\\
&\left.+\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^2\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\tilde{e},\tilde{e}}A(\zeta_{\tilde{V}}+\pi)\right)\right\}\\
&=\frac{\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2\times\left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\Pi,\Pi}A(\zeta_{\tilde{V}}+\pi)\right)}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^4}\\
&+\frac{\left[\begin{array}{c}\left(\begin{array}{c}-\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)(A(\zeta_{\tilde{V}}+\pi))\\ \left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)(A(\zeta_{\tilde{V}}+\pi))\end{array}\right)^\top\begin{pmatrix}\Omega_z^{\tilde{e},\tilde{e}} & \Omega_z^{\tilde{e},\tilde{V}}\\ \Omega_z^{\tilde{V},\tilde{e}} & \Omega_z^{\tilde{V},\tilde{V}}\end{pmatrix}\\ \times\left(\begin{array}{c}-\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)(A(\zeta_{\tilde{V}}+\pi))\\ \left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)(A(\zeta_{\tilde{V}}+\pi))\end{array}\right)\end{array}\right]}{\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)^4},
\end{aligned}$$

with

$$\begin{pmatrix}\zeta_{\tilde{e}}\\ \zeta_{\tilde{V}}\end{pmatrix}\overset{d}{=}\mathcal{N}\left(0,\begin{pmatrix}\Omega_z^{\tilde{e},\tilde{e}} & \Omega_z^{\tilde{e},\tilde{V}}\\ \Omega_z^{\tilde{V},\tilde{e}} & \Omega_z^{\tilde{V},\tilde{V}}\end{pmatrix}\right).$$

Finally, we note that

$$\left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)^2 \times \left((\zeta_{\tilde{V}}+\pi)^\top A\Omega_z^{\Pi,\Pi}A(\zeta_{\tilde{V}}+\pi)\right) \geqslant 0,$$

and

$$\left[\begin{array}{c} \begin{pmatrix} -\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)(A(\zeta_{\tilde{V}}+\pi)) \\ \left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)(A(\zeta_{\tilde{V}}+\pi)) \end{pmatrix}^\top \begin{pmatrix} \Omega_z^{\tilde{e},\tilde{e}} & \Omega_z^{\tilde{e},\tilde{V}} \\ \Omega_z^{\tilde{V},\tilde{e}} & \Omega_z^{\tilde{V},\tilde{V}} \end{pmatrix} \\ \times \begin{pmatrix} -\left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)(A(\zeta_{\tilde{V}}+\pi)) \\ \left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)(A(\zeta_{\tilde{V}}+\pi)) \end{pmatrix} \end{array}\right] \geqslant 0,$$

and the equalities hold if and only if

$$\begin{aligned} \left((\zeta_{\tilde{V}}+\pi)^\top A(\zeta_{\tilde{V}}+\pi)\right)(A(\zeta_{\tilde{V}}+\pi)) &= 0, \\ \left((\zeta_{\tilde{V}}+\pi)^\top A\zeta_{\tilde{e}}\right)(A(\zeta_{\tilde{V}}+\pi)) &= 0. \end{aligned}$$

Given that A is positive definite, the above two equalities hold if and only if $\zeta_{\tilde{V}}+\pi=0$, which has probability zero since $\zeta_{\tilde{V}}$ has a non-degenerate normal distribution. In addition, the denominator of $\bar{\Phi}_1$ is positive with probability one, which implies

$$1/\dot{\Phi}_1^{1/2} = O_P(1).$$

This concludes the proof. □

C.9 Proof of Lemma A.9

By Lemma A.8, we have $\hat{\beta} \xrightarrow{p} \beta$ and $(\hat{\beta}-\beta)^2\Pi^\top\Pi/\sqrt{K} = o_P(1)$.

For $T(\beta_0)$, by Lemma A.7,

$$\frac{\Psi}{\hat{\bar{\Psi}}} \xrightarrow{p} 1,$$

which, by (C.20), implies that

$$\frac{\Phi_1}{\hat{\bar{\Phi}}_1} \xrightarrow{p} 1.$$

We have

$$T(\beta_0) = \frac{(X^\top z\hat{A}_n z^\top X)^{-1}(X^\top z\hat{A}_n z^\top e)}{\sqrt{\hat{\Phi}_1}} + a_1\delta + \left(\frac{d_n}{\sqrt{\hat{\Phi}_1}} - a_1\right)\delta$$
$$= \frac{(X^\top z\hat{A}_n z^\top X)^{-1}(X^\top z\hat{A}_n z^\top e)}{\sqrt{\hat{\Phi}_1}} + a_1\delta + o_P(1),$$

where the second equality holds because

$$\frac{d_n}{\sqrt{\hat{\Phi}_1}} - a_1 = \frac{d_n}{\sqrt{\Phi_1}}\frac{\sqrt{\Phi_1}}{\sqrt{\hat{\Phi}_1}} - a_1 = o_p(1).$$

In addition, we have

$$\begin{aligned}\frac{(X^\top z\hat{A}_n z^\top X)^{-1}(X^\top z\hat{A}_n z^\top e)}{\sqrt{\hat{\Phi}_1}} &= \frac{(X^\top z\hat{A}_n z^\top X)^{-1}}{\sqrt{(X^\top z\hat{A}_n z^\top X)^{-2}}} \times \frac{(X^\top z\hat{A}_n z^\top e)}{\sqrt{\hat{\Psi}}} \\ &= \sqrt{\frac{\Psi}{\hat{\Psi}}} \times \frac{(X^\top z\hat{A}_n z^\top e)}{\sqrt{\Psi}} \times (1 + o_P(1)) \\ &= \frac{(\Pi^\top z A_n z^\top \tilde{e})}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} \times (1 + o_P(1)) \\ &= \frac{(\Pi^\top z A_n z^\top \tilde{e})}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} + o_P(1),\end{aligned}$$

whence

$$T(\beta_0) = \frac{1}{\sqrt{\Psi}} \sum_{g\in[G]} \acute{\Pi}_{[g]}^\top \tilde{e}_{[g]} + a_1\delta + o_P(1).$$

For $LM(\beta_0)$, if $\Pi^\top\Pi/\sqrt{K} \to \infty$, we have

$$\begin{aligned}LM(\beta_0) &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + a_2\delta + \left(\frac{d_n(X^\top(P-\bar{P})X)}{\sqrt{\hat{\Sigma}}} - a_2\right)\delta \\ &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + a_2\delta + \left(\frac{d_n}{\sqrt{\Phi_2}} \times (1 + o_P(1)) - a_2\right)\delta \\ &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + a_2\delta + o_P(1),\end{aligned}$$

where the second equality holds because by Lemma A.7,

$$\frac{\Sigma}{\hat{\Sigma}} \xrightarrow{p} 1,$$

which, by (C.26), implies that

$$\frac{\Phi_2(X^\top(P-\bar{P})X)^2}{\hat{\Sigma}} = \frac{\Phi_2(X^\top(P-\bar{P})X)^2}{\Sigma}\frac{\Sigma}{\hat{\Sigma}} = \frac{(X^\top(P-\bar{P})X)^2}{(\Pi^\top(P-\bar{P})\Pi)^2}\frac{\Sigma}{\hat{\Sigma}} \xrightarrow{p} 1.$$

Alternatively, if $\Pi^\top\Pi/\sqrt{K} = O(1)$, we have

$$\begin{aligned} LM(\beta_0) &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + \frac{d_n\delta(X^\top(P-\bar{P})X)}{\sqrt{\hat{\Sigma}}} \\ &= \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + o_P(1), \end{aligned}$$

where we use the fact that

$$\frac{X^\top(P-\bar{P})X}{\sqrt{\hat{\Sigma}}} = \frac{X^\top(P-\bar{P})X}{\sqrt{\Sigma}} \times \frac{\Sigma}{\hat{\Sigma}} = \frac{\Pi^\top(P-\bar{P})\Pi + O_P(\sqrt{\Sigma})}{\sqrt{\Sigma}} \times O_P(1) = O_P(1)$$

by (C.24), $\Pi^\top\Pi/\sqrt{K} = O(1)$, and

$$\frac{1}{\sqrt{\Phi_2}} = \sqrt{\frac{(\Pi^\top(P-\bar{P})\Pi)^2}{\Sigma}} = O(\frac{\Pi^\top\Pi}{\sqrt{K}}) = O(1),$$

