A Valid Anderson-Rubin Test under Both Fixed and Diverging Number of Weak Instruments

Dennis Lim* Wenjie Wang† Yichong Zhang‡

Click here for the latest draft

Abstract

The conventional and jackknife Anderson-Rubin (AR) Tests are developed separately to conduct weak-identification-robust inference when the number of instrumental variables (IVs) is fixed or diverging to infinity with the sample size, respectively. These two tests compare distinct test statistics with distinct critical values. To implement them, researchers first need to take a stance on the asymptotic behaviour of the number of IVs, which is ambiguous when this number is just moderate. Instead, in this paper, we propose two analytical and one bootstrap-based weak-identification-robust AR tests, all of which control asymptotic size whether the number of IVs is fixed or diverging. We further analyze the power properties of these uniformly valid AR tests under both fixed and diverging number of IVs.

Keywords: Many instruments, size, weak identification

JEL Classification: C12, C36, C55

*Singapore Management University. E-mail address: dennis.lim.2019@phdecons.smu.edu.sg.

†Division of Economics, School of Social Sciences, Nanyang Technological University. HSS-04-65, 14 Nanyang Drive, Singapore 637332. E-mail address: wang.wj@ntu.edu.sg.

‡Singapore Management University. E-mail address: yczhang@smu.edu.sg.

1 Introduction

Existing literature on hypothesis testing for Instrumental Variable (IV) models focuses on either fixed number of instruments asymptotics (e.g. Andrews, Moreira, and Stock (2006), Kleibergen (2005)) or diverging instruments asymptotics (e.g. Angrist, Imbens, and Krueger (1999), Chao and Swanson (2005), Andrews and Stock (2007), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Mikusheva and Sun (2022)). To fully understand the problem at hand, we first restrict our attention to the Anderson-Rubin (AR) statistic. The reason for this restriction is as follows: Andrews et al. (2006)[Lemma 1(d)] showed that $Z'Y$ is a sufficient statistic for the parameter of interest β in the general Instrumental Variable IV framework (see (2.1)). They considered the Anderson-Rubin (AR) statistic[1], which is a bijective transformation of the sufficient statistic $Z'Y$. Since a statistic is a sufficient statistic if and only if their bijective transformation is itself a sufficient statistic[2], it follows that the AR-statistic is a sufficient statistic for the parameter of interest β. It is therefore reasonable to simply restrict our attention to this particular statistic and draw out its most salient features.

Going back to the problem, classical IV models assume that the number of instruments is fixed, and with it, the two-staged-least-square (2SLS) estimation was proposed. However, Sawa (1969) and Phillips and Hale (1977), among many others, have shown that the usual 2SLS estimation is biased whenever the number of instruments (K) diverge to infinity. To overcome this, Angrist et al. (1999) proposed running a first-stage regression n times, once for each observation, leaving out one observation at a time, where n is the number of sample size. This is commonly referred to as "Jackknifing" of a given statistic. In particular, Chao et al. (2012) derived the asymptotic property of the Jackknifed-AR test under the case of $K \to \infty$, showing that the estimator converges to a standard normal distribution under some appropriate re-scaling. However, when K is moderate, it is unclear which statistic the researcher should use. On one hand the researcher could use the classical AR-test for fixed instrument (defined as $AR_{classical}$ in section 5.1), which has size-control for fixed instruments but has power-deficit when the number of instruments is large (See Lemma B.5). On the other hand, the researcher could instead use the Jackknifed-AR test (defined as $AR_{standard}$ and AR_{cf} in section 5.1), which provides good size-control whenever the number of instruments is large, but has size-distortion when the number of instruments is small. A simple

[1] They denoted this statistic as S in equation (2.6) of their paper

[2] This follows straightforwardly from the Factorization Theorem, see for instance Lehmann and Romano (2006)[Corollary 2.6.1]

simulation illustrates this issue.[3]

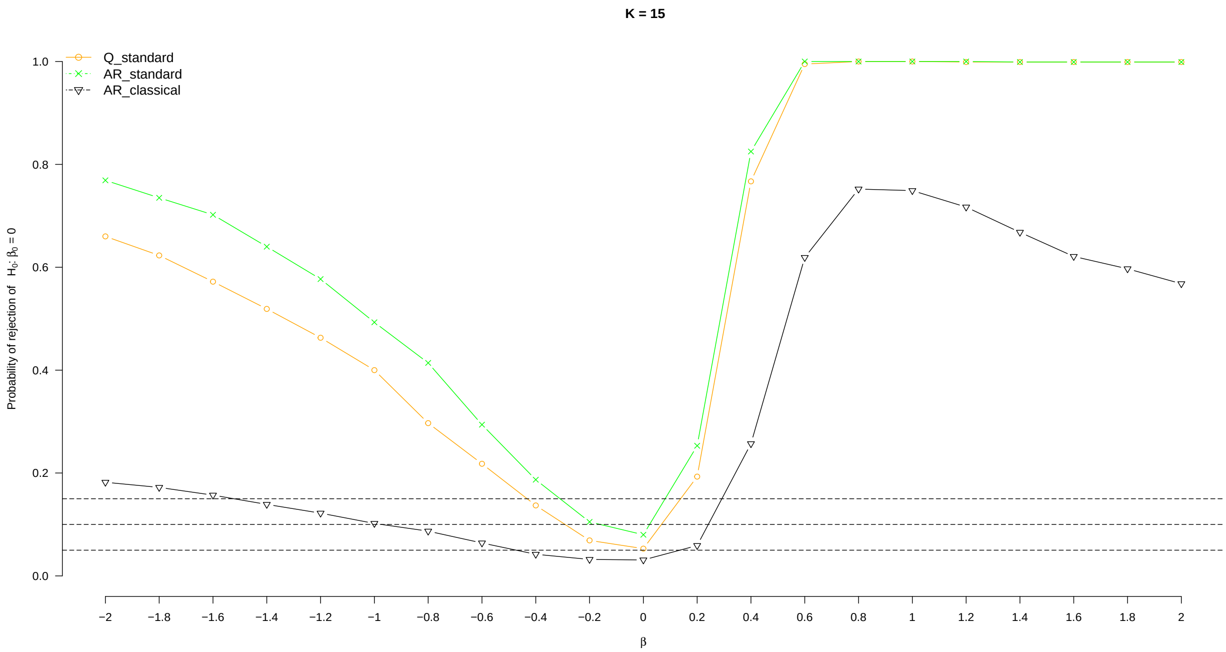

Figure 1: Power curve for $K = 15$

Note: The orange line with circle represents $Q_{standard}$; the green line with a cross represents $AR_{standard}$; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

Figure 1 demonstrates the case of moderate instruments, with the number of instruments being 15 and sample size equal 200. We propose two tests in the paper, one of which is $Q_{standard}$ (see section 5.1 for the description of this test). At the true parameter $\beta = 0$, $AR_{standard}$ has a size-distortion of 8%, while the sizes of $Q_{standard}$ and $AR_{classical}$ are 5.3% and 3.1% respectively. We can see that the power of $AR_{classical}$ is low throughout, while $Q_{standard}$ has the added advantage of mirroring $AR_{standard}$'s power while controlling for size.

Our proposed test takes into account this mismatch between fixed and diverging instrument asymptotics, and provide a critical-value that converges in both cases to the correct asymptotic limit distribution under the null, regardless of identification strength, so long as the number of

[3]The tests in Figure 1 are simulated based on the design of section 5.2, except we have reduced the sample size to 200 (from 400), and defined $U_2 \sim exp(0.2) - 5, v_1 = diag(z_{11}, z_{21}, ..., z_{n1})(Beta(0.5, 0.5) - 0.5)$. We also increased the concentration parameter $\overline{\mathcal{C}} \approx 140$, which is about eight times higher than that specified in section 5.2. Using a different (higher or lower) concentration parameter does not change the size, shape, power-ranking, and percentage difference in power among the tests. We increased $\overline{\mathcal{C}}$ to "stretch-out" the power-curves in order to see this difference clearer.

controls grow slower than the fourth root of the number of instruments[4]. The critical-value defined in (2.7) is related to Anatolyev and Solvsten (2023),[5] and we extend their result to the problem of weak instruments.

Structure of the paper: Section 2 makes precise the model setup and provides the testing procedure for our statistic under full-vector inference for both fixed and diverging instruments. It further motivates and introduces the robust critical-value for our test statistic. Section 3 provides a new strong approximation result for any 'AR-type' tests. Section 4 provides the asymptotic size and power properties of our test. Specifically, this section demonstrates that our test consistently differentiates the null from the alternative under strong identification, for both fixed and diverging instruments. Furthermore, that our test have exact asymptotic size-control for both fixed and diverging instruments is also shown. As an additional result, we derive in this section the exact distribution of a generic Jackknifed-AR statistic under fixed K setting. Section 5.2 provides simulation results for our power-curve based on calibrated data, which lends itself to our theory. Section 5.3 provides an application of our theory to empirical data. Proofs of Theorems, Lemmas, and Corollaries stated in the main text are shown in Appendix A, while Auxiliary Lemmas are provided in Appendix B. In Appendix C we provide details on the two estimators satisfying (2.9). In Appendix D we discuss general limit problems under fixed and diverging instruments.

Notation: We write $[n]$ to mean $\{1, ..., n\}$ and $\mathbb{N} := \{1, 2, ...\}$. In this paper, n is generally taken to be the sample size, unless otherwise stated. For any vector or matrix A, $||A||_F := \sqrt{trace(A'A)}$ is taken to be the Frobenius-norm. When there is no room for confusion, we simply write it as $||A||$. The spectral norm is denoted as $||A||_S := \sqrt{\lambda_{max}(A'A)}$, where $\lambda_{min}(B)$ and $\lambda_{max}(B)$ are defined as the minimum and maximum eigenvalue of a square matrix B. For any real numbers $a, b \in \mathbb{R}$, we write $a \leq Cb$ to mean that a is less than or equal b times a constant C that is independent of sample size n. For any index j, integer m and constant $\mathbb{C} > 0$, we write $\chi^2_{m,j}(\mathbb{C})$ to mean the jth chi-square random variable with m-degrees-of-freedom and non-centrality parameter $\mathbb{C}$. At times we do not include the index j, and write simply as $\chi^2_m(\mathbb{C})$ to mean a generic chi-square random variable with m-degrees-of-freedom and non-centrality parameter $\mathbb{C}$. We also write $\chi^2_{m,j}$ to mean $\chi^2_{m,j}(0)$, i.e. centrality parameter equal zero. We write WPA1 to mean 'with probability approaching one'. We define ι_i to be a vector of zeros, with the value 1 only on the ith element. For any set S, we write S^c to mean the complement of the set. We use the symbol '$\otimes$' to denote Kronecker product. We

[4] Chao, Swanson, and Woutersen (2023) showed that when the dimension of controls are large, partialling these controls out leads to inconsistent estimates under weak identification. They assumed $\frac{\sqrt{d_W}}{n} = o(1)$, where d_W is the dimension of the controls, and showed that this condition is sufficient for consistent hypothesis testing. We have a similar type of assumption here (see assumption 2)

[5] In particular, they showed that a weighted chi-bar distribution is able to mirror statistics of the AR-type (We say that a statistic T is of an AR-type if we can express $T = \varepsilon A \varepsilon$ for some deterministic symmetric matrix A and ε is a random vector with zero mean and well-defined (or finite) covariance matrix)

write $\mathcal{Z}_K(J)$ to represent a standard Gaussian plus $J \in \mathbb{R}^K$, i.e. $\mathcal{Z}_K(J) := \mathcal{N}(J, I_K)$. For any statistic T with some given probability law, denote $q_{1-\alpha}(T)$ to be the $(1-\alpha)$-quantile of the law of T.