Assumption 4 implies that $a_2 = \lim_{n\to\infty} d_n/\sqrt{\Phi_2} = 0$ in this case. In either case, we obtain

$$LM(\beta_0) = \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} + a_2\delta + o_P(1).$$

In addition, we have

$$\begin{aligned} \frac{X^\top(P-\bar{P})e}{\sqrt{\hat{\Sigma}}} &= \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{1}{\sqrt{\Sigma}}\left(\hat{\Pi}^\top\tilde{e} + \tilde{V}^\top(P-\bar{P})\tilde{e} + \tilde{V}^\top P_W\bar{P}\tilde{e} + \tilde{V}^\top\bar{P}P_W\tilde{e} - \tilde{V}^\top P_W\bar{P}P_W\tilde{e}\right) \\ &= \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{1}{\sqrt{\Sigma}}\left(\hat{\Pi}^\top\tilde{e} + \tilde{V}^\top(P-\bar{P})\tilde{e} + o_P(\sqrt{\Sigma})\right), \end{aligned}$$

$$= \frac{1}{\sqrt{\Sigma}} \left(\hat{\Pi}^\top \tilde{e} + \tilde{V}^\top (P - \bar{P}) \tilde{e} \right) + o_P(1),$$

where the second equality holds by $\Sigma \geq C(\Pi^\top \Pi + K) \to \infty$ as shown in Lemmas A.3 and A.5, and the last equality holds by Lemma A.7. It follows that

$$LM(\beta_0) = \frac{1}{\sqrt{\Sigma}} \left(\sum_{g \in [G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h \in [G]^2, g \neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) + a_2 \delta + o_P(1).$$

For AR, we have

$$e^\top (P - \bar{P}) e - \tilde{e}^\top (P - \bar{P}) \tilde{e} = \tilde{e}^\top P_W \bar{P} \tilde{e} + \tilde{e}^\top \bar{P} P_W \tilde{e} - \tilde{e}^\top P_W \bar{P} P_W \tilde{e} = O_P(1)$$

by Lemma A.3, and $\Upsilon \geq CK$ by (C.19). This implies that

$$\frac{1}{\sqrt{\Upsilon}} \left(\sum_{g,h \in [G]^2, g \neq h} e_{[g]}^\top P_{[g,h]} e_{[h]} - \sum_{g,h \in [G]^2, g \neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) = O_P(K^{-1/2}) = o_P(1).$$

We also have

$$\begin{aligned}
\hat{e}^\top (P - \bar{P}) \hat{e} - e^\top (P - \bar{P}) e &= (\hat{\beta} - \beta)^2 X^\top (P - \bar{P}) X - 2(\hat{\beta} - \beta) X^\top (P - \bar{P}) e \\
&= (\hat{\beta} - \beta)^2 \left(\Pi^\top (P - \bar{P}) \Pi + O_P(\sqrt{\Sigma}) \right) \\
&\quad - 2(\hat{\beta} - \beta) \left(\hat{\Pi}^\top \tilde{e} + \tilde{V}^\top (P - \bar{P}) \tilde{e} + o_P(\sqrt{\Sigma}) \right) \\
&= (\hat{\beta} - \beta)^2 O_P(\Pi^\top \Pi + \sqrt{\Sigma}) + (\hat{\beta} - \beta) O_P(\sqrt{\Sigma}) \\
&= \left((\hat{\beta} - \beta)^2 \Pi^\top \Pi + (\hat{\beta} - \beta)^2 \sqrt{K} + (\hat{\beta} - \beta) \sqrt{K} \right) \times O_P(1)
\end{aligned}$$

by (C.22) and (C.24), $\Pi^\top \Pi = O(\sqrt{K})$, and $\Sigma = O(\Pi^\top \Pi + K) = O(K)$. Then by $\hat{\beta} \xrightarrow{p} \beta$ and $(\hat{\beta} - \beta)^2 \Pi^\top \Pi / \sqrt{K} = o_P(1)$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{\Upsilon}} \left(\sum_{g,h \in [G]^2, g \neq h} \hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} - \sum_{g,h \in [G]^2, g \neq h} e_{[g]}^\top P_{[g,h]} e_{[h]} \right) \\
&= \left(\frac{(\hat{\beta} - \beta)^2 \Pi^\top \Pi}{\sqrt{K}} + (\hat{\beta} - \beta)^2 + (\hat{\beta} - \beta) \right) \times O_P(1) \\
&= o_P(1).
\end{aligned}$$

It follows that, by Lemma A.7, we have

$$\begin{aligned}
AR &= \sqrt{\frac{\Upsilon}{\hat{\Upsilon}}} \times \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \\
&= \sqrt{\frac{\Upsilon}{\hat{\Upsilon}}} \times \frac{1}{\sqrt{\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} - \sum_{g,h\in[G]^2, g\neq h} e_{[g]}^\top P_{[g,h]} e_{[h]} \right) \\
&+ \sqrt{\frac{\Upsilon}{\hat{\Upsilon}}} \times \frac{1}{\sqrt{\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} e_{[g]}^\top P_{[g,h]} e_{[h]} - \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \\
&+ \sqrt{\frac{\Upsilon}{\hat{\Upsilon}}} \times \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \\
&= \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} + o_P(1).
\end{aligned}$$

This concludes the proof. □

C.10 Proof of Lemma A.10

Note that by Assumption 1.2, we have $G \to \infty$ as $n \to \infty$. The proof follows the same steps as in Chao et al. (2012) to check the conditions for the martingale central limit theorem; see, for example, Hall and Heyde (1980).

Step 1: Construct martingale difference array. Let

$$\begin{aligned}
\rho_{1n} &= \operatorname{cov}\left[\frac{1}{\sqrt{\Psi}} \sum_{g\in[G]} \acute{\Pi}_{[g]}^\top \tilde{e}_{[g]}, \frac{1}{\sqrt{\Sigma}} \left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \right] \\
&= \frac{1}{\sqrt{\Psi\Sigma}} \sum_{g\in[G]} \mathbb{E}\left[\left(\acute{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \right],
\end{aligned}$$

such that $\rho_1 = \lim_{n\to\infty} \rho_{1n}$, and

$$\begin{aligned}
\rho_{2n} &= \operatorname{cov}\left[\frac{1}{\sqrt{\Sigma}} \left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} + \sum_{g,h\in[G]^2, g\neq h} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right), \frac{1}{\sqrt{\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right] \\
&= \frac{2}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left[\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \right],
\end{aligned}$$

such that $\rho_2 = \lim_{n\to\infty} \rho_{2n}$, and note that

$$\operatorname{cov}\left[\frac{1}{\sqrt{\Psi}}\sum_{g\in[G]} \acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}, \frac{1}{\sqrt{\Upsilon}}\sum_{g,h\in[G]^2,g\neq h} \tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right] = 0.$$

The assumptions for ρ_1 and ρ_2 in Theorem 4.1 ensure that

$$\begin{pmatrix} 1 & \rho_{1n} & 0 \\ \rho_{1n} & 1 & \rho_{2n} \\ 0 & \rho_{2n} & 1 \end{pmatrix}^{-1}$$

exists for all n large enough, and by the Slutsky theorem, the result would follow if

$$\begin{pmatrix} 1 & \rho_{1n} & 0 \\ \rho_{1n} & 1 & \rho_{2n} \\ 0 & \rho_{2n} & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} \frac{1}{\sqrt{\Psi}}\sum_{g\in[G]} \acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} \\ \frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} + \sum_{g,h\in[G]^2,g\neq h} \tilde{V}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right) \\ \frac{1}{\sqrt{\Upsilon}}\sum_{g,h\in[G]^2,g\neq h} \tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]} \end{pmatrix}$$
$$\rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right).$$