2 Setup and Testing Procedure

2.1 Setup

Consider the model

$$\begin{aligned} \widetilde{Y} &= \widetilde{X}\beta + W\mathbb{\Gamma} + \widetilde{e} \\ \widetilde{X} &= \widetilde{\Pi} + \widetilde{v} \end{aligned} \tag{2.1}$$

where $\widetilde{X} \in \mathbb{R}^{n\times d_X}$, $W \in \mathbb{R}^{n\times d_W}$, d_X is of some fixed finite dimension, $\widetilde{Y}, \widetilde{e} \in \mathbb{R}^{n\times 1}$, $\widetilde{\Pi}_i \equiv \mathbb{E}(\widetilde{X}_i|\widetilde{Z}_i, W_i) \in \mathbb{R}^{1\times d_X}$ where $\widetilde{Z} \in \mathbb{R}^{n\times K}$ is the matrix of instrument with full-rank. Also, $\beta \in \mathbb{R}^{d_X}$ and $\mathbb{\Gamma} \in \mathbb{R}^{d_W\times 1}$. We observe $(\widetilde{Y}, \widetilde{X}, W, \widetilde{Z})$, and assume that W is a full-ranked **exogenous** control matrix with $d_W \leq n$, implying that its projection matrix $P_W := W(W'W)^{-1}W'$ is well-defined. Furthermore, the error terms $\widetilde{e}_i$ are assumed to be independent across i. We assume throughout this paper that $d_X = 1$ in order to highlight the most salient features of our test, but we remark here that it can be extended to higher dimensions (i.e. d_X to be of dimension greater than one) so that β can be multivariate.[6]

We are interested in testing

$$H_0 : \beta = \beta_0 \quad versus \quad H_1 : \beta \neq \beta_0. \tag{2.2}$$

To this end, we want to obtain a test that has size control under the null, irregardless of identification strength. We allow the dimensions of the instruments and control, d_Z and d_W, to diverge to infinity as $n \to \infty$ (these dimensions can be fixed as well), with the added allowance that whenever they do diverge, d_Z can grow at the same rate as the sample size, while d_W must grow at a slower rate than the sample size.

To simplify matters, we first partial out the controls W and rewrite the model as

$$\begin{aligned} Y &= X\beta + e \\ X &= \Pi + v \end{aligned} \tag{2.3}$$

where $Y = M_W\widetilde{Y}$, $X = M_W\widetilde{X}$, $\Pi = M_W\widetilde{\Pi}$, $e = M_W\widetilde{e}$, $v = M_W\widetilde{v}$, $Z = M_W\widetilde{Z}$, $M_W = I_n -$

[6] See Remark 1

P^W, where $P^W := W(W'W)^{-1}W'$. Throughout the text, we denote $\widetilde{\sigma}_i^2 := \mathbb{E}\widetilde{e}_i^2, \widetilde{\varsigma}_i^2 := \mathbb{E}\widetilde{v}_i^2, \sigma_i^2 := \mathbb{E}e_i^2, \varsigma_i^2 := \mathbb{E}v_i^2, \widetilde{\gamma}_i := Cov(\widetilde{e}_i, \widetilde{v}_i)$ and $P := Z(Z'Z)^{-1}Z'$. We define $e_i(\beta_0) := Y - X\beta_0 = e + \Delta X$, where $\Delta := \beta - \beta_0$. We define $\sigma_i^2(\beta_0) := \widetilde{\sigma}_i^2 + 2\Delta\widetilde{\gamma}_i + \Delta^2\widetilde{\varsigma}_i^2$ and $\varsigma_i^2(\beta_0) := \widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i + \Delta^2\widetilde{\sigma}_i^2$. For notational simplicity, we write $e := (e_1, ..., e_n)'$ instead $e(\beta_0)$ whenever $\beta = \beta_0$. Furthermore, define $U := Z(Z'Z)^{-1/2} \in \mathbb{R}^{n\times K}$ and $Q_{a,b} := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}a_i b_j}{\sqrt{K}}$ for any two vectors $a, b \in \mathbb{R}^n$, where P_{ij} is the (i,j)-th element of P. We make the following assumptions thought-out the rest of the paper.

Assumption 1. *Suppose that the errors* $(\widetilde{e}_i, \widetilde{v}_i)$ *are mean zero and independent over* i.

Assumption 2 (Moment conditions)**.** *Suppose* $\frac{p_n}{K} = o(1)$ *and* $p_n \leq \delta < 1$*, where* $p_n := \max_i P_{ii}$*. Furthermore, assume* $p_n^W := \max_i P_{ii}^W = o(1)$*, and* $d_W = O(K^{(1-\eta)/4})$ *for any* $\eta > 0$*. Let the errors and* $|\Pi_i|$ *be bounded in the eighth moment and bounded away from zero in the second moment, i.e.* $\max_i(\Pi_i^8 + \mathbb{E}\widetilde{e}_i^8 + \mathbb{E}\widetilde{v}_i^8) < \overline{C} < \infty$ *and* $(\Pi'\Pi)^2, \sigma_i^2(\beta_0), \varsigma_i^2(\beta_0) \geq \underline{C} > 0$*. Furthermore, suppose* $\underline{C} \leq \lambda_{\min}(W'W/n) \leq \lambda_{\max}(W'W/n) \leq \overline{C}$ *and that* Z *has full rank.*

For a balanced instrument design without controls, $p_n = \frac{K}{n}$. Hence, for both fixed and diverging K, $\frac{p_n}{K} = \frac{1}{n} = o(1)$. Note that $p_n > 0$ by the full rank assumption of Z, since $\sum_{i\in[n]} P_{ii} = K$. Furthermore, $p_n \leq 1$ since each element on the diagonal of a projection matrix is always bounded by one. We allow the number of controls to diverge to infinity. However, in order for p_n^W to shrink to zero in assumption 2, the increase in dimension of the control d_W must be slower than n (i.e. $d_W = o(n)$), since by definition, $p_n^W \geq \frac{d_W}{n}$. In fact, we require a weaker assumption, that is, $d_W = O(K^{(1-\eta)/4})$ for any arbitrarily small $\eta > 0$. This assumption ensures that we can strongly approximate our statistic.[7] In the case of fixed K,

$$\frac{p_n d_W^2}{K^{1/2}} = \frac{p_n^{1/2}}{K^{1/2}}(p_n^{1/2} \cdot O(1) \cdot K^{-(1-\eta)/2}) = \frac{p_n^{1/2}}{K^{1/2}}O(1) = o(1)O(1) = o(1)$$

Under diverging K,

$$\frac{p_n d_W^2}{K^{1/2}} \leq \frac{d_W^2}{K^{1/2}} = O(1) \cdot K^{-(1-\eta)/2}K^{1/2} = o(1)$$

2.2 Some Background and Motivation

In this section we briefly discuss the general difficulties of constructing a test that has simultaneous size-control for both fixed and diverging instruments. Consider first the classical case of homoskedastic variance and fixed instruments. For simplicity, we assume for the moment that control matrices are not present in the model of (2.1). Under the null, a consistent estimator of the variance σ^2 can be given by $\widehat{\sigma}^2 := \frac{1}{n}\sum_{i\in[n]} e_i^2$. Then under the usual regularity assumptions,

[7] See Theorem 1 and the discussion after.

by continuous mapping theorem the estimator

$$\frac{e'Pe}{K\widehat{\sigma}^2} = \frac{1}{K\sigma^2 + o_p(1)}(n^{-1/2}Z'e)'(n^{-1}Z'Z)^{-1}(n^{-1/2}Z'e) \rightsquigarrow \frac{1}{K}\chi^2_K.$$

Consider now the case of diverging instruments. Note that by Chao et al. (2012)[Lemma A2], $\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j}{\sqrt{2K\widehat{\sigma}^2}} \rightsquigarrow \mathcal{N}(0,1)$. Furthermore, WPA1 we have $\frac{\sum_{i\in[n]} P_{ii}e_i^2}{K\widehat{\sigma}^2} = \frac{\sum_{i\in[n]} P_{ii}\sigma^2}{K\sigma^2} = \frac{\sum_{i\in[n]} P_{ii}}{K} = 1$ (See Lemma B.1). Therefore we have

$$\frac{e'Pe}{K\widehat{\sigma}^2} = \frac{1}{\sqrt{K}}\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j}{\sqrt{K}\widehat{\sigma}^2} + \frac{\sum_{i\in[n]} P_{ii}e_i^2}{K\widehat{\sigma}^2} \xrightarrow{p} 1.$$

Observe then that there are two distinct limiting distributions for the same (classical) statistic under two different cases of instruments. In fact, for the diverging K case, $e'Pe$ itself would diverge to infinity, so that the denominator K acts as a form of normalization. This normalization has the same order as the diagonal elements. To see this, note that the diagonal elements $\sum_{i\in[n]} P_{ii}e_i^2 = O(K)$, while the non-diagonal elements $\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j = O(\sqrt{K})$, so that the order of the diagonal terms dominate the non-diagonals. Note that the non-diagonals have a smaller order due to it being centered. At this stage, we conclude that the statistic $\frac{e'Pe}{K\widehat{\sigma}^2}$ does not work simultaneously for both cases of instruments, due to the diagonal elements. This highlights the importance of removing the diagonals under diverging K. Therefore, in order to consider both cases of fixed and diverging instruments simultaneously, a natural idea would be to focus on the Jackknifed statistic, where the diagonals are removed, i.e. the statistic

$$\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j}{\sqrt{2K}\widehat{\sigma}^2},$$

which converges weakly to a $\frac{\chi^2_K - K}{\sqrt{2K}}$-distribution under fixed K. As $K \to \infty$, we see that $\frac{\chi^2_K - K}{\sqrt{2K}} \rightsquigarrow \mathcal{N}(0,1)$. A researcher would therefore be inclined to use the following test under homoskedasticity: Reject whenever

$$\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j}{\sqrt{2K}\widehat{\sigma}^2} > q_{1-\alpha}(\frac{\chi^2_K - K}{\sqrt{2K}})$$

As a matter of fact, they would have exact asymptotic-size control in either case of fixed or diverging instruments. However, under general heteroskedasticity, we see that this matter is further complicated because the variances of errors are generally of unknown form, so that consistent estimation of these variances is impossible whenever instruments diverge. Nevertheless, as we explain in the next section, even under diverging controls and heteroskedastic errors, our method provides exact asymptotic size-control simultaneously for both fixed and diverging instruments.

2.3 Analytical Test Statistic and Critical Value

Our test statistic is denoted as $\widehat{Q}(\beta_0)$ and defined as

$$\widehat{Q}(\beta_0) := \frac{e(\beta_0)' P e(\beta_0)}{\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)} \tag{2.4}$$

Our test compares the test statistic $\widehat{Q}(\beta_0)$ with a robust critical value $C_\alpha(\widehat{\Phi}_1(\beta_0))$, where $\alpha \in (0,1)$ is the significance level and $\widehat{\Phi}_1(\beta_0)$ is a consistent estimator of $\Phi_1(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$ under the null, with more details provided in section 2.4. We will reject $H_0 : \beta = \beta_0$ at α significance level if

$$\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0)).$$

To see the exact formula of the critical value, we need to explain the limit distribution of our test statistic $\widehat{Q}(\beta_0)$ under the null, in which case the $e_i(\beta_0)$ has mean zero and variance $\sigma_i^2(\beta_0)$ for $\beta = \beta_0$. When K is fixed, under regularity conditions, we can show that

$$\widehat{Q}(\beta_0) \rightsquigarrow \mathcal{Z}' D_n \mathcal{Z} = \sum_{k\in[K]} w_{n,i}\chi_{1,k}^2, \tag{2.5}$$

where $\mathcal{Z} \sim \mathcal{N}(0, I_K)$ and $D_n := diag(w_{1,n}, ..., w_{K,n})$ are the eigenvalues of

$$\Omega(\beta_0) := \frac{(Z'\Lambda(\beta_0)Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda(\beta_0)Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}, \tag{2.6}$$

where $\Lambda(\beta_0) = \text{diag}(\sigma_1^2(\beta_0), \cdots, \sigma_n^2(\beta_0))$, and $\{\chi_{1,i}^2\}_{i\in[K]}$ are K independent chi-squared random variables with 1 degree of freedom. The denominator of $\Omega(\beta_0)$ (i.e., $\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)$) is chosen so that $trace(\Omega(\beta_0)) = 1$. Also note that $\Omega(\beta_0)$ is positive semi-definite, implying that its eigenvalues $(\omega_1, \cdots, \omega_K)$ are nonnegative and sum up to 1.