Let $v = (v_1, v_2, v_3)^{\top}$ with $\|v\|_2 = 1$, and

$$c \equiv (c_1, c_2, c_3)^{\top} = \begin{pmatrix} 1 & \rho_{1n} & 0 \\ \rho_{1n} & 1 & \rho_{2n} \\ 0 & \rho_{2n} & 1 \end{pmatrix}^{-1/2} (v_1, v_2, v_3)^{\top},$$

and note that $\|c\|_2$ is bounded for all n large enough. By Cramer-Wald device, it suffices to show that

$$c^{\top}\begin{pmatrix} \frac{1}{\sqrt{\Psi}}\sum_{g\in[G]} \acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} \\ \frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} + \sum_{g,h\in[G]^2,g\neq h} \tilde{V}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right) \\ \frac{1}{\sqrt{\Upsilon}}\sum_{g,h\in[G]^2,g\neq h} \tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]} \end{pmatrix} \rightsquigarrow \mathcal{N}(0,1).$$

Denote

$$M_n = \frac{c_1}{\sqrt{\Psi}}\sum_{g\in[G]} \acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} + \frac{c_2}{\sqrt{\Sigma}}\left(\sum_{g\in[G]} \hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]} + \sum_{g,h\in[G]^2,g\neq h} \tilde{V}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}\right)$$

$$+\frac{c_3}{\sqrt{\Upsilon}}\sum_{g,h\in[G]^2,g\neq h}\tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]},$$

and note that M_n can be written as sum of martingale difference array. To see this, define

$$M_{1G}=\frac{c_1}{\sqrt{\Psi}}\acute{\Pi}_{[1]}^{\top}\tilde{e}_{[1]}+\frac{c_2}{\sqrt{\Sigma}}\hat{\Pi}_{[1]}^{\top}\tilde{e}_{[1]},$$

and for $g\geq 2$ define

$$\begin{aligned}M_{gG}=&\frac{c_1}{\sqrt{\Psi}}\acute{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}+\frac{c_2}{\sqrt{\Sigma}}\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}+\frac{c_2}{\sqrt{\Sigma}}\sum_{h<g}\left(\tilde{V}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}+\tilde{V}_{[h]}^{\top}P_{[h,g]}\tilde{e}_{[g]}\right)\\&+\frac{c_3}{\sqrt{\Upsilon}}\sum_{h<g}\left(\tilde{e}_{[g]}^{\top}P_{[g,h]}\tilde{e}_{[h]}+\tilde{e}_{[h]}^{\top}P_{[h,g]}\tilde{e}_{[g]}\right),\end{aligned}$$

and we have $M_n=\sum_{g\in[G]}M_{gG}$. Further define $\varepsilon_{[g]}=\left(\tilde{e}_{[g]}^{\top},\tilde{V}_{[g]}^{\top}\right)^{\top}$ and the sequence of σ-fields $\mathcal{F}_{gG}=\{\varepsilon_{[1]},\cdots,\varepsilon_{[g]}\}$ such that $\mathcal{F}_{(g-1)G}\subset\mathcal{F}_{gG}$ with $\mathcal{F}_{0G}=\{\emptyset,\Omega\}$. It is clear that $\{M_{gG}\}_{g=1}^{G}$ is a sequence of martingale difference array with respect to $\{\mathcal{F}_{gG}\}_{g=1}^{G}$. Note that

$$\sum_{g\in[G]}\mathbb{E}(M_{gG}^2)=\mathbb{E}(M_n^2)=1$$

by the property of martingale difference array.

In order to show

$$\sum_{g\in[G]}M_{gG}\rightsquigarrow\mathcal{N}(0,1),$$

it remains to check the Lindeberg's condition

$$\sum_{g\in[G]}\mathbb{E}\left(M_{gG}^2\mathbf{1}_{\{|M_{gG}|\geq c\}}\right)\to 0$$

for every $c>0$, and the stability condition

$$\sum_{g\in[G]}\mathbb{E}\left(M_{gG}^2|\mathcal{F}_{(g-1)G}\right)\xrightarrow{p}\sum_{g\in[G]}\mathbb{E}(M_{gG}^2).$$

Step 2: Check Lindeberg's condition. It suffices to show that

$$\sum_{g\in[G]} \mathbb{E}(M_{gG}^{(2+\delta)}) \to 0$$

for some $\delta > 0$, and we will verify that for $\delta = 2$ in what follows. We have

$$\begin{aligned}\sum_{g\in G} \mathbb{E}(M_{gG}^4) \leqslant \frac{C}{\Psi^2} \sum_{g\in[G]} \mathbb{E}\left(\acute{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^4 + \frac{C}{\Sigma^2} \sum_{g\in[G]} \mathbb{E}\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^4 + \frac{C}{\Sigma^2} \sum_{g\in[G]} \mathbb{E}\left(\sum_{h<g} \tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)^4 \\ + \frac{C}{\Sigma^2} \sum_{g\in[G]} \mathbb{E}\left(\sum_{h<g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]}\right)^4 + \frac{C}{\Upsilon^2} \sum_{g\in[G]} \mathbb{E}\left(\sum_{h<g} \tilde{e}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]}\right)^4.\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\frac{1}{\Psi^2} \sum_{g\in[G]} \mathbb{E}\left(\acute{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^4 \\ &\leqslant \frac{C}{\Psi^2} \sum_{g\in[G]} \left(\acute{\Pi}_{[g]}^\top \acute{\Pi}_{[g]}\right)^2 \\ &\leqslant \frac{C \max_{g\in[G]} \left\|\acute{\Pi}_{[g]}\right\|_2^2 \acute{\Pi}^\top \acute{\Pi}}{\left(\acute{\Pi}^\top \acute{\Pi}\right)^2} \\ &= o(1).\end{aligned}$$

Here we use the fact that $\acute{\Pi}_{[g]} = z_{[g]} A_n z^\top \Pi$, so that

$$\begin{aligned}\frac{\max_{g\in[G]} \left\|\acute{\Pi}_{[g]}\right\|_2^2}{\acute{\Pi}^\top \acute{\Pi}} &= \frac{\max_{g\in[G]} \left\|\acute{\Pi}_{[g]}\right\|_2^2 / n}{\Pi^\top z A_n (z^\top z/n) A_n z^\top \Pi} \\ &\leq \frac{C \max_{g\in[G]} \Pi^\top z A_n (z_{[g]}^\top z_{[g]}/n) A_n z^\top \Pi}{\|A_n z^\top \Pi\|_2^2} \\ &\leq \frac{C \max_{g\in[G]} n_g \times \max_{i\in I_g, g\in[G]} \|z_{i,g}\|_2^2}{n} \\ &= o(1).\end{aligned}$$

For the second term, we have

$$\frac{1}{\Sigma^2} \sum_{g\in[G]} \mathbb{E}\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^4$$

$$
\begin{aligned}
&\leqslant \frac{C}{\Sigma^2} \sum_{g\in[G]} \left(\hat{\Pi}_{[g]}^\top \hat{\Pi}_{[g]}\right)^2 \\
&\leqslant \frac{C \max_{g\in[G]} \left\|\hat{\Pi}_{[g]}\right\|_2^2 \hat{\Pi}^\top \hat{\Pi}}{\left(\Pi^\top \Pi + K\right)^2} \\
&= o(1)
\end{aligned}
$$

as $K \to \infty$.

For the third term, we have

$$
\begin{aligned}
\frac{1}{\Sigma^2} \sum_{g\in[G]} \mathbb{E}\left(\sum_{h<g} \tilde{e}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]}\right)^4 \leqslant{}& \frac{C}{\Sigma^2} \sum_{g\in[G]} \sum_{h<g} \mathbb{E}\left(\tilde{e}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]}\right)^4 \\
&+ \frac{C}{\Sigma^2} \sum_{g\in[G]} \sum_{h,k<g,h\neq k} \mathbb{E}\left(\tilde{e}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]}\right)^2 \left(\tilde{e}_{[k]}^\top P_{[k,g]} \tilde{V}_{[g]}\right)^2,
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{1}{\Sigma^2} \sum_{g\in[G]} \sum_{h<g} \mathbb{E}\left(\tilde{e}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]}\right)^4 \\
&\leqslant \frac{1}{\Sigma^2} \sum_{g\in[G]} \sum_{h<g} \lambda_{\max}^2\left(P_{[h,g]} P_{[g,h]}\right) \\
&\leqslant \frac{C}{K^2} \sum_{g\in[G]} \sum_{h<g} \operatorname{trace}\left(P_{[h,g]} P_{[g,h]}\right) \\
&= o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{\Sigma^2} \sum_{g\in[G]} \sum_{h,k<g,h\neq k} \mathbb{E}\left(\tilde{e}_{[h]}^\top P_{[h,g]} \tilde{V}_{[g]}\right)^2 \left(\tilde{e}_{[k]}^\top P_{[k,g]} \tilde{V}_{[g]}\right)^2 \\
&\leqslant \frac{C}{\Sigma^2} \sum_{g\in[G]} \sum_{h,k<g,h\neq k} \lambda_{\max}\left(P_{[g,h]} P_{[h,g]}\right) \lambda_{\max}\left(P_{[g,k]} P_{[k,g]}\right) \\
&\leqslant \frac{C}{\Sigma^2} \sum_{g\in[G]} \sum_{h,k<g,h\neq k} \operatorname{trace}\left(P_{[g,h]} P_{[h,g]}\right) \operatorname{trace}\left(P_{[g,k]} P_{[k,g]}\right) \\
&\leqslant \frac{C}{K^2} \sum_{g\in[G]} \sum_{h<g} \operatorname{trace}\left(P_{[g,h]} P_{[h,g]}\right) \\
&= o(1).
\end{aligned}
$$

The last two terms can be handled similarly.

Step 3: Check stability condition. The variance and conditional variance can be written as

$$\mathbb{E}\left(M_{1G}^2\right) = \mathbb{E}\left(M_{1G}^2|\mathcal{F}_{0G}\right) = \frac{c_1^2}{\Psi}\acute{\Pi}_{[1]}^\top\Omega_1^{\tilde{e},\tilde{e}}\acute{\Pi}_{[1]} + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\hat{\Pi}_{[1]}^\top\Omega_1^{\tilde{e},\tilde{e}}\acute{\Pi}_{[1]} + \frac{c_2^2}{\Sigma}\hat{\Pi}_{[1]}^\top\Omega_1^{\tilde{e},\tilde{e}}\hat{\Pi}_{[1]},$$

and for $g \geq 2$,

$$\begin{aligned}
\mathbb{E}\left(M_{gG}^2\right) = {} & \frac{c_1^2}{\Psi}\acute{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]} + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\hat{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]} + \frac{c_2^2}{\Sigma}\hat{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\hat{\Pi}_{[g]} \\
& + \frac{c_2^2}{\Sigma}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,h]}\right) + \frac{c_2^2}{\Sigma}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right) \\
& + \frac{4c_3^2}{\Upsilon}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right) + \frac{2c_2^2}{\Sigma}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{V}}P_{[g,h]}\right) \\
& + \frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,h]}\right) + \frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\sum_{h<g}\text{trace}\left(\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(M_{gG}^2|\mathcal{F}_{(g-1)G}\right) = {} & \frac{c_1^2}{\Psi}\acute{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]} + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\hat{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]} + \frac{c_2^2}{\Sigma}\hat{\Pi}_{[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}\hat{\Pi}_{[g]} \\
& + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]} + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]} \\
& + \frac{4c_1c_3}{\sqrt{\Psi\Upsilon}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]} + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\hat{\Pi}_{[g]} \\
& + \frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}\hat{\Pi}_{[g]} + \frac{4c_1c_3}{\sqrt{\Psi\Upsilon}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\hat{\Pi}_{[g]} \\
& + \frac{c_2^2}{\Sigma}\sum_{h,k<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,k]}\tilde{e}_{[k]} + \frac{c_2^2}{\Sigma}\sum_{h,k<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]} \\
& + \frac{4c_3^2}{\Upsilon}\sum_{h,k<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]} + \frac{2c_2^2}{\Sigma}\sum_{h,k<g}e_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]} \\
& + \frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\sum_{h,k<g}V_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]} + \frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\sum_{h,k<g}e_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]}.
\end{aligned}$$

We thus obtain

$$
\begin{aligned}
&\mathbb{E}\left(M_{gG}^2|\mathcal{F}_{(g-1)G}\right)-\mathbb{E}\left(M_{gG}^2\right)\\
&=\left\{\begin{array}{c}\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]}+\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}\acute{\Pi}_{[g]}+\frac{4c_1c_3}{\sqrt{\Psi\Upsilon}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]}\\+\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\hat{\Pi}_{[g]}+\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{h<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}\hat{\Pi}_{[g]}+\frac{4c_1c_3}{\sqrt{\Psi\Upsilon}}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\hat{\Pi}_{[g]}\end{array}\right\}\\
&+\left\{\begin{array}{c}\frac{c_2^2}{\Sigma}\left[\sum_{h,k<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,k]}\tilde{e}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,h]}\right)\right]\\+\frac{c_2^2}{\Sigma}\left[\sum_{h,k<g}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right]\\+\frac{4c_3^2}{\Upsilon}\left[\sum_{h,k<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right]\\+\frac{2c_2^2}{\Sigma}\left[\sum_{h,k<g}e_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{V}}P_{[g,h]}\right)\right]\\+\frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\left[\sum_{h,k<g}V_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,h]}\right)\right]\\+\frac{4c_2c_3}{\sqrt{\Sigma\Upsilon}}\left[\sum_{h,k<g}e_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}P_{[g,k]}\tilde{e}_{[k]}-\sum_{h<g}\mathrm{trace}\left(\Omega_h^{\tilde{e},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right]\end{array}\right\}\\
&\equiv M_{gG}^{(1)}+M_{gG}^{(2)},
\end{aligned}
$$

and it suffices to show that

$$
\begin{aligned}
M_n^{(1)}&\equiv\sum_{g\in[G]}M_{gG}^{(1)}=o_P(1),\\
M_n^{(2)}&\equiv\sum_{g\in[G]}M_{gG}^{(2)}=o_P(1).
\end{aligned}
$$

Consider first $M_n^{(1)}$, we shall only compute the sum of the first term in $M_{gG}^{(1)}$, as the other terms can be handled similarly. Recall that $\tilde{P}$ is the block lower triangular matrix corresponding to $P-\bar{P}$, we have

$$
\mathbb{E}\left(\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{g\in[G]}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]}\right)=0,
$$

and

$$
\begin{aligned}
\mathbb{V}\left(\frac{2c_1c_2}{\sqrt{\Psi\Sigma}}\sum_{g\in[G]}\sum_{h<g}\tilde{e}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{V},\tilde{e}}\acute{\Pi}_{[g]}\right)&=\frac{4c_1^2c_2^2}{\Psi\Sigma}\mathbb{E}\left(\tilde{e}^\top\tilde{P}^\top\Omega_{\tilde{V},\tilde{e}}\acute{\Pi}\right)^2\\
&\leqslant\frac{C\lambda_{\max}\left(\tilde{P}\tilde{P}^\top\right)\lambda_{\max}\left(\Omega_{\tilde{e},\tilde{V}}\Omega_{\tilde{V},\tilde{e}}\right)\acute{\Pi}^\top\acute{\Pi}}{(\acute{\Pi}^\top\acute{\Pi})K}\\
&\leqslant\frac{C\lambda_{\max}\left(\tilde{P}\tilde{P}^\top\right)}{K}
\end{aligned}
$$

$$\leqslant \frac{C\left\|\tilde{P}\tilde{P}^{\top}\right\|_F}{K}$$
$$= o(1),$$

where we use Lemma A.2 in the last equality.