Suppose $\hat{\Lambda}(\beta_0) = \text{diag}(e_1^2(\beta_0), \cdots, e_n^2(\beta_0))$. Then, when K is fixed, we can consistently estimate the eigenvalues $(w_{1,n}, ..., w_{K,n})$ by the eigenvalues of

$$\widehat{\Omega}(\beta_0) := \frac{(Z'\hat{\Lambda}(\beta_0)Z)^{1/2}(Z'Z)^{-1}(Z'\hat{\Lambda}(\beta_0)Z)^{1/2}}{\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)},$$

which are denoted as $\widetilde{w}_n = (\widetilde{w}_{1,n}, \cdots, \widetilde{w}_{K,n})'$. This motivates us to consider the $1-\alpha$ quantile of weighted chi-squares random variable with weights $\widetilde{w}_n$ (i.e., $F_{\widetilde{w}_n} = \sum_{i\in[K]} \widetilde{w}_{i,n}\chi_{1,i}^2$), which is denoted as $q_{1-\alpha}(F_{\widetilde{w}_n})$ and can be simulated given $\tilde{w}$. However, the eigenvalue estimators are not consistent if K is diverging as fast as the sample size n. Fortunately, in this case, we can show that

that

$$\Phi^{-1/2}(\beta_0)\left[\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)\right](\widehat{Q}(\beta_0)-1)\rightsquigarrow\mathcal{N}(0,1)$$

and

$$\left(\sum_{k\in[K]}2\tilde{w}_{n,k}^2\right)^{-1/2}(F_{\tilde{w}}-1)\rightsquigarrow\mathcal{N}(0,1).$$

where $\Phi_1(\beta_0)=\frac{2}{K}\sum_{i\in[n]}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$.

Given a consistent estimator $\widehat{\Phi}_1(\beta_0)$ of $\Phi_1(\beta_0)$, we can adjust the critical value $q_{1-\alpha}(F_{\widetilde{w}_n})$ as

$$C_\alpha(\widehat{\Phi}_1(\beta_0)):=1+\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)}\left(\frac{q_{1-\alpha}(F_{\widetilde{w}_n})-1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}}\right). \tag{2.7}$$

This adjustment guarantees the asymptotic size control of our test under diverging K case.

Lastly, we note that the critical value $C_\alpha(\widehat{\Phi}_1(\beta_0))$ can be rearranged as

$$q_{1-\alpha}(F_{\widetilde{w}_n})+(q_{1-\alpha}(F_{\widetilde{w}_n})-1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}}-1\right). \tag{2.8}$$

When K is fixed, we are able to show that, under the null,

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}}-1\xrightarrow{p}0,$$

implying that the adjustment of the critical value is asymptotically negligible. This guarantees the asymptotic size control of our test under the fixed K case.

2.4 Estimator for Critical Values

We provide further details of $\widehat{\Phi}_1(\beta_0)$ discussed in the previous section. We assume that $\widehat{\Phi}_1(\beta_0)$ is some estimator satisfying

$$\widehat{\Phi}_1(\beta_0)=\Phi(\beta_0)+\mathcal{D}(\Delta)+o_p(1+\sum_{i\in[4]}\Delta^i) \tag{2.9}$$

where

$$\Phi(\beta_0) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$$

and

$$\mathcal{D}(\Delta) = \begin{cases} O_p(1) & \text{if } \Delta \neq 0 \text{ is fixed} \\ o_p(1) & \text{if } \Delta = o(1) \end{cases}$$

We introduce two estimators that satisfy (2.9) – this is shown in Appendix C. The first estimator is due to Crudu, Mellace, and Sándor (2021), which we denote as

$$\widehat{\Phi}_1^{standard}(\beta_0) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2(\beta_0)e_j^2(\beta_0)$$

In this case, its accompanying function for $\mathcal{D}(\Delta)$ is given as[8]

$$\mathcal{D}^{standard}(\Delta) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(2\Delta^2\Pi_j^2\sigma_i^2(\beta_0) + \Delta^4\Pi_i^2\Pi_j^2).$$

In order to decrease the size of the variance estimator under the alternative, we further consider the cross-fit variance estimator due to Mikusheva and Sun (2022).

$$\widehat{\Phi}_1^{cf}(\beta_0) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} \widetilde{P}_{ij}^2[e_i(\beta_0)M_i'e(\beta_0)][e_j(\beta_0)M_j'e(\beta_0)]$$

where $M := I_n - Z(Z'Z)^{-1}Z'$ and $\widetilde{P}_{ij}^2 := \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$, which is the second estimator satisfying (2.9). Its corresponding asymptotic property as well as the expression of $\mathcal{D}^{cf}(\Delta)$ is provided in Theorem C.0.2.[9] To see why the cross-fit estimator works, under the alternative, we can express $e_i(\beta_0) = e_i + \Delta\Pi_i + \Delta v_i$. Consider the case where $\widetilde{\Pi} \equiv \widetilde{Z}\theta_0$. Then $\Pi = M_W\widetilde{\Pi} = M_W\widetilde{Z}\theta_0$, so that $M\Pi = MM_W\widetilde{Z}\theta_0 = MZ\theta_0 = 0$ as $Z = M_W\widetilde{Z}$. Hence we can remove the effects of Δ from Π_i. The bias of the standard variance estimator $\widehat{\Phi}_1^{standard}(\beta_0)$ grows the at fourth power of Δ, so that removing this component leads to higher power. Note that whenever the controls W are dropped out of the model (2.1), the cross-fit estimator is exactly Mikusheva and Sun (2022)'s cross-fit estimator and $\mathbb{E}\widehat{\Phi}_1^{cf}(\beta_0) = \Phi_1(\beta_0)$ under the null. However, when there are exogenous controls in the model, $\mathbb{E}\widehat{\Phi}_1^{cf}(\beta_0) \neq \Phi_1(\beta_0)$ due to the effects of partial-ling out the controls M_W from the error terms $\widetilde{e}$, which leads to dependence among the error terms e_i in the reduced-form model (2.3). The reason we are still able to obtain a consistent cross-fit estimator under the null lies in

[8]This is shown in Theorem C.0.1

[9]Note that the cross-fit estimator is more 'costly' than the standard estimator in the sense that the former requires that $\max_i P_{ii} \leq \delta < 1$, while the latter does not have this requirement.

the assumption that $p_n^W := \max_i P_{ii}^W = o(1)$.

2.5 Bootstrap-based Test

The Bootstrap-based statistic is defined as

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_i(\beta_0)e_j(\beta_0)}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \tag{2.10}$$

with $\widehat{\Phi}_1(\beta_0)$ satisfying (2.9) and has the additional requirement that it can be constructed from using only $e(\beta_0)$ and P. The two estimators $\widehat{\Phi}_1(\beta_0)^{standard}$ and $\widehat{\Phi}_1(\beta_0)^{cf}$ discussed in section 2.4 satisfy this requirement. We will reject $H_0 := \beta = \beta_0$ at α significance level if

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_\alpha^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L}),$$

where $C_\alpha^B(\widehat{\Phi}_1(\beta_0), \mathcal{L})$ is the critical value that depends (1) on some large positive integer B, (2) significance-level α, (3) identically and independently distributed (i.i.d.) random variables $\{\kappa_i\}_{i\in[n]}$ following the probability law $\mathcal{L}$ with the property that its mean is zero and variance is one, and (4) the bootstrapped estimator $\widehat{\Phi}_1^{BS}(\beta_0)$. The critical-value is computed in the following manner: Fix β_0, a large B, and some $\alpha \in (0,1)$. Fix any $\ell \in \{1, ..., B\}$, and generate independent random variables $\{\kappa_{i,\ell}\}_{i\in[n]}$ following the law $\mathcal{L}$. We then multiply each $e_i(\beta_0)$ by $\kappa_{i,\ell}$, denoting the new random variable $\eta_{i,\ell} := \kappa_{i,\ell}e_i(\beta_0)$. Since $\widehat{\Phi}_1(\beta_0)$ is assumed to be constructed by using only $e(\beta_0)$ and P, we construct $\widehat{\Phi}_1^{BS,\ell}(\beta_0)$ in exactly the same way that $\widehat{\Phi}_1(\beta_0)$ was constructed, but replacing $(e(\beta_0), P)$ with (η_ℓ, P), where $\eta_\ell = (\eta_{1,\ell}, ..., \eta_{n,\ell})'$. Once this is done, construct the statistic

$$\widehat{J}^{BS,\ell} := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_{i,\ell}\eta_{j,\ell}}{\sqrt{K\widehat{\Phi}_1^{BS,\ell}(\beta_0)}}$$

By repeating this process for every $\ell \in [n]$, we obtain a collection of statistics $\{\widehat{J}^{BS,\ell}\}_{\ell\in[B]}$. Then

$$C_\alpha^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L}) := inf\left\{ z \in \mathbb{R} : 1-\alpha \leq \frac{\sum_{\ell\in[B]} 1\left\{\widehat{J}^{BS,\ell} \leq z\right\}}{B} \right\} \tag{2.11}$$

3 Strong Approximation

This section is concerned with the conditions for which we can view the error terms $(\widetilde{e}_i, \widetilde{v}_i)$ as being normally distributed. This is important for understanding the limit distribution of (2.4) under fixed instruments, as well as generic Jackknifed-AR tests under fixed instruments.

Consider a sequence of independent random variables $\{\varepsilon_i\}_{i\in[n]}$ such that $\varepsilon_i \sim \mathcal{N}(0, \widetilde{\sigma}_i^2)$, so that ε_i mirrors the first and second moment of $\widetilde{e}_i$. We assume that $\{\varepsilon_i\}_{i\in[n]}$ is independent of $\{(\widetilde{e}_i, \widetilde{v}_n)\}_{i\in[n]}$. We have the following result which tells us that under the null, whether our statistic is Jackknifed or of the AR-type, we can always treat our errors as being normally distributed.

Theorem 1 (Strong approximation)**.** *Suppose assumption 1 holds and* $\sup_{i\in\mathbb{N}} \mathbb{E}(\widetilde{e}_i)^4 < \infty$. *Then we have*

$$\frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ij} e_i e_j \stackrel{d}{=} \frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ij} \mathcal{E}_i \mathcal{E}_j + O_p\left(\left[\frac{(p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2} d_W)}{K^{1/2}} \right]^{1/3} + \frac{p_n d_W^2}{K^{1/2}} \right)$$

where $p_n := \max_i P_{ii}$ *and* $\mathcal{E} := M_W \varepsilon$. *Furthermore,*

$$\frac{1}{K} e' P e \stackrel{d}{=} \frac{1}{K} \mathcal{E}' P \mathcal{E} + O_p\left(\frac{p_n^{1/2}}{K^{1/2}} \right)$$

The requirement for strong approximation is very weak, namely that $\frac{p_n}{K} = o(1)$ and $\frac{p_n d_W^2}{K^{1/2}} = o(1)$. In the simple case where d_W is bounded, i.e. $d_W \leq C$ for some $C < \infty$, we only require that $\frac{p_n}{K} = o(1)$, since then

$$\frac{d_W p_n^{1/2}}{K^{1/4}} \leq C p_n^{1/4} \frac{p_n^{1/4}}{K^{1/4}} \leq C \frac{p_n^{1/4}}{K^{1/4}} = o(1)$$

In view of Theorem 1, we can view errors to be normally distributed under assumption 2. The requirement for the eighth-moment of errors to be bounded is used only to control the size of our test statistic under the diverging K case, specifically when K diverges at the same order as n (see Theorem 2 and Lemma B.3, diverging K case).

4 Asymptotic properties

4.1 Asymptotic size

We discuss the size properties of our test in this section, and provide details on why the Jackknifed AR test will fail under fixed K asymptotics, not just under homoskedasticity (which was discussed in section 2.2), but in the presence of heteroskedasticity. In fact, we establish necessary and sufficient conditions for which general Jackknifed-AR can obtain exact asymptotic size under general conditions. We have the following assumption.

Assumption 3. *Suppose $p_n \leq \overline{C}\frac{K}{n}$ for some $\overline{C} < \infty$*

Assumption 3 ensures that we have size control. Intuitively, it states that the largest value on the diagonal of the projection matrix P is regular in the sense that the order of p_n is equal to the fraction of instruments over the number of observations, $\frac{K}{n}$. This follows from the fact that, by definition, $\frac{K}{n} \leq p_n$. In the case of balanced instruments, we have that $p_n = \frac{K}{n}$. Furthermore, note that this assumption automatically implies the first part of Assumption 2, since then $\frac{p_n}{K} \leq \overline{C}\frac{K}{n}\frac{1}{K} = \frac{\overline{C}}{n} = o(1)$.