Now consider $M_n^{(2)}$, we shall only compute the sum of the first term in $M_{gG}^{(2)}$, as the other terms can be handled similarly. We have

$$\mathbb{E}\left(\frac{c_2^2}{\Sigma}\sum_{g\in[G]}\left[\sum_{h,k<g}\tilde{e}_{[h]}^{\top}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,k]}\tilde{e}_{[k]}-\sum_{h<g}\operatorname{trace}\left(\Omega_h^{\tilde{e},\tilde{e}}P_{[h,g]}\Omega_g^{\tilde{V},\tilde{V}}P_{[g,h]}\right)\right]\right)=0,$$

and

$$\begin{aligned}
&\mathbb{V}\left(\frac{c_2^2}{\Sigma}\sum_{g\in[G]}\left[\sum_{h,k<g}\tilde{V}_{[h]}^{\top}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}-\sum_{h<g}\operatorname{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right]\right)\\
\leqslant\ &\frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g\in[G]}\sum_{h<g}\left(\tilde{V}_{[h]}^{\top}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\tilde{V}_{[h]}-\operatorname{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right)\right)^2\\
&+\frac{C}{\Sigma^2}\mathbb{E}\left(\sum_{g\in[G]}\sum_{h,k<g,h\neq k}\tilde{V}_{[h]}^{\top}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}\right)^2.
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g\in[G]}\sum_{h<g}\left(\tilde{V}_{[h]}^{\top}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\tilde{V}_{[h]}-\operatorname{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right)\right)^2\\
=\ &\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{h\in[G]}\tilde{V}_{[h]}^{\top}\left(\sum_{g>h}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\tilde{V}_{[h]}-\operatorname{trace}\left(\Omega_h^{\tilde{V},\tilde{V}}\left(\sum_{g>h}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right)\right)^2\\
\leqslant\ &\frac{1}{\Sigma^2}\sum_{h\in[G]}\mathbb{E}\left(\tilde{V}_{[h]}^{\top}\left(\sum_{g>h}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\tilde{V}_{[h]}\right)^2\\
\leqslant\ &\frac{1}{\Sigma^2}\sum_{h\in[G]}\lambda_{\max}^2\left(\left(\tilde{P}^{\top}\Omega_{\tilde{e}}\tilde{P}\right)_{[h,h]}\right)\\
\leqslant\ &\frac{C}{\Sigma^2}\sum_{h\in[G]}\operatorname{trace}\left(\left(\tilde{P}^{\top}\Omega_{\tilde{e}}\tilde{P}\right)_{[h,h]}\right)\\
\leqslant\ &\frac{C}{\Sigma^2}\operatorname{trace}\left(\tilde{P}^{\top}\Omega_{\tilde{e}}\tilde{P}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \frac{C}{K^2}\text{trace}\left(\tilde{P}^\top \tilde{P}\right) \\
&= o(1),
\end{aligned}
$$

where we use the fact that

$$
\text{trace}\left(\tilde{P}^\top \tilde{P}\right) = ||\tilde{P}||_F^2 \leq ||P||_F^2 = O(K).
$$

For the second term, we have

$$
\begin{aligned}
&\frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{g\in[G]}\sum_{h,k<g,h\neq k}\tilde{V}_{[h]}^\top P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\tilde{V}_{[k]}\right)^2 \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{h,k\in[G]^2,h\neq k}\tilde{V}_{[h]}^\top\left(\sum_{g>h\vee k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right)^2 \\
&= \frac{1}{\Sigma^2}\mathbb{E}\left(\sum_{h,k\in[G]^2,h<k}\tilde{V}_{[h]}^\top\left(\sum_{g>k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right. \\
&\qquad \left. + \sum_{h,k\in[G]^2,h>k}\tilde{V}_{[h]}^\top\left(\sum_{g>h}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{h,k\in[G]^2,h\neq k}\mathbb{E}\left(\tilde{V}_{[h]}^\top\left(\sum_{g>h\vee k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\tilde{V}_{[k]}\right)^2 \\
&\leqslant \frac{C}{\Sigma^2}\sum_{h,k\in[G]^2,h\neq k}\text{trace}\left(\left(\sum_{g>h\vee k}P_{[h,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,k]}\right)\left(\sum_{g>h\vee k}P_{[k,g]}\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\right)\right) \\
&\leqslant \frac{C}{\Sigma^2}\sum_{h,k\in[G]^2}\text{trace}\left(\left(\tilde{P}^\top\Omega_{\tilde{e}}\tilde{P}\right)_{[h,k]}\left(\tilde{P}^\top\Omega_{\tilde{e}}\tilde{P}\right)_{[k,h]}\right) \\
&\leqslant \frac{C}{K^2}\text{trace}\left(\tilde{P}^\top\Omega_{\tilde{e}}\tilde{P}\tilde{P}^\top\Omega_{\tilde{e}}\tilde{P}\right) \\
&\leqslant \frac{C\left\|\tilde{P}\tilde{P}^\top\right\|_F^2}{K^2} \\
&= o(1),
\end{aligned}
$$

where we use Lemma A.2 in the last equality. This concludes the proof. □

C.11 Proof of Lemma A.11

For the first result, for $j \in [d_z]$, let v_j be the d_z-dimensional unit vector with j-th element one and other elements zero, and denote $\bar{z}_j = z\Omega^{-1/2}v_j$, then

$$\bar{z}_j^\top \bar{z}_j = v_j^\top (\Omega/n)^{-1/2}(z^\top z/n)(\Omega/n)^{-1/2}v_j = O(1),$$
$$\max_{g\in[G]} \bar{z}_{j,[g]}^\top \bar{z}_{j,[g]} = \max_{g\in[G]} v_j^\top (\Omega/n)^{-1/2}(z_{[g]}^\top z_{[g]}/n)(\Omega/n)^{-1/2}v_j = o(1).$$

It follows that

$$\left|\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\mathbb{E}\left[\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right]\right| \le \left(\sum_{g\in[G]}\mathbb{E}\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)^2\right)^{1/2}\left(\frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^2\right)^{1/2} = O(1),$$

and as d_z is fixed, we obtain

$$\frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\mathbb{E}\left[\left(z_{[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right] = O(1). \tag{C.32}$$

In addition, let $\hat{V} = M_W(P-\bar{P})V = Q\tilde{V}$, we have

$$\begin{aligned}\frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right)\right] &= \frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \hat{e}_{[g]}\right)\right] \\ &\quad + \frac{1}{\sqrt{\Sigma}}\Omega^{-1/2}\sum_{g\in[G]}\left[\left(z_{[g]}^\top \hat{e}_{[g]}\right)\left(\hat{V}_{[g]}^\top \hat{e}_{[g]}\right)\right].\end{aligned}$$

Thus, it suffices to show that, for any j,

$$\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left[\left(\bar{z}_{j,[g]}^\top \hat{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \hat{e}_{[g]}\right)\right] - \frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\mathbb{E}\left[\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right] = o_P(1), \tag{C.33}$$
$$\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left[\left(\bar{z}_{j,[g]}^\top \hat{e}_{[g]}\right)\left(\hat{V}_{[g]}^\top \hat{e}_{[g]}\right)\right] = o_P(1). \tag{C.34}$$

For (C.33), the left-hand side can be written as

$$\frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^\top \hat{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \hat{e}_{[g]}\right) - \sum_{g\in[G]}\mathbb{E}\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)\right)$$

$$= \underbrace{\frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)-\sum_{g\in[G]}\mathbb{E}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\right)}_{R_{27}}$$
$$+ \underbrace{\frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}e_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}e_{[g]}\right)-\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\right)}_{R_{28}}$$
$$+ \underbrace{\frac{1}{\sqrt{\Sigma}}\left(\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\hat{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\hat{e}_{[g]}\right)-\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}e_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}e_{[g]}\right)\right)}_{R_{29}}.$$

For R_{27}, it has mean zero and

$$\begin{aligned}
\mathbb{V}\left(R_{27}\right) &= \frac{1}{\Sigma}\mathbb{E}\left(\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)-\sum_{g\in[G]}\mathbb{E}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\right)^{2} \\
&\leqslant \frac{1}{\Sigma}\sum_{g\in[G]}\mathbb{E}\left(\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)\right)^{2} \\
&\leqslant \frac{C\max_{g\in[G]}\left\|\hat{\Pi}_{[g]}\right\|_{2}^{2}\bar{z}_{j}^{\top}\bar{z}_{j}}{(\Pi^{\top}\Pi+K)} \\
&= o(1).
\end{aligned}$$