By the results of the previous sections, we can show uniform size-control of our test under any identification strength, simultaneously for both fixed and diverging instruments. Let $\lambda_n \in \Lambda_n$ be the data generating process of n observations for $(\widetilde{e}, \widetilde{v}, Z, W)$. We impose the following restriction on the sequence of classes of DGPs $(\{\Lambda_n\}_{n\geq 1})$:

$$\left(\begin{array}{c} \{\widetilde{e}_i, \widetilde{v}_i\}_{i\in[n]} \text{ are independent, } \mathbb{E}\widetilde{e}_i = \mathbb{E}\widetilde{v}_i = 0, \\ \frac{p_n}{K} = o(1), p_n^W = o(1), d_W = O(K^{(1-\eta)/4}) \text{ for any } \eta > 0, \\ \max_i \Pi_i^2 + \max_i \mathbb{E}\widetilde{e}_i^8 + \max_i \mathbb{E}\widetilde{v}_i^8 \leq \overline{C} < \infty, \\ \Pi'\Pi, \sigma_i^2(\beta_0), \zeta_i^2(\beta_0) \geq \underline{C} > 0 \text{ under the null,} \\ \underline{C} \leq \text{mineig}(\frac{W'W}{n}) \leq \text{maxeig}(\frac{W'W}{n}) \leq \overline{C}, \\ 0 \leq P_{ii} \leq \delta < 1, \\ \widehat{\Phi}_1(\beta_0) \text{ satisfies (2.9) under the null,} \\ \text{where } 0 < \underline{C}, \overline{C}, \delta < \infty \text{ are some fixed constants} \end{array}\right) \tag{4.1}$$

Then our test has size-control uniformly over the set of DGPs that satisfy (4.1). We formalize the statement as follows:

Theorem 2. *Suppose $\{\Lambda_n\}_{n\geq 1}$ satisfies (4.1) and assumption 3 holds. Then under the null, for both fixed and diverging instruments, we have exact size control for the proposed tests, i.e.*

$$\liminf_{n\to\infty} \inf_{\lambda_n\in\Lambda_n} \mathbb{P}_{\lambda_n}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = \limsup_{n\to\infty} \sup_{\lambda_n\in\Lambda_n} \mathbb{P}_{\lambda_n}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = \alpha$$

and

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{\lambda_n\in\Lambda_n} \mathbb{P}_{\lambda_n}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_\alpha^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) \\ &\quad = \limsup_{n\to\infty} \sup_{\lambda_n\in\Lambda_n} \mathbb{P}_{\lambda_n}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_\alpha^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) = \alpha \end{aligned}$$

Remark 1. *Note that Theorem 2 still holds when β is multivariate (instead of a scalar in (2.1)). This is because under the null, the true error $\widetilde{e}$ can be taken as known, with the remaining computation of our test depending only on the controls W and instrument Z, both of which are observed.*

Therefore, repeating the proof under the null yields size control for any $\beta \in \mathbb{R}^{d_X}$ with fixed $d_X \geq 1$.

4.2 Asymptotic power

In this section we show that under strong identification, for both fixed and diverging instruments, our test consistently differentiates the null from the alternative, where strong identification means $\overline{\mathcal{C}} := Q_{\Pi,\Pi} \to \infty$. The concentration parameter $\overline{\mathcal{C}}$ was introduced by Mikusheva and Sun (2022).[10] To motivate this concentration parameter, note that under the linear IV setting where $\Pi_i = \pi' Z_i$, for $K \to \infty$ it was shown in Mikusheva and Sun (2022)[Theorem 1] that whenever $\frac{\pi' Z' Z \pi}{\sqrt{K}}$ is bounded, no test can consistently differentiate the null from the alternative. Furthermore, Chao et al. (2012)'s consistent estimator was based on the assumption that $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$.[11] Taken together, one can expect that the requirement of $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$ in the linear IV setting is important to ensuring that our test consistently differentiates the null from the alternative. In fact, this requirement is equal to requiring that $\overline{\mathcal{C}} \to \infty$, which explains why $\overline{\mathcal{C}}$ should be the right measure of identification strength. [12]

4.2.1 Diverging instruments

We want to evaluate the power of our test $\widehat{Q}(\beta_0)$ under permutations of different scenarios. In particular, we consider three cases for some sequence $d_n \to 0$: (1) Strong identification and local alternative, where $d_n \overline{\mathcal{C}} = \widetilde{\mathcal{C}}$ and $\Delta = \widetilde{\Delta} d_n^{1/2}$ for some fixed $\widetilde{\Delta}, \widetilde{\mathcal{C}} \in \mathbb{R}$; (2) Strong identification and fixed alternative, where $d_n \overline{\mathcal{C}} = \widetilde{\mathcal{C}}$ and $\Delta = \widetilde{\Delta}$; (3) Weak identification and fixed alternative, where $\overline{\mathcal{C}} = \widetilde{\mathcal{C}}$ and $\Delta = \widetilde{\Delta}$.

Theorem 3. *Suppose Assumption 1, 2, 3 and (D.1) holds and $K \to \infty$. For any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (2.9), we have under strong identification and fixed alternative*

$$\lim_{n \to \infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = 1$$

Theorem 3 shows that whenever identification strength diverges to infinity, our test consistently differentiates the null from the alternative. Note that in general, for any fixed alternative Δ not

[10] Section D provides more detail regarding the concentration parameter $\overline{\mathcal{C}}$

[11] See Assumption 2 of their paper

[12] To see this, note that we can express the concentration parameter as $\overline{\mathcal{C}} = \frac{\pi' Z' Z \pi}{\sqrt{K}} - \frac{\sum_{i \in [n]} P_{ii} (\pi' Z_i)^2}{\sqrt{K}}$, so that by assumption 2, $(1-\delta) \frac{\pi' Z' Z \pi}{\sqrt{K}} \leq \overline{\mathcal{C}} \leq \frac{\pi' Z' Z \pi}{\sqrt{K}}$. We can then see that the order between $\frac{\pi' Z' Z \pi}{\sqrt{K}}$ and $\overline{\mathcal{C}}$ are the same.

necessarily zero, for diverging K we have that[13]

$$\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}} \rightsquigarrow \mathcal{N}(0,1)$$

Therefore, under weak identification with fixed alternatives, we have the following result:

Theorem 4. *Suppose Assumption 1, 2, 3 and (D.1) holds. For $K \to \infty$ and any estimator $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$, we have under weak identification and fixed alternative that*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}\right)$$

where $F(\cdot)$ denotes the cumulative distribution function (CDF) of a standard normal distribution. In particular, if we assume $\Pi' M \Pi \leq \frac{\Pi'\Pi}{K} \to 0$, then $\widehat{\Phi}_1(\beta_0)$ can be taken as $\widehat{\Phi}_1^\ell(\beta_0)$ for $\ell = \{standard, cf\}$ given in section 2.4.

The assumption of $\frac{\Pi'\Pi}{K} \to 0$ automatically ensures that $\widehat{\Phi}_1^{standard}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$, while the additional requirement of $\Pi' M \Pi \leq \frac{\Pi'\Pi}{K}$ is made to ensure that $\widehat{\Phi}_1^{cf}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$ as well. Next, we have the asymptotic power for our test under strong-identification and local-alternative, which is similar to the case of weak identification and fixed alternative.

Theorem 5. *Suppose Assumption 1, 2, 3 and (D.1) holds. For $K \to \infty$ and any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (2.9), under strong identification and local alternative we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}\right)$$

4.2.2 Fixed instruments

We introduce a measure of identification strength for a fixed number of instruments, defined as

$$\widetilde{\mu}_n^2 := ||\mu_{K,n}||_F^2$$

where $\mu_{K,n} := n^{-1/2} Z'\Pi$. For notational simplicity we drop the dependence on n and simply denote $\mu_{K,n}$ by μ_K. Note that there is an intimate relationship between the concentration parameter defined above for the fixed K case (i.e. $\widetilde{\mu}_n^2$) and the concentration parameter $\overline{\mathcal{C}}$ defined for the diverging K case discussed earlier: $\widetilde{\mu}_n^2$ and $\overline{\mathcal{C}}$ have the same order. To see this, note that under the

[13] See the proof of Theorem 3

assumption that $Z'Z/n \xrightarrow{p} Q_{ZZ}$, a positive-definite matrix, we have that with WPA1,

$$\widetilde{\mu}_n^2 \leq \lambda_{max}\left(\frac{Z'Z}{n}\right) \cdot \mu_K' \left(\frac{Z'Z}{n}\right)^{-1} \mu_K = \lambda_{max}(Q_{ZZ})\Pi' P \Pi \leq \frac{\lambda_{max}(Q_{ZZ})}{\lambda_{min}(Q_{ZZ})}\widetilde{\mu}_n^2$$

where we note that $\widetilde{\mu}_n^2 = \mu_K'\mu_K$. Since $0 < \lambda_{min}(Q_{ZZ}) \leq \lambda_{max}(Q_{ZZ}) \leq C$, $\widetilde{\mu}_n^2$ has the same order as $\Pi' P \Pi$; as K is fixed, $\widetilde{\mu}_n^2$ has the same order as $\frac{\Pi' P \Pi}{\sqrt{K}}$. Furthermore, observe $\frac{\sum_{i\in[n]} P_{ii}\Pi_i^2}{\sqrt{K}} \leq \max_i \Pi_i^2 \frac{\sum_{i\in[n]} P_{ii}}{\sqrt{K}} \leq C\sqrt{K} \leq C$ under fixed instruments, so that $\frac{\Pi' P \Pi}{\sqrt{K}} = \overline{\mathcal{C}} + \frac{\sum_{i\in[n]} P_{ii}\Pi_i^2}{\sqrt{K}}$ has the same order as $\overline{\mathcal{C}}$. Combining these facts yield the result that $\widetilde{\mu}_n^2$ has the same order as $\overline{\mathcal{C}}$.

We say that there is strong identification whenever $\widetilde{\mu}_n^2 \to \infty$. Otherwise we say that there is weak identification. To be precise we consider three cases for some sequence $d_n \to 0$: (1) Strong identification and local alternative, where $\Delta = \widetilde{\Delta} d_n$ for some fixed $\widetilde{\Delta}$ and $\widetilde{\mu}_n^2 = \widetilde{\mu}^2/d_n^2$ for some positive and finite constant $\widetilde{\mu}^2$; (2) Strong identification and fixed alternative whereby $\widetilde{\mu}_n^2 = \widetilde{\mu}^2/d_n^2$ and $\Delta = \widetilde{\Delta}$; (3) Weak identification and fixed alternative where $\Delta = \widetilde{\Delta}$ and $\widetilde{\mu}_n^2 \to \widetilde{\mu}^2$, where $\widetilde{\mu}^2$ is some finite positive value. Note that weak identification and local alternative is not discussed since it has no power. Defining $\Lambda_{0,i}(\Delta) := \mathbb{E}(\widetilde{e}_i, \Delta\widetilde{v}_i)(\widetilde{e}_i, \Delta\widetilde{v}_i)'$, we make the following assumption:

Assumption 4. *For every sequence of $\Delta_n \to \Delta^\dagger \in \mathbb{R}$, suppose $\frac{1}{n}\sum_{i\in[n]} \Lambda_{0,i}(\Delta_n) \otimes Z_i Z_i' \to \Sigma(\Delta^\dagger)$ and $\frac{Z'Z}{n} \to Q_{ZZ}$, where $\Sigma(\Delta^\dagger)$ and Q_{ZZ} are positive-definite matrices. Furthermore, assume that $\sup_i ||Z_i||_F < \infty$.*

Under the assumption that the errors in the DGP of (2.1) are independent and identically distributed, the assumption that $\frac{1}{n}\sum_{i\in[n]} \Lambda_{0,i}(\Delta_n) \otimes Z_i Z_i' \to \Sigma(\Delta^\dagger)$ in assumption 4 can be removed.

Recall from (2.8) that the power of our proposed test involves the critical value that is itself random. This randomness comes from the limit of eigenvalues of $D_{\widetilde{w}_n} := diag(\widetilde{w}_{1,n}, ..., \widetilde{w}_{K,n})$. Since this is generally unknown, in order to derive the power properties of our test under fixed K, we begin by showing some intermediate asymptotic properties pertaining to the critical value (2.7).