For R_{28}, we have

$$\begin{aligned}
R_{28} = &\underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\hat{\Pi}_{[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)}_{R_{28,1}} \\
&-\underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\hat{\Pi}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{28,2}} \\
&-\underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{\Pi}_{[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)}_{R_{28,3}},
\end{aligned}$$

where

$$|R_{28,1}| \leqslant \max_{g\in[G]} \left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2^2 \times \sqrt{\bar{z}_j^\top \bar{z}_j} \times \sqrt{\frac{\hat{\Pi}^\top\hat{\Pi}}{\Sigma}} = o_P(1),$$
$$|R_{28,2}| \leqslant \max_{g\in[G]} \left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2 \times \sqrt{\bar{z}_j^\top \bar{z}_j} \times \sqrt{\frac{1}{\Sigma}\sum_{g\in[G]}\left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]}\right)^2} = o_P(1),$$
$$|R_{28,3}| \leqslant \max_{g\in[G]} \left\|W_{[g]}\hat{\gamma}_{\tilde{e}}\right\|_2 \times \sqrt{\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)^2} \times \sqrt{\frac{\hat{\Pi}^\top\hat{\Pi}}{\Sigma}} = o_P(1),$$

whence $R_{28} = o_P(1)$. For R_{29}, we have

$$R_{29} = \hat{\Delta}^2 \times \underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^\top X_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top X_{[g]}\right)}_{R_{29,1}}$$
$$- \hat{\Delta} \times \underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^\top X_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top e_{[g]}\right)}_{R_{29,2}}$$
$$- \hat{\Delta} \times \underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^\top e_{[g]}\right)\left(\hat{\Pi}_{[g]}^\top e_{[g]}\right)}_{R_{29,3}},$$

where $\hat{\Delta} = o_P(1)$ and

$$|R_{29,1}| \leqslant \sqrt{\sum_{g\in[G]} \bar{z}_{j,[g]}^\top \bar{z}_{j,[g]} X_{[g]}^\top X_{[g]}} \times \sqrt{\frac{1}{\Sigma}\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \hat{\Pi}_{[g]} X_{[g]}^\top X_{[g]}} = O_P(1),$$
$$|R_{29,2}| \leqslant \sqrt{\sum_{g\in[G]} \bar{z}_{j,[g]}^\top \bar{z}_{j,[g]} X_{[g]}^\top X_{[g]}} \times \sqrt{\frac{1}{\Sigma}\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \hat{\Pi}_{[g]} e_{[g]}^\top e_{[g]}} = O_P(1),$$
$$|R_{29,3}| \leqslant \sqrt{\sum_{g\in[G]} \bar{z}_{j,[g]}^\top \bar{z}_{j,[g]} e_{[g]}^\top e_{[g]}} \times \sqrt{\frac{1}{\Sigma}\sum_{g\in[G]} \hat{\Pi}_{[g]}^\top \hat{\Pi}_{[g]} e_{[g]}^\top e_{[g]}} = O_P(1),$$

since $\max_{g\in[G]} \mathbb{E}\left(X_{[g]}^\top X_{[g]}\right) = O(1)$ and $\max_{g\in[G]} \mathbb{E}\left(e_{[g]}^\top e_{[g]}\right) = O(1)$ by Lemma A.1. It follows that $R_{29} = o_P(1)$.

For (C.34), the left-hand side can be written as

$$\begin{aligned}
&\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\hat{e}_{[g]}\right)\left(\hat{V}_{[g]}^{\top}\hat{e}_{[g]}\right)\\
&=\frac{\hat{\Delta}^2}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}X_{[g]}\right)\left(\hat{V}_{[g]}^{\top}X_{[g]}\right)-\frac{\hat{\Delta}}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}X_{[g]}\right)\left(\hat{V}_{[g]}^{\top}e_{[g]}\right)\\
&-\frac{\hat{\Delta}}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}e_{[g]}\right)\left(\hat{V}_{[g]}^{\top}X_{[g]}\right)+\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}e_{[g]}\right)\left(\hat{V}_{[g]}^{\top}e_{[g]}\right).
\end{aligned}$$

Here we only show that the last term is $o_P(1)$, which is most difficult since it does not involve $\hat{\Delta}$. We have

$$\begin{aligned}
\underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}e_{[g]}\right)\left(\hat{V}_{[g]}^{\top}e_{[g]}\right)}_{R_{30}}=&\underbrace{\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{V}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{30,1}}\\
&\underbrace{-\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\hat{V}_{[g]}^{\top}\tilde{e}_{[g]}\right)}_{R_{30,2}}\\
&\underbrace{-\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}\tilde{e}_{[g]}\right)\left(\hat{V}_{[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)}_{R_{30,3}}\\
&\underbrace{+\frac{1}{\sqrt{\Sigma}}\sum_{g\in[G]}\left(\bar{z}_{j,[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)\left(\hat{V}_{[g]}^{\top}W_{[g]}\hat{\gamma}_{\tilde{e}}\right)}_{R_{30,4}}.
\end{aligned}$$

Note that

$$\begin{aligned}
\hat{V}_{[g]}=&\sum_{h\in[G],h\neq g}P_{[g,h]}\tilde{V}_{[h]}+\sum_{h\in[G]}P_{W,[g,h]}P_{[h,h]}\tilde{V}_{[h]}\\
&+\sum_{h\in[G]}P_{[g,g]}P_{W,[g,h]}\tilde{V}_{[h]}-\sum_{h\in[G]}\left(\sum_{k\in[G]}P_{W,[g,k]}P_{[k,k]}P_{W,[k,h]}\right)\tilde{V}_{[h]}.
\end{aligned}$$

Therefore, we have

$$\frac{1}{\Sigma}\sum_{g\in[G]}\left(\hat{V}_{[g]}^{\top}\tilde{e}_{[g]}\right)^2\leq\frac{C}{\Sigma}\sum_{g\in[G]}\left(\sum_{h\in[G],h\neq g}\tilde{V}_{[h]}^{\top}P_{[h,g]}\tilde{e}_{[g]}\right)^2$$

$$
\begin{aligned}
&+ \frac{C}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top P_{[h,h]} P_{W,[h,g]} \tilde{e}_{[g]} \right)^2 \\
&+ \frac{C}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top P_{W,[h,g]} P_{[g,g]} \tilde{e}_{[g]} \right)^2 \\
&+ \frac{C}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top \left(\sum_{k\in[G]} P_{W,[h,k]} P_{[k,k]} P_{W,[k,g]} \right) \tilde{e}_{[g]} \right)^2 \\
&= \frac{C}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 + o_P(1) \\
&= \frac{C}{\Sigma} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G],h\neq g} \tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right)^2 + o_P(1) \\
&= O_P(1),
\end{aligned}
$$

by Lemma A.4 and the facts that

$$
\begin{aligned}
\frac{1}{\Sigma} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top P_{[h,h]} P_{W,[h,g]} \tilde{e}_{[g]} \right)^2 &= \frac{1}{\Sigma} \sum_{g,h\in[G]^2} \mathbb{E} \left(\tilde{V}_{[h]}^\top P_{[h,h]} P_{W,[h,g]} \tilde{e}_{[g]} \right)^2 \\
&\leqslant \frac{C \operatorname{trace}\left(P_W \bar{P}^2 P_W \right)}{\Sigma} \\
&= o(1),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\Sigma} \sum_{g\in[G]} \mathbb{E} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top P_{W,[h,g]} P_{[g,g]} \tilde{e}_{[g]} \right)^2 &= \frac{1}{\Sigma} \sum_{g,h\in[G]^2} \mathbb{E} \left(\tilde{V}_{[h]}^\top P_{W,[h,g]} P_{[g,g]} \tilde{e}_{[g]} \right)^2 \\
&\leqslant \frac{C \operatorname{trace}\left(\bar{P} P_W^2 \bar{P} \right)}{\Sigma} \\
&= o(1),
\end{aligned}
$$

and

$$
\frac{1}{\Sigma} \sum_{g\in[G]} \left(\sum_{h\in[G]} \tilde{V}_{[h]}^\top \left(\sum_{k\in[G]} P_{W,[h,k]} P_{[k,k]} P_{W,[k,g]} \right) \tilde{e}_{[g]} \right)^2
$$

$$
\begin{aligned}
&= \frac{1}{\Sigma} \sum_{g,h \in [G]^2} \left(\tilde{V}_{[h]}^\top \left(\sum_{k \in [G]} P_{W,[h,k]} P_{[k,k]} P_{W,[k,g]} \right) \tilde{e}_{[g]} \right)^2 \\
&\leqslant \frac{C \operatorname{trace}\left(P_W \bar{P} P_W^2 \bar{P} P_W \right)}{\Sigma} \\
&= o(1).
\end{aligned}
$$