Lemma 4.1. *Suppose Assumption 1, 2, 4 holds and we are under fixed K. Consider any estimator $\widehat{\Phi}_1(\beta_0)$ satisfying (2.9). Then for fixed Δ we have*

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}} = O_p(1)$$

Under the alternative, for fixed K, the limiting distribution of the critical value C_α (see (A.20) for its expression) becomes that of a weighted chi-square F_{w^*}-distribution. Given that the limit w^* is unknown in practice, in order to discuss the power properties of our test, one straightforward method is to find the worst-case power property, i.e. we want to examine the values of $w^* = (w_1^*, ..., w_K^*)$ such that $||w^*||_F = 1$, $w_i^* \geq 0$ and $q_{1-\alpha}(F_{w^*})$ is the largest it can be. We have the following result due to Fleiss (1971):

Lemma 4.2. *For any vector $a \in \mathbb{R}^K$ for some fixed dimension K such that $\sum_{i\in[K]} a_i = 1$ and each $a_i \geq 0$, we have*

$$q_{1-\alpha}(\chi_1^2) \geq q_{1-\alpha}\left(\sum_{\ell\in[K]} a_\ell \chi_{1,\ell}^2\right)$$

where the $\chi_{1,\ell}^2$ are independent chi-squares with one-degree-of-freedom

Note that for fixed K, by expression (A.20), Lemma 4.1 and 4.2, we can obtain an upper bound for the power of our test under the worst-case scenario's power

$$\mathbb{P}\left(\widehat{Q}(\beta_0) > q_{1-\alpha}(\chi^2(1)) + O_p(1)\right) \leq \mathbb{P}\left(\widehat{Q}(\beta_0) > q_{1-\alpha}(F_{\widetilde{w}_n}) + O_p(1)\right)$$

Combining lemmas 4.1 and 4.2, we can show that our test consistently differentiates the null from the alternative. The requirement is that the concentration parameter $\widetilde{\mu}_n^2$ diverges to infinity. This requirement is similar to Mikusheva and Sun (2022)[Theorem 1] (this was established for diverging instruments), which shows that for any set of bounded concentration parameter, there is no test that can consistently differentiate the null from the alternative. This result is formally given as:

Theorem 6. *Suppose Assumption 1, 2, 4 holds and we are under fixed K. For any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (2.9), our test consistently differentiates the null from alternative, i.e.*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = 1$$

for any fixed $\Delta \neq 0$, whenever $\widetilde{\mu}_n^2 \to \infty$.

To simplify the discussion for the power properties of the remaining cases, we assume without loss of generality that under weak identification, $\mu_K \equiv \widetilde{\mu}$,[14] while under strong identification, $d_n\mu_K \equiv \widetilde{\mu}$, where $\widetilde{\mu} \in \mathbb{R}^K$ is some constant. Denote $\Omega^*(\beta_0) := \lim_{n\to\infty} \Omega(\beta_0)$ defined in (2.6). We have the following result:

Theorem 7. *Suppose Assumption 1, 2, 4 holds and we are under fixed K. Furthermore, let $\frac{p_n \Pi'\Pi}{K} = O(1)$ and suppose $\Omega^*(\beta_0)$ is well-defined. Then under strong-identification and local alternative, for*

[14]Under weak identification, $\mu_K'\mu_K \equiv \widetilde{\mu}_n^2 \to \widetilde{\mu}^2 \in \mathbb{R}$. This implies that μ_K must be bounded. By Bolzano-Weierstrass, for every sub-sequence of μ_K, there exists a further sub-sequence μ_{K_j} that converges to μ, where $\mu'\mu = \widetilde{\mu}^2$. Therefore, instead of arguing along sub-sequences, the simplification that $\mu_K \equiv \widetilde{\mu}$ allows us to argue along the full sequence.

any estimator $\widehat{\Phi}_1(\beta_0)$ *that satisfies (2.9),*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = \mathbb{P}\left(\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)' \Omega^*(\beta)\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*})\right)$$

where $w^* = (w_1^*, ..., w_K^*)$ *are the eigenvalues of* $\Omega^*(\beta_0)$

Note that $w_i^* \geq 0$ and $\sum_{i\in[K]} w_i^* = 1$. We can diagonalize $\Omega^*(\beta_0) = Q^{*'} D^* Q^*$ such that $Q^* Q^{*'} = Q^{*'} Q^* = I_K$, with $D^* = diag(w_1^*, ..., w_K^*)$. Then we can express the asymptotic power under strong-identification and local alternative as

$$\mathbb{P}\left(\sum_{i\in[K]} w_i^* \chi^2_{1,i}(\mathbb{M}_i) > q_{1-\alpha}(\sum_{i\in[K]} w_i^* \chi^2_{1,i})\right)$$

where $\mathbb{M}_i := \widetilde{\Delta}^2(\iota_i' Q^* \Sigma(0)\widetilde{\mu})^2$ is the non-centrality parameter, by which the power of the test depends on. Furthermore, we can show that our test has certain desirable properties; in particular, our test is admissible in some class of tests. Consider the test $\phi_{\alpha,w^*} := 1\left\{\sum_{i\in[K]} w_i^* \chi^2_{1,i}(\mathbb{M}_i) > q_{1-\alpha}(\sum_{i\in[K]} w_i^* \chi^2_{1,K})\right\}$. Then we have the following result.

Corollary 4.1. *Let* Φ_α *be the class of size-*α *tests for* $H_0 : \mathbb{M}_1 = ... = \mathbb{M}_K = 0$ *constructed based on* K *independent chi-squares* $(\chi^2_{1,i}, ..., \chi^2_{1,K})$*. Then* ϕ_{α,w^*} *is an admissible test within* Φ_α*.*

Corollary 4.1 relates back to Theorem 7 in the sense that our test is admissible over tests in some class under strong-identification and local-alternative. Finally, we can express the asymptotic power of our test under weak-identification and fixed alternative:

Theorem 8. *Suppose Assumption 1, 2, 4 holds and we are under fixed* K*. Assume* $\Omega^*(\beta_0)$ *is well-defined. Then under weak-identification and fixed alternative, if we further assume that* $\Pi'\Pi = O(1)$*, then for any estimator* $\widehat{\Phi}_1(\beta_0)$ *that satisfies (2.9),*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) = \mathbb{P}\left(\mathcal{Z}\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right)' \Omega^*(\beta_0)\mathcal{Z}\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*})\right)$$

where w^* *are the eigenvalues of* $\Omega^*(\beta_0)$ *defined in (2.6).*

Note that the assumption of $\Pi'\Pi = O(1)$ automatically implies weak-identification for fixed K. To see this, observe that WPA1,

$$\widetilde{\mu}_n^2 = \mu_K' \mu_K \leq \lambda_{max}(Q_{ZZ}) \cdot \mu_K' \left(\frac{Z'Z}{n}\right)^{-1} \mu_K = \lambda_{max}(Q_{ZZ})\Pi' P \Pi \leq \lambda_{max}(Q_{ZZ}) \cdot \Pi'\Pi,$$

so that $\widetilde{\mu}_n^2 \leq C$ for some constant $C < \infty$. As before, we can re-write the asymptotic power given in Theorem 8 as

$$\mathbb{P}\left(\sum_{i\in[K]} w_i^* \chi^2_{1,i}(\overline{\mathbb{M}}_i) > q_{1-\alpha}(\sum_{i\in[K]} w_i^* \chi^2_{1,i})\right)$$

where $\overline{\mathbb{M}}_i := \widetilde{\Delta}^2(\iota_i' Q^* \Sigma(\widetilde{\Delta})\widetilde{\mu})^2$ is the non-centrality parameter. This ensures that our test statistic has power strictly greater than α. The asymptotic rejection criteria can be written as a test $\overline{\phi}_{\alpha,w^*} := 1\left\{\sum_{i\in[K]} w_i^* \chi^2_{1,i}(\overline{\mathbb{M}}_i) > q_{1-\alpha}(\sum_{i\in[K]} w_i^* \chi^2_{1,i}\right\}$. Analogous to Theorem 7, we have the result that under weak-identification and fixed-alternative, our test statistic is admissible in some class of test. This follows from the following corollary.

Corollary 4.2. *Let $\overline{\Phi}_\alpha$ be the class of size-α tests for $H_0 : \overline{\mathbb{M}}_1 = ... = \overline{\mathbb{M}}_K = 0$ constructed based on K independent chi-squares $(\chi^2_{1,i}, ..., \chi^2_{1,K})$. Then $\overline{\phi}_{\alpha,w^*}$ is an admissible test within $\overline{\Phi}_\alpha$.*

5 Simulation and Application

In this section, we compare the difference in power and size between existing tests and our test, under two different data generating processes (DGP). To begin, we explicitly define these tests and their corresponding critical-values.

5.1 Description of Tests

We consider the following tests:

(1) Our proposed test using the standard estimator which rejects whenever

$$\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1^{standard}(\beta_0))$$

(2) Our proposed test using the cross-fit estimator, which rejects whenever

$$\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1^{cf}(\beta_0))$$

(3) the Jackknifed AR-statistic for diverging K provided by Mikusheva and Sun (2022), which rejects whenever

$$\frac{1}{\sqrt{\widehat{\Phi}_1^{cf}(\beta_0)}\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ij} e_i(\beta_0) e_j(\beta_0) > q_{1-\alpha}\left(\mathcal{N}(0,1)\right);$$

(4) the standard estimator for diverging K by Crudu et al. (2021) which rejects whenever

$$\frac{1}{\sqrt{\widehat{\Phi}_1^{standard}(\beta_0)}\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}e_i(\beta_0)e_j(\beta_0) > q_{1-\alpha}\left(\mathcal{N}(0,1)\right);$$

(5) The classical AR-statistic for fixed K, i.e. we reject whenever

$$J_n'\widehat{\Omega}_n^{-1}J_n > q_{1-\alpha}(\chi^2_K), \;\; where\ J_n := n^{-1/2}Z'e(\beta_0)\ and\ \widehat{\Omega}_n := \frac{1}{n}Z'\{diag(e_1^2(\beta_0),...,e_n^2(\beta_0))\}Z$$

(6) the Jackknifed-AR for fixed K and homoskedastic errors given by Mikusheva and Sun (2022)[Supplementary Appendix, Lemma S4.1], which rejects whenever

$$\frac{1}{\sqrt{\widehat{\Phi}_1^{cf}(\beta_0)}\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}e_i(\beta_0)e_j(\beta_0) > q_{1-\alpha}\left(\frac{\chi^2_K-K}{\sqrt{2K}}\right)$$

We denote the tests $(1),(2),(3),(4),(5),(6)$ by $Q_{standard}, Q_{cf}, AR_{cf}, AR_{standard}, AR_{classical}$ and JAR_{homo} respectively. We conduct $1,000$ simulation replications to obtain stable results.

5.2 Simulation Based on Hausman, Newey, Woutersen, Chao, and Swanson (2012)

We consider the following model based on the DGP given by Hausman et al. (2012), with sample size $n = 400$, and vary the number of instruments $K \in \{1,2,3,4,5,6,8,10,15,20,40,100,200,300\}$. Let

$$\begin{aligned}
&Y = \beta X + W\Gamma + D_{z_1}U_1\\
&X = \pi_K z_1 + U_2\\
&W = (1,...,1)' \in \mathbb{R}^n\\
&U_1 = \rho_1 U_2 + \sqrt{\frac{1-\rho_1^2}{\phi^2+0.86^4}}(\phi v_1 + 0.86 v_2),\\
&z_{i1} \sim \mathcal{N}(0.5,1), \quad v_{1i} \sim \mathcal{N}(0, z_{i1}^2), \quad v_{2i} \sim \mathcal{N}(0, 0.86^2),\\
&D_{z_1} := diag(\sqrt{1+z_{11}^2}, \sqrt{1+z_{21}^2},..., \sqrt{1+z_{n1}^2})\\
&U_{2i} \sim exponential(0.5) - 2, \quad \phi = 0.3, \quad \rho_1 = 0.3
\end{aligned}$$

We assume that the errors across different i are independent. Furthermore, $z_1 = (z_{11}, z_{21}, ..., z_{n1})$ are independent from any error terms, and $\pi_K \in \mathbb{R}$ is chosen to be such that the identification strength is small; since the value of K affects identification strength, we have different values of π_K for different instruments. We consider values of π_K such that for each K, the concentration parameter $\overline{\mathcal{C}} \approx 17.5$.[15] The diagonal matrix D_{z1} allows U_1 to be dependent on z_1 but at the same time has variance bounded away from zero, in the event some elements of z_1 are close to zero. We assume $\beta = 0$ and $\mathbb{\Gamma} = 1$ to be the true parameters.

The ith instrument observation for $K \geq 6$ is given by

$$Z_i' := (z_{1i}, z_{1i}^2, z_{1i}^3, z_{1i}^4, z_{1i}^5, z_{1i}D_{i1}, ..., z_{1i}D_{i,K-5}),$$

where $D_{ik} \in \{0, 1\}$ is a dummy variable with $\mathbb{P}(D_{ik} = 1) = 1/2$, so that $Z_i \in \mathbb{R}^K$. For $K \leq 5$, the ith instrument observation is

$$\begin{aligned}
&Z_i' := z_{i1} \quad \text{for } K = 1,\\
&Z_i' := (z_{i1}, z_{i2}) \text{ for } K = 2,\\
&Z_i' := (z_{i1}, z_{i2}, z_{i1}z_{i2}) \text{ for } K = 3,\\
&Z_i' := (z_{i1}, z_{i2}, z_{i1}z_{i2}, z_{i1}^2) \text{ for } K = 4,\\
&Z_i' := (z_{i1}, z_{i2}, z_{i1}z_{i2}, z_{i1}^2, z_{i2}^2) \text{ for } K = 5,\\
&z_{i2} \sim \mathcal{N}(0.5, 1) \text{ independent of } z_{i1}
\end{aligned}$$

Note that $z_2 := (z_{12}, z_{22}, ..., z_{n2})'$ does not affect the DGP, so that in some sense it is a 'spurious' instrument. It is added for smaller instruments to ensure that the $\overline{C}$ in assumption 3 is not too large. We detail the probability of rejection under the null of $\beta = \beta_0$ in the following table.

[15] We used the command 'set.seed(1)' for our simulation in R programming so that Z can be pinned down without changing. After this was done, we calibrated the value of π so that $\overline{\mathcal{C}} := \frac{(\pi z_1)' P_0 (\pi z_1)}{\sqrt{K}} = 17.5$ for each K, where $P_0 := P - diag(P)$ and $P := M^W Z(Z' M^W Z)^{-1}(M^W)Z'$. Note that π changes with K. Furthermore, through extensive simulation, the results will not change much when $\overline{\mathcal{C}}$ changes by a little, say ± 2.

Table 1: **Rejection Probability under Null**

	$AR_{standard}$ (5%)	$\mathbf{Q_{standard}}$ (5%)	AR_{cf} (5%)	$\mathbf{Q_{cf}}$ (5%)	$AR_{classical}$ (5%)	JAR_{homo} (5%)
$K = 1$	**0.071**	0.048	**0.074**	0.052	0.048	0.052
$K = 2$	0.064	0.05	0.065	0.053	0.056	0.051
$K = 3$	**0.086**	0.067	**0.091**	**0.073**	0.057	**0.072**
$K = 4$	**0.083**	0.058	**0.094**	0.069	0.05	**0.07**
$K = 5$	**0.08**	0.058	**0.092**	0.066	0.062	**0.077**
$K = 6$	**0.075**	0.046	**0.131**	**0.104**	0.039	**0.106**
$K = 8$	**0.079**	0.048	**0.121**	**0.1**	0.033	**0.108**
$K = 10$	**0.086**	0.061	**0.131**	**0.107**	0.031	**0.109**
$K = 15$	0.066	0.035	**0.095**	**0.071**	0.034	**0.079**
$K = 20$	0.0567	0.047	**0.094**	**0.073**	0.022	**0.077**
$K = 40$	0.056	0.038	**0.085**	0.068	0.014	**0.073**
$K = 100$	0.056	0.042	**0.85**	0.065	0.002	**0.075**
$K = 200$	0.061	0.04	**0.112**	**0.085**	0	**0.101**
$K = 300$	0.053	0.043	**0.115**	**0.96**	0	**0.109**

Note: $AR_{standard}, Q_{standard}, AR_{cf}, Q_{cf}, AR_{classic}, JAR_{homo}$ corresponds to (4), (1), (3), (2), (5), (6) discussed in Section 5.1 respectively. We reject at the 95% confidence-level (i.e. $\alpha = 0.05$) and **bold** values which are greater than or equal to **0.07**.

Table 1 provides the probability of rejection under the null for different values of K. The $AR_{standard}$ suffers from size issues when the number of instrument is small. Our corresponding proposed test $Q_{standard}$ resolves this. Similarly, severe size distortion occurs for AR_{cf}[16]; our corresponding test Q_{cf} tries to resolve this, albeit partially successful. However, notice that Q_{cf} reduces the size distortion by about $20\% - 30\%$. The classical AR test for fixed K, denoted $AR_{classical}$ does not suffer size distortion; however, we see that it suffers from substantial power decline for larger values of instruments, say $K \geq 6$, as seen from Figure 4–8. Finally, the Jackknifed-AR denoted by JAR_{homo} suffers from size-distortion even for small instruments, say $K = 3$. This is expected since the critical value of the test is based on homoskedastic errors, while the errors of the DGP are heteroskedastic.

[16] The size-distortion of AR_{cf} persists even under large K (say $K \geq 200$) due to $p_n := \max_i P_{ii}$ being very close to one (it is roughly 0.992 in the simulation when $K = 300$). Recall from Theorem C.0.2 that one of the key assumptions in assuring $\widehat{\Phi}_1^{cf}(\beta_0)$ satisfies (2.9) is that $p_n \leq \delta < 1$ for some δ. Note that even though this assumption was made in Theorem C.0.1, it is actually not needed for the consistency of $\widehat{\Phi}_1^{standard}(\beta_0)$, which explains why $AR_{standard}$ has reasonable size for larger K.

In order to obtain a fair power-comparison between the tests due to size-distortion, for each given K we compute the $(1-\alpha)$-quantile of each distribution under the null, using $1,000$ sample points from the test statistic. We then reject the tests whenever the test statistic is greater than this null-computed quantile.[17]

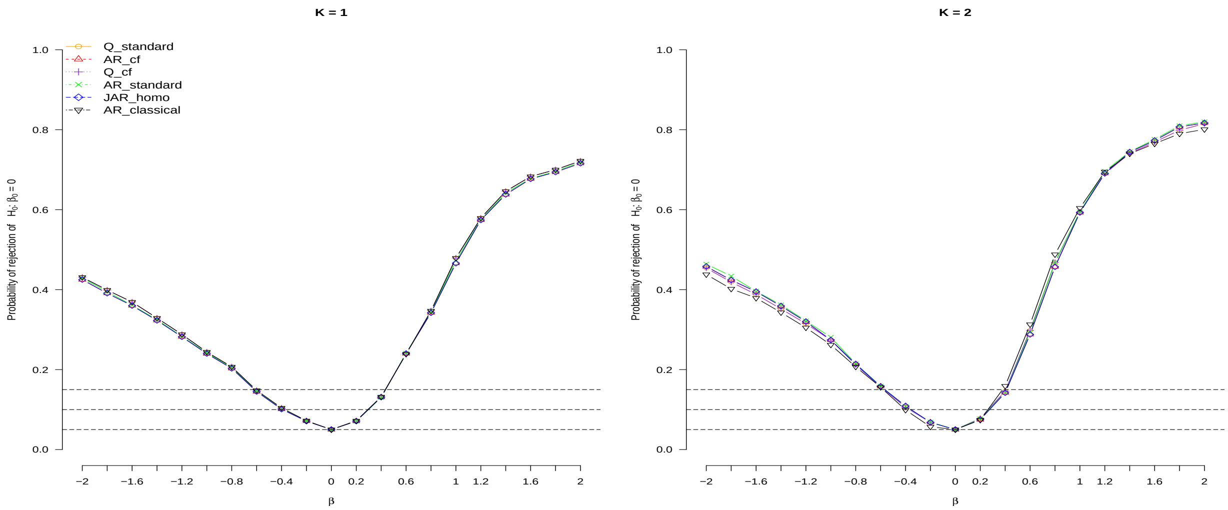

Figure 2: Power curve for $K = 1, 2$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

[17] Note that these null-computed quantiles are in general infeasible in the sense that they cannot be constructed without knowing the true DGP and parameters

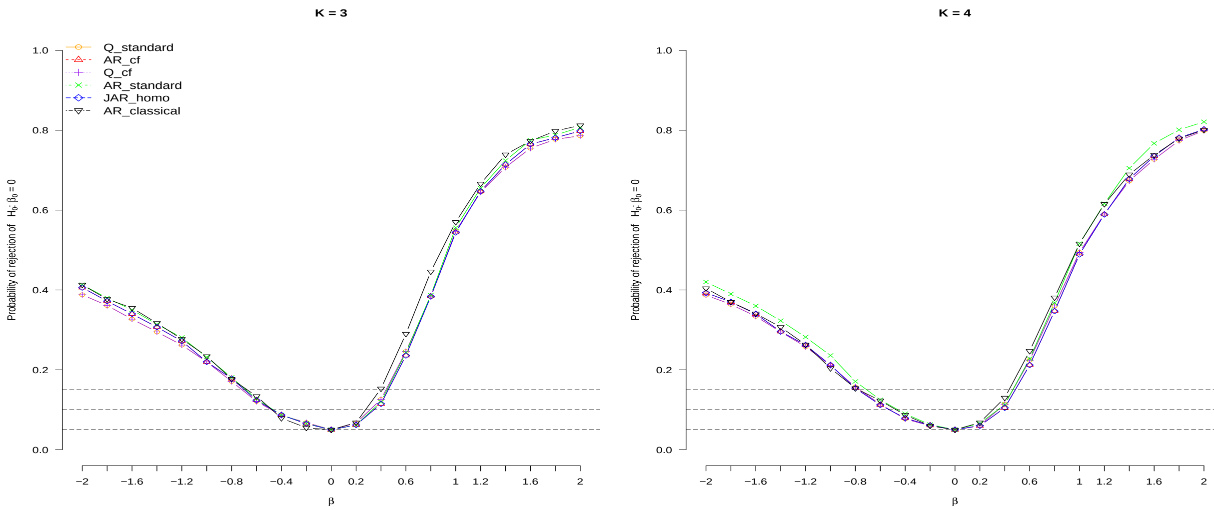

Figure 3: Power curve for $K = 3, 4$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

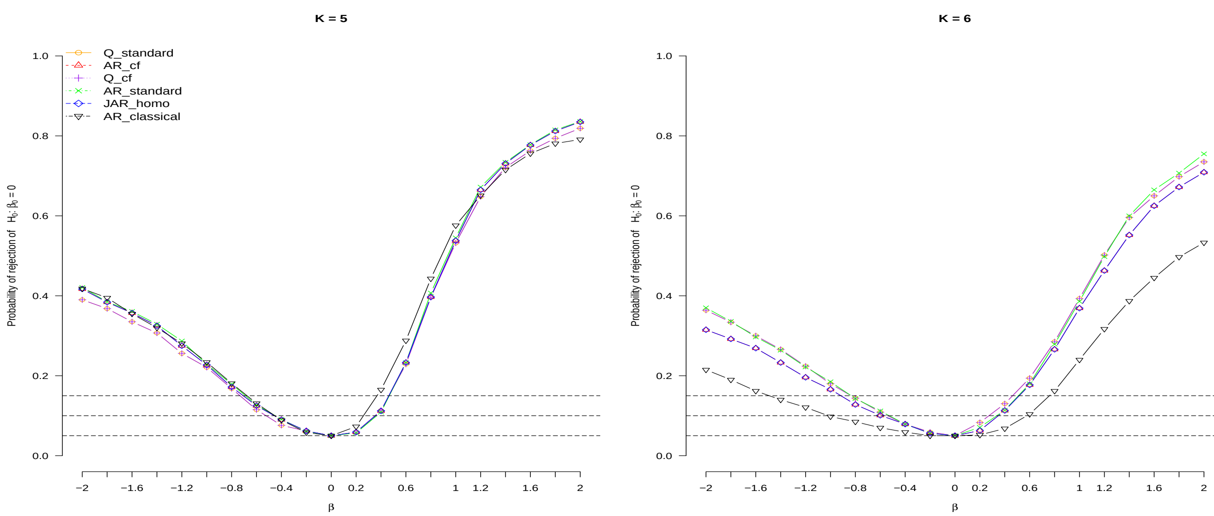

Figure 4: Power curve for $K = 5, 6$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