This implies

$$
|R_{30,2}| \leqslant \max_{g \in [G]} \left\| W_{[g]} \hat{\gamma}_{\tilde{e}} \right\|_2 \times \sqrt{\bar{z}_j^\top \bar{z}_j} \times \sqrt{\frac{1}{\Sigma} \sum_{g \in [G]} \left(\hat{V}_{[g]}^\top \tilde{e}_{[g]} \right)^2} = o_P(1).
$$

In addition, we have

$$
\begin{aligned}
|R_{30,3}| &\leqslant \max_{g \in [G]} \left\| W_{[g]} \hat{\gamma}_{\tilde{e}} \right\|_2 \times \sqrt{\sum_{g \in [G]} \left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]} \right)^2} \times \sqrt{\frac{\tilde{V}^\top Q^2 \tilde{V}}{\Sigma}} = o_P(1), \\
|R_{30,4}| &\leqslant \max_{g \in [G]} \left\| W_{[g]} \hat{\gamma}_{\tilde{e}} \right\|_2^2 \times \sqrt{\bar{z}_j^\top \bar{z}_j} \times \sqrt{\frac{\tilde{V}^\top Q^2 \tilde{V}}{\Sigma}} = o_P(1).
\end{aligned}
$$

It follows that

$$
R_{30} = R_{30,1} + o_P(1),
$$

and for $R_{30,1}$, note that similar to the proof above, we also have

$$
R_{30,1} = \frac{1}{\sqrt{\Sigma}} \sum_{g,h \in [G]^2 g \neq h} \left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) + o_P(1),
$$

where

$$
\mathbb{E}\left(\frac{1}{\sqrt{\Sigma}} \sum_{g,h \in [G]^2 g \neq h} \left(z_{[g]}^\top \tilde{e}_{[g]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \right) = 0,
$$

and

$$
\mathbb{V}\left(\frac{1}{\sqrt{\Sigma}} \sum_{g,h \in [G]^2 g \neq h} \left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]} \right) \left(\tilde{V}_{[h]}^\top P_{[h,g]} \tilde{e}_{[g]} \right) \right)
$$

$$
\begin{aligned}
&= \frac{1}{\Sigma}\mathbb{E}\left(\sum_{g,h\in[G]^2 g\neq h}\left(\bar{z}_{j,[g]}^\top \tilde{e}_{[g]}\right)\left(\tilde{V}_{[h]}^\top P_{[h,g]}\tilde{e}_{[g]}\right)\right)^2 \\
&\leqslant \frac{C}{\Sigma}\mathbb{E}\left(\sum_{g,h\in[G]^2 g\neq h}\bar{z}_{j,[g]}^\top\left(\tilde{e}_{[g]}\tilde{e}_{[g]}^\top - \Omega_g^{\tilde{e},\tilde{e}}\right)P_{[g,h]}\tilde{V}_{[h]}\right)^2 \\
&+ \frac{C}{\Sigma}\mathbb{E}\left(\sum_{g,h\in[G]^2 g\neq h}\bar{z}_{j,[g]}^\top\Omega_g^{\tilde{e},\tilde{e}}P_{[g,h]}\tilde{V}_{[h]}\right)^2 \\
&\leq \frac{C}{\Sigma}\sum_{g,h\in[G]^2 g\neq h}\mathbb{E}\left(\bar{z}_{j,[g]}^\top\tilde{e}_{[g]}\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{V}_{[h]}\right)^2 \\
&+ \frac{C}{\Sigma}\mathbb{E}\left(\bar{z}_j^\top\Omega_{\tilde{e}}(P-\bar{P})\tilde{V}\right)^2 \\
&\leq \frac{C\max_{g\in[G]}\left\|\bar{z}_{j,[g]}\right\|_2^2}{\Sigma}\sum_{g,h\in[G]^2 g\neq h}\operatorname{trace}\left(P_{[g,h]}P_{[h,g]}\right) + \frac{C\bar{z}_j^\top\bar{z}_j}{\Sigma} \\
&= o(1).
\end{aligned}
$$

For the second result, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{\Sigma\Upsilon}}\left[\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\right. \\
&\qquad\left. - \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\right] \\
&= \frac{1}{\sqrt{\Sigma\Upsilon}}\sum_{g,h\in[G]^2,g\neq h}\left(\Pi_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right) \\
&+ \frac{1}{\sqrt{\Sigma\Upsilon}}\left[\sum_{g,h\in[G]^2,g\neq h}\left(V_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\right. \\
&\qquad\left. - \sum_{g,h\in[G]^2,g\neq h}\mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]}\tilde{e}_{[h]}\right)\right]. \qquad \text{(C.35)}
\end{aligned}
$$

The first term on the RHS of (C.35) can be written as

$$
\frac{1}{\sqrt{\Sigma\Upsilon}}\sum_{g,h\in[G]^2,g\neq h}\left(\Pi_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^\top P_{[g,h]}\hat{e}_{[h]}\right)
$$

$$= \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)}_{R_{31}}$$

$$+ \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} e_{[h]}\right)\left(e_{[g]}^\top P_{[g,h]} e_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right)\left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \right)}_{R_{32}}$$

$$+ \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} e_{[h]}\right)\left(e_{[g]}^\top P_{[g,h]} e_{[h]}\right) \right)}_{R_{33}}.$$

By using the same argument as in the proof of Lemma A.6, we can show that $R_{31} = o_P(1)$ and $R_{32} = o_P(1)$. For R_{33}, we have

$$\begin{aligned}
R_{33} = &-\hat{\Delta}^3 \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} X_{[h]}\right)\left(X_{[g]}^\top P_{[g,h]} X_{[h]}\right)}_{R_{33,1}} \\
&+ \hat{\Delta}^2 \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} e_{[h]}\right)\left(X_{[g]}^\top P_{[g,h]} X_{[h]}\right)}_{R_{33,2}} \\
&+ \hat{\Delta}^2 \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} X_{[h]}\right)\left(e_{[g]}^\top P_{[g,h]} X_{[h]}\right)}_{R_{33,3}} \\
&+ \hat{\Delta}^2 \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} X_{[h]}\right)\left(X_{[g]}^\top P_{[g,h]} e_{[h]}\right)}_{R_{33,4}} \\
&- \hat{\Delta} \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} X_{[h]}\right)\left(e_{[g]}^\top P_{[g,h]} e_{[h]}\right)}_{R_{33,5}} \\
&- \hat{\Delta} \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} e_{[h]}\right)\left(X_{[g]}^\top P_{[g,h]} e_{[h]}\right)}_{R_{33,6}} \\
&- \hat{\Delta} \times \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \sum_{g,h\in[G]^2, g\neq h} \left(\Pi_{[g]}^\top P_{[g,h]} e_{[h]}\right)\left(e_{[g]}^\top P_{[g,h]} X_{[h]}\right)}_{R_{33,7}},
\end{aligned}$$

and by using a similar argument as in the proof of Lemma A.6, we can show that

$$R_{33,i} = O_P(1), \quad i = 1, \ldots, 7,$$

which implies that $R_{33} = o_P(1)$.