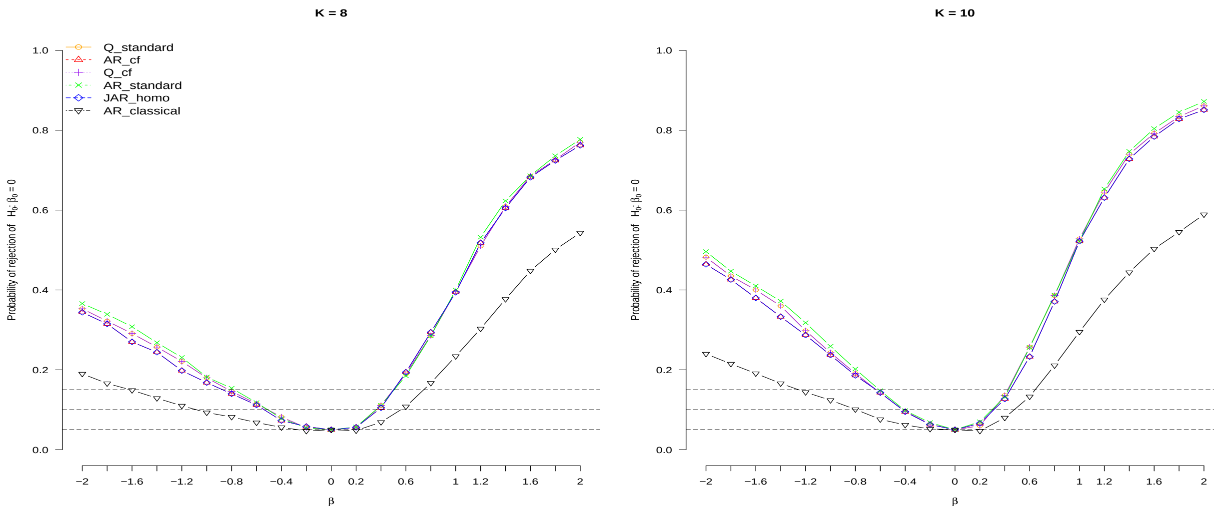

Figure 5: Power curve for $K = 8, 10$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

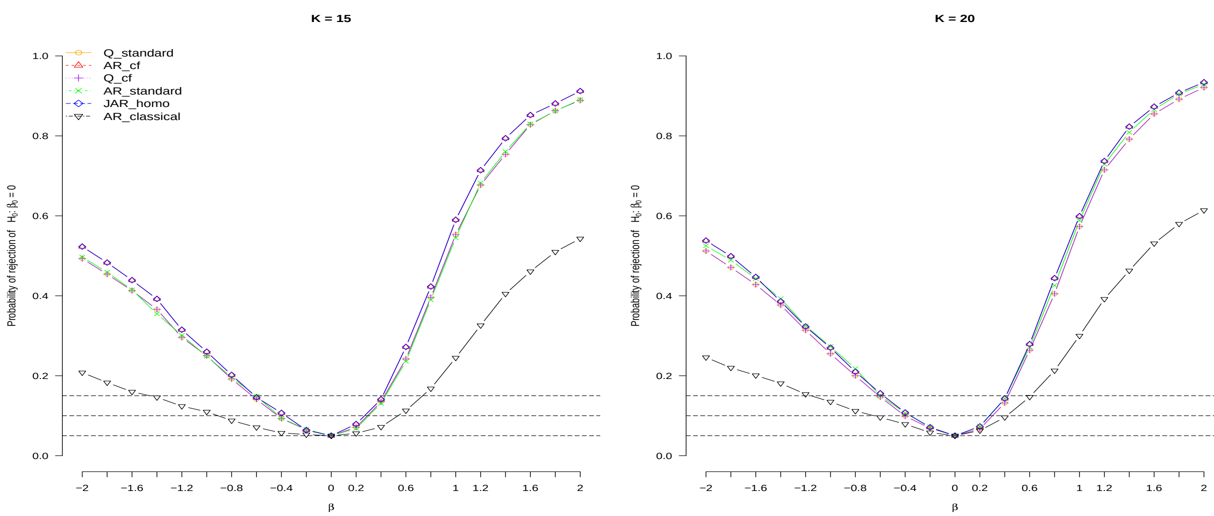

Figure 6: Power curve for $K = 15, 20$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

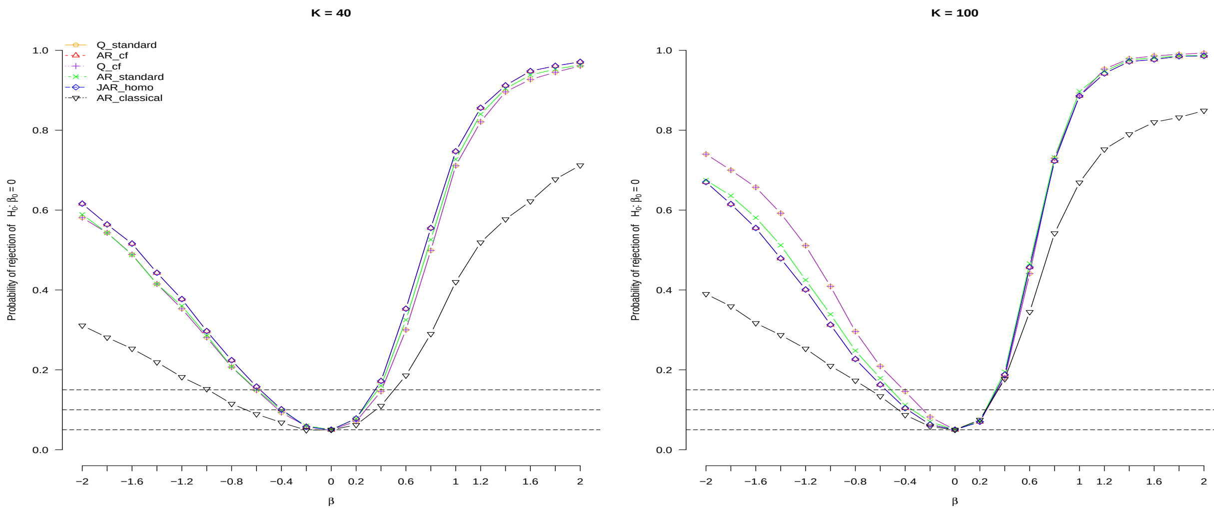

Figure 7: Power curve for $K = 40, 100$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

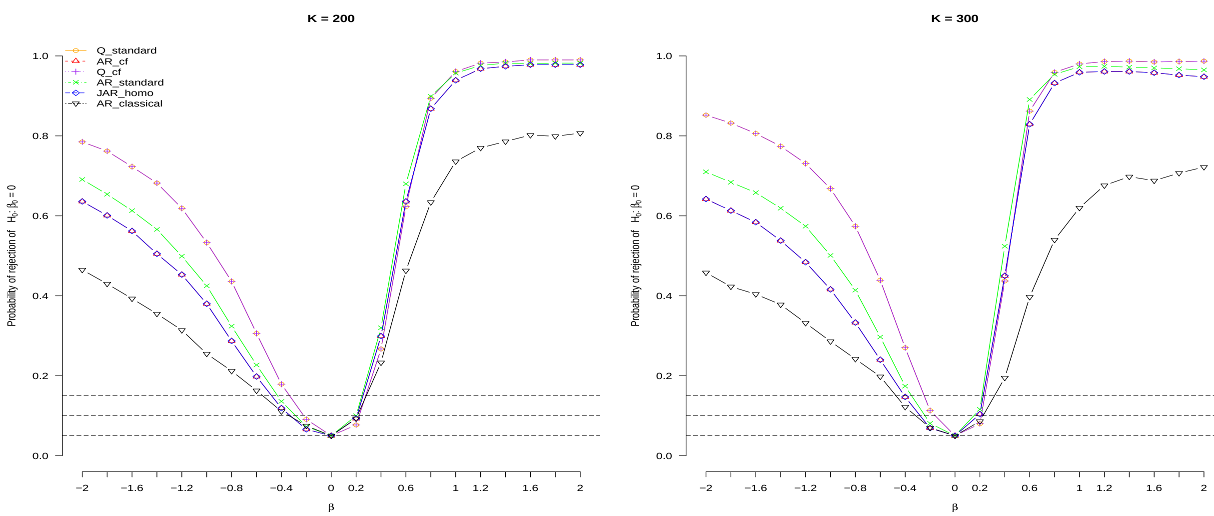

Figure 8: Power curve for $K = 200, 300$

Note: The orange line with circle represents $Q_{standard}$; the red line with upward-pointing triangle represents AR_{cf}; the purple line with a cross represents Q_{cf}; the green line with a cross represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with downward-pointing triangle represents $AR_{classical}$. The first horizontal dotted black line represents 5%; the second represents 10%; the third represents 15%.

Figures 2-8 plot the size-adjusted power curve for the aforementioned tests. We make a few observations. First, the two proposed tests $Q_{standard}$ and Q_{cf} (which we simply call Q-tests) have similar power over different instruments, which is expected as their rejection rate are asymptotically equal under every alternative. Second, the size-adjusted power of the Q-tests is as good as the $AR_{standard}$ and AR_{cf} when the number of instruments is large (say $K \geq 100$), while all tests have approximately equal power whenever the number of instruments is small (say $K \leq 5$). Third, for moderate to large number of instruments (say $K \geq 6$), the power of the $AR_{classical}$ is comparatively lower than the other tests. Finally, when the number of instruments is large, the power curves for AR_{cf} and JAR_{homo} are similar because the two tests differ only in the critical value used (i.e. $q_{1-\alpha}(\mathcal{N}(0,1))$ for the former and $q_{1-\alpha}(\frac{\chi^2_K - K}{\sqrt{2K}})$ for the latter). As $K \to \infty$, $\frac{\chi^2_K - K}{\sqrt{2K}} \rightsquigarrow \mathcal{N}(0,1)$, so that eventually, for larger instruments, the rejection rate of these two tests should be equal.

5.3 Empirical Application

In this section, we consider the linear IV regression with underlying specification based on Angrist and Krueger (1991), using the full original dataset.[18] In particular, we consider the 1980s census of 329,509 men born in 1930-1939 based on Angrist and Krueger's (1991) dataset. The model follows Mikusheva and Sun (2022), which can be written explicitly as

$$\begin{aligned} ln\ W_i &= Constant + H_i^\top \zeta + \sum_{c=30}^{38} YOB_{i,c}\xi_c + \sum_{s \neq 56} POB_{i,s}\eta_s + \beta E_i + \gamma_i \\ E_i &= Constant + H_i^\top \lambda + \sum_{c=30}^{38} YOB_{i,c}\mu_c + \sum_{s \neq 56} POB_{i,s}\alpha_s + Z_{i,K} + \varepsilon_i \end{aligned} \tag{5.1}$$

where W_i is the weekly wage, E_i is the education of the i-th individual, H_i is a vector of covariates,[19] $YOB_{i,c}$ is a dummy variable indicating whether the individual was born in year $c = \{30, 31, ..., 39\}$, while $QOB_{i,j}$ is a dummy variable indicating whether the individual was born in quarter-of-birth $j \in \{1, 2, 3, 4\}$. $POB_{i,s}$ is the dummy variable indicating whether the individual was born in state $s \in \{51\ states\}$.[20] Both γ_i and ε_i are the error terms. We consider twenty-one varying numbers of instruments; in particular,

$$K = \{3, 10, 20, 30, 50, 100, 150, 180, 200, 250, 300, 350, 400, 450, 600, 765, 918, 1071, 1224, 1377, 1530\},$$

[18] The dataset can be downloaded from MIT Economics, Angrist Data Archive, https://economics.mit.edu/faculty/angrist/data1/data/angkru1991.

[19] The covariates we consider are: RACE, MARRIED, SMSA, NEWENG, MIDATL, ENOCENT, WNOCENT, SOATL, ESOCENT, WSOCENT, and MT.

[20] The state numbers are from 1 to 56, excluding (3,7,14,43,52), corresponding to U.S. state codes.

so that $Z_{i,K}$ varies with K. Specifically, we have

$$Z_{i,3} = \sum_{j=1}^{3} QOB_{i,j}\delta_j,$$

$$Z_{i,10} = \sum_{j=1}^{1}\sum_{c=30}^{39} QOB_{i,j}YOB_{i,c}\theta_{j,c}, ..., Z_{i,30} = \sum_{j=1}^{3}\sum_{c=30}^{39} QOB_{i,j}YOB_{i,c}\theta_{j,c},$$

$$Z_{i,50} = \sum_{j=1}^{1}\sum_{s\neq 56} QOB_{i,j}POB_{i,s}\delta_{j,s}, ..., Z_{i,150} = \sum_{j=1}^{3}\sum_{s\neq 56} QOB_{i,j}POB_{i,s}\delta_{j,s},$$

$$Z_{i,180} = \sum_{j=1}^{3}\sum_{s\neq 56} QOB_{i,j}POB_{i,s}\delta_{j,s} + \sum_{j=1}^{3}\sum_{c=30}^{39} QOB_{i,j}YOB_{i,c}\theta_{j,c},$$

$$Z_{i,200} = \sum_{c=30}^{33}\sum_{s\neq 56} YOB_{i,j}POB_{i,s}QOB_{1,j}\psi_{c,s}, ..., Z_{i,450} = \sum_{c=30}^{38}\sum_{s\neq 56} YOB_{i,j}POB_{i,s}QOB_{1,j}\psi_{c,s},$$

$$Z_{i,600} = \sum_{c=30}^{38}\sum_{s\neq 56} YOB_{i,j}POB_{i,s}\psi_{c,s} + \sum_{j=1}^{3}\sum_{s\neq 56} QOB_{i,j}POB_{i,s}\delta_{j,s},$$

$$Z_{i,765} = \sum_{c=30}^{34}\sum_{j=1}^{3}\sum_{s\in\{51\ states\}} QOB_{i,j}YOB_{i,c}POB_{i,s}\delta_{j,c,s}, ...$$

$$..., Z_{i,1071} = \sum_{c=30}^{39}\sum_{j=1}^{3}\sum_{s\in\{51\ states\}} QOB_{i,j}YOB_{i,c}POB_{i,s}\delta_{j,c,s}$$

The coefficient β is the return to education. We vary this β across 1,000 equidistant grid-points from -0.5 to 0.5 (i.e., $\beta \in \{-0.5, -0.499, -0.498, ..., 0, ..., 0.499, 0.5\}$) and solve for the range of β where the null hypothesis cannot be rejected, according to section 5.1. Specifically, we can write the above model as

$$ln\ W_i = C_i\Gamma + \beta E_i + \gamma_i \tag{5.2}$$

$$E_i = C_i\tau + Z_i\Theta + \varepsilon_i, \tag{5.3}$$

where C_i is a (329,509×71)-matrix of controls containing the first four terms on the right-hand of (5.1). We can then partial out the controls C_i by multiplying each equation (5.2) and (5.3) by the residual matrix $I - C(C^\top C)^{-1}C^\top$ to obtain a form analogous to that in the main text:

$$Y_i = X_i\beta + e_i,$$

$$X_i = \Pi_i + v_i$$

Then, at each grid-point we take $\beta_0 = \beta$ and compute $AR_{standard}, Q_{standard}, AR_{cf}, Q_{cf}, AR_{classical}$ and JAR_{homo}. We reject the chosen value of β_0 for if it exceeds the one-sided 5%-quantile of the corresponding critical-value (i.e. $\alpha = 0.05$ with the tests and their critical-value described in Section 5.1). Note that the full QOB, YOB, POB or their interactions are not used in order to avoid multicollinearity. We report the upper and lower bounds of the confidence set for which the null cannot be rejected in Table 2 below.

Table 2: **Confidence Interval**

	$AR_{standard}$ (5%)	$\mathbf{Q_{standard}}$ (5%)	AR_{cf} (5%)	$\mathbf{Q_{cf}}$ (5%)	$AR_{classical}$ (5%)	JAR_{homo} (5%)
$K = 3$	[0.056,0.147]	[0.053,0.15]	[0.056,0.147]	[0.053,0.15]	[0.053,0.151]	[0.052,0.151]
$K = 10$	[-0.007,0.16]]	[-0.011,0.165]	[-0.007,0.16]]	[-0.011,0.165]	[-0.011,0.166]	[-0.011,0.165]
$K = 20$	[0.017,0.174]	[0.014,0.178]	[0.017,0.174]	[0.014,0.178]	[0.014,0.18]	[0.014,0.178]
$K = 30$	[0,0.169]	[-0.002,0.172]	[0,0.169]	[-0.002,0.172]	[-0.002,0.177]	[-0.002,0.172]
$K = 50$	[0.005,0.183]	[0.002,0.188]	[0.005,0.183]	[0.002,0.188]	[-0.01,0.188]	[0.002,0.188]
$K = 100$	[0.018,0.2]	[0.017,0.202]	[0.018,0.2]	[0.017,0.202]	[0.009,0.203]	[0.017,0.202]
$K = 150$	[0.023,0.208]	[0.022,0.21]	[0.023,0.208]	[0.022,0.21]	[0.022,0.212]	[0.022,0.21]
$K = 180$	[0.008,0.201]	[0.007,0.202]	[0.008,0.202]	[0.007,0.202]	[0.007,0.207]	[0.007,0.202]
$K = 200$	[-0.216,0.23]	[-0.223,0.233]	[-0.218,0.23]	[-0.224,0.233]	[-0.214,0.236]	[-0.224,0.233]
$K = 250$	[-0.118,0.258]	[-0.122,0.261]	[-0.12,0.258]	[-0.123,0.261]	[-0.111,0.256]	[-0.122,0.261]
$K = 300$	[-0.097,0.24]	[-0.1,0.242]	[-0.098,0.24]	[-0.1,0.242]	[-0.085,0.238]	[-0.1,0.242]
$K = 350$	[-0.107,0.28]	[-0.11,0.283]	[-0.108,0.28]	[-0.111,0.283]	[-0.092,0.274]	[-0.11,0.283]
$K = 400$	[-0.078,0.305]	[-0.081,0.308]	[-0.079,0.305]	[-0.081,0.308]	[-0.058,0.298]	[-0.081,0.308]
$K = 450$	[-0.105,0.29]	[-0.107,0.293]	[-0.106,0.29]	[-0.108,0.293]	[-0.092,0.281]	[-0.107,0.293]
$K = 600$	[-0.018,0.228]	[-0.019,0.229]	[-0.019,0.228]	[-0.019,0.229]	[-0.013,0.224]	[-0.019,0.229]
$K = 765$	[-0.09,0.192]	[-0.093,0.194]	[-0.09,0.192]	[-0.093,0.194]	[-0.125,0.163]	[-0.092,0.194]
$K = 918$	[-0.055,0.182]	[-0.058,0.183]	[-0.055,0.182]	[-0.057,0.184]	[-0.076,0.157]	[-0.056,0.183]
$K = 1071$	[-0.042,0.19]	[-0.044,0.192]	[-0.041,0.19]	[-0.043,0.192]	[-0.064,0.168]	[-0.042,0.191]
$K = 1224$						
$K = 1377$						
$K = 1530$						

Note: $AR_{standard}, Q_{standard}, AR_{cf}, Q_{cf}, AR_{classic}, JAR_{homo}$ corresponds to (4), (1), (3), (2), (5), (6) discussed in Section 5.1 respectively.

Table 2 highlights a few salient features of our proposed method, which we discuss in detail. First of all, notice that the proposed statistic $Q_{standard}$ and Q_{cf} has similar confidence intervals (C.I.). Recall from Table 1 that the size-control for Q_{cf} was slightly distorted due to p_n being extremely close to one, a requirement for the validity of the cross-fit variance estimator $\widehat{\Phi}_1^{cf}(\beta_0)$.

In the empirical application p_n is bounded away from one, so that $Q_{standard}$ and Q_{cf} should be expected to be close to each other. This can also be evidenced from $AR_{standard}$'s C.I. being close to AR_{cf} over all values of K. Second, $AR_{classical}$'s C.I. is quite different from all other statistics for larger K, which is to be expected since $AR_{classical}$ is meant for testing under fixed K. However, notice that $Q_{standard}$ and Q_{cf}'s C.I. are close to $AR_{classical}$ for smaller K values, while $Q_{standard}$ and Q_{cf} differs from $AR_{standard}$ and AR_{cf} at these values, which suggests that the C.I. for both $AR_{standard}$ and AR_{cf} may not be valid for smaller K. For large K (say $K \geq 350$), $Q_{standard}$ and Q_{cf}'s C.I. converges to that of $AR_{standard}$ and AR_{cf}. We can therefore see that our proposed test ensures that the C.I. we obtain is correct. Third, JAR_{homo}'s C.I. converges to $AR'_{cf}s$ C.I. as the number of instruments increase. This is expected since the test JAR_{homo} converges to AR_{cf} as $K \to \infty$.

Fourth, comparing Q_{cf} and JAR_{homo} for small K, we see that their C.I. are very similar. We can infer from this that the data seems to be exhibiting homoskedastic variance. This requires some explanation. Consider a fixed Δ not necessarily zero. Note that under some additional assumptions, we can show that under fixed K, WPA1, we have[21]

$$||\widetilde{w}_n - w_n|| \approx 0$$

This implies that WPA1, $F_{\widetilde{w}} \rightsquigarrow F_w$ approximately. Under homoskedasticity, $w_{i,n} = \frac{1}{K}$, so that $F_w = \frac{\chi^2_K}{K}$. Therefore, WPA1 approximately,

$$\frac{q_{1-\alpha}(F_{\widetilde{w}}) - 1}{\sqrt{2}||\widetilde{w}_n||_F} \to q_{1-\alpha}\left(\frac{\chi^2_K/K - 1}{\sqrt{2}\sqrt{\sum_{i\in[K]}\frac{1}{K^2}}}\right) = q_{1-\alpha}\left(\frac{\chi^2_K - K}{\sqrt{2K}}\right)$$

By rearrangement, the rejection criteria for Q_{cf} becomes: reject whenever

$$\frac{1}{\sqrt{K\widehat{\Phi}_1^{cf}(\beta_0)}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)(\widehat{Q}(\beta_0) - 1) > q_{1-\alpha}\left(\frac{q_{1-\alpha}(F_{\widetilde{w}}) - 1}{\sqrt{2}||\widetilde{w}_n||_F}\right) \approx q_{1-\alpha}\left(\frac{\chi^2_K - K}{\sqrt{2K}}\right)$$

Furthermore, recall that the rejection criteria for JAR_{homo} is given as

$$\frac{1}{\sqrt{K\widehat{\Phi}_1^{cf}(\beta_0)}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)(\widehat{Q}(\beta_0) - 1) > q_{1-\alpha}\left(\frac{\chi^2_K - K}{\sqrt{2K}}\right)$$

We therefore conclude that under homoskedasticity, for fixed K, the rejection rate of Q_{cf} and JAR_{homo} should be approximately equal. Since the C.I. of both tests are similar, we can infer

[21] In particular, if we impose the additional assumption that $\max_{i\in[n]} \frac{\Delta^2\Pi_i^2}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \approx 0$, then we can see that this result follows from Lemma B.3

somewhat that the variance is homoskedastic. As a form of robustness check, note that $AR_{classical}$ and JAR_{homo} has similar C.I. for small K, where we recall $AR_{classical}$ is robust to heteroskedasticity under fixed K. This further confirms our intuition. To summarize point four, our proposed tests $Q_{standard}$ and Q_{cf} can serve to check for homoskedastic variance.

References

Anatolyev, S. and M. Solvsten (2023). Testing many restrictions under heteroskedasticity. *Journal of Econometrics 236*(1), 105473.

Andrews, D. W., X. Cheng, and P. Guggenberger (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics 218*(2), 496–531.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Andrews, D. W. K. and J. H. Stock (2007). Testing with many weak instruments. *Journal of Econometrics 138*(1), 24–46.

Angrist, J., G. Imbens, and A. Krueger (1999). Jackknife instrumental variables estimates. *Journal of Applied Econometrics 14*(1), 57–67.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earning? *Quarterly Journal of Economics 106*(4), 979–1014.

Chao, J. C. and N. R. Swanson (2005). Consistent Estimation with a Large Number of Weak Instruments. *Econometrica 73*(5), 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*(1), 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 235*(2), 1747–1769.

Crudu, F., G. Mellace, and Z. Sándor (2021). Inference in instrumental variable models with heteroskedasticity and many instruments. *Econometric Theory