The second term on the RHS of (C.35) can be written as

$$\begin{aligned}
&\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(V_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right) \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \right) \\
&= \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \mathbb{E}\left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \right)}_{R_{34}} \\
&+ \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(V_{[g]}^\top P_{[g,h]} e_{[h]}\right) \left(e_{[g]}^\top P_{[g,h]} e_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]}\right) \right)}_{R_{35}} \\
&+ \underbrace{\frac{1}{\sqrt{\Sigma\Upsilon}} \left(\sum_{g,h\in[G]^2, g\neq h} \left(V_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right) \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]}\right) - \sum_{g,h\in[G]^2, g\neq h} \left(V_{[g]}^\top P_{[g,h]} e_{[h]}\right) \left(e_{[g]}^\top P_{[g,h]} e_{[h]}\right) \right)}_{R_{36}}.
\end{aligned}$$

By using the same argument as in the proof of Lemma A.6, we can show that $R_{34} = o_P(1)$ and $R_{35} = o_P(1)$. In addition, we can show that $R_{36} = o_P(1)$ as in the proof for R_{33}. This concludes the proof. □

C.12 Proof of Lemma A.12

If Assumptions 1–4 hold, then by Lemma A.8, we have $\hat{\beta} \xrightarrow{p} \beta$. Consider first $\hat{\rho}_1$. It suffices to show that $\hat{\rho}_1 - \rho_{1n} = o_P(1)$, and we note that $\hat{\rho}_1$ can be written as

$$\begin{aligned}
\hat{\rho}_1 &= \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}} \sum_{g\in[G]} \left[\left(\acute{X}_{[g]}^\top \hat{e}_{[g]}\right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right) \right] \\
&= \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}} \sum_{g\in[G]} \left[\left((z\hat{A}_n z^\top X)_{[g]}^\top \hat{e}_{[g]}\right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right) \right] \\
&= \frac{1}{\sqrt{\hat{\Psi}\hat{\Sigma}}} X^\top z \hat{A}_n \sum_{g\in[G]} \left[\left(z_{[g]}^\top \hat{e}_{[g]}\right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]}\right) \right]
\end{aligned}$$

$$= \sqrt{\frac{\Sigma}{\hat{\Sigma}}} \times \frac{1}{\sqrt{X^\top z \hat{A}_n \hat{\Omega} \hat{A}_n z^\top X}} X^\top z \hat{A}_n \hat{\Omega}^{1/2}$$
$$\times \hat{\Omega}^{-1/2} \Omega^{1/2} \times \frac{1}{\sqrt{\Sigma}} \Omega^{-1/2} \sum_{g \in [G]} \left[\left(z_{[g]}^\top \hat{e}_{[g]} \right) \left(\hat{X}_{[g]}^\top \hat{e}_{[g]} \right) \right].$$

By Lemma A.7, we have

$$\sqrt{\frac{\Sigma}{\hat{\Sigma}}} = 1 + o_P(1),$$
$$\hat{\Omega}^{-1/2} \Omega^{1/2} = I_{d_z} + o_P(1),$$

and by (C.17) and (C.18) we have

$$\begin{aligned}
&\frac{1}{\sqrt{X^\top z \hat{A}_n \hat{\Omega} \hat{A}_n z^\top X}} X^\top z \hat{A}_n \hat{\Omega}^{1/2} \\
&= \frac{1}{\sqrt{(X^\top z / r_n)(\hat{A}_n / \lambda_n)(\hat{\Omega}/n)(\hat{A}_n/\lambda_n)(z^\top X / r_n)}} (X^\top z / r_n)(\hat{A}_n/\lambda_n)(\hat{\Omega}/n)^{1/2} \\
&= \frac{1}{\sqrt{(\Pi^\top z / r_n)(A_n / \lambda_n)(\Omega/n)(A_n/\lambda_n)(z^\top \Pi / r_n)}} (\Pi^\top z / r_n)(A_n/\lambda_n)(\Omega/n)^{1/2} + o_P(1) \\
&= \frac{1}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} \Pi^\top z A_n \Omega^{1/2} + o_P(1).
\end{aligned}$$

Further recall that

$$\begin{aligned}
\rho_{1n} &= \frac{1}{\sqrt{\Psi \Sigma}} \sum_{g \in [G]} \mathbb{E} \left[\left(\acute{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \right] \\
&= \frac{1}{\sqrt{\Pi^\top z A_n \Omega A_n z^\top \Pi}} \Pi^\top z A_n \Omega^{1/2} \times \frac{1}{\sqrt{\Sigma}} \Omega^{-1/2} \sum_{g \in [G]} \mathbb{E} \left[\left(z_{[g]}^\top \tilde{e}_{[g]} \right) \left(\hat{\Pi}_{[g]}^\top \tilde{e}_{[g]} \right) \right].
\end{aligned}$$

The consistency of $\hat{\rho}_1$ then follows by (C.32) and Lemma A.11.

Next, consider $\hat{\rho}_2$. It suffices to show that $\hat{\rho}_2 - \rho_{2n} = o_P(1)$, we have

$$\begin{aligned}
\hat{\rho}_2 - \rho_{2n} = \frac{2}{\sqrt{\Sigma \Upsilon}} \Bigg[&\sum_{g,h \in [G]^2, g \neq h} \left(X_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \left(\hat{e}_{[g]}^\top P_{[g,h]} \hat{e}_{[h]} \right) \\
&- \sum_{g,h \in [G]^2, g \neq h} \mathbb{E} \left(\tilde{V}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \left(\tilde{e}_{[g]}^\top P_{[g,h]} \tilde{e}_{[h]} \right) \Bigg]
\end{aligned}$$

$$+\left(\sqrt{\frac{\Sigma\Upsilon}{\hat{\Sigma}\hat{\Upsilon}}}-1\right)\times\left(\frac{2}{\sqrt{\Sigma\Upsilon}}\sum_{g,h\in[G]^2,g\neq h}\left(X_{[g]}^{\top}P_{[g,h]}\hat{e}_{[h]}\right)\left(\hat{e}_{[g]}^{\top}P_{[g,h]}\hat{e}_{[h]}\right)\right).$$

The consistency of $\hat{\rho}_2$ then follows by Lemmas A.7 and A.11.

Lastly, consider $\hat{\alpha}_1$ and $\hat{\alpha}_2$. If Assumptions 1-3 hold with $\Pi^{\top}\Pi/\sqrt{K}\to\infty$, then by (C.20), (C.26) and Lemma A.7, we have $\widehat{\Phi}_1/\Phi_1\xrightarrow{p}1$ and $\widehat{\Phi}_2/\Phi_2\xrightarrow{p}1$. Therefore, if the assumptions for d_n in Assumption 4 hold, then by the continuous mapping theorem, we have $\hat{\alpha}_1\xrightarrow{p}\alpha_1$ and $\hat{\alpha}_2\xrightarrow{p}\alpha_2$. Alternatively, if Assumption 1-3 hold with $\Pi^{\top}\Pi/\sqrt{K}=O(1)$, we have $a_2=0$, so that $\alpha_2=0$ and $\alpha_1=1$. By Lemma A.7, we have $\hat{\Sigma}/\Sigma\xrightarrow{p}1$, and note that $\Sigma/\Gamma_{\tilde{V},\tilde{e}}\to 1$ since $\Pi^{\top}\Pi/K\to 0$. With $\Gamma_{\tilde{V},\tilde{e}}/K\to\Gamma_{22}>0$, we can show that $\widehat{\Phi}_2\rightsquigarrow\bar{\Phi}_2$ for some random variable $\bar{\Phi}_2$ such that $\bar{\Phi}_2>0$ with probability one, as in the proof of Step 3 of Lemma A.8. Combining this with the fact that $\widehat{\Phi}_1\xrightarrow{p}0$ by (C.20) and Lemma A.7, we have $\hat{\alpha}_2\xrightarrow{p}0$ and $\hat{\alpha}_1\xrightarrow{p}1$, by the continuous mapping theorem. This concludes the proof. □

References

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): "Asymptotic Distribution Of JIVE In A Heteroskedastic IV Regression With Many Instruments," *Econometric Theory*, 28, 42–86.

Djogbenou, A. A., J. G. MacKinnon, and M. O. Nielsen (2019): "Asymptotic theory and wild bootstrap inference with clustered errors," *Journal of Econometrics*, 212, 393–412.

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

Hansen, B. E. and S. Lee (2019): "Asymptotic theory for clustered samples," *Journal of Econometrics*, 210, 268–290.

Lim, D., W. Wang, and Y. Zhang (2024): "A conditional linear combination test with many weak instruments," *Journal of Econometrics*, 238, 105602.

Magnus, J. R. and H. Neudecker (2019): *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons.

Müller, U. K. (2011): "Efficient tests under a weak convergence assumption," *Econometrica*, 79, 395–435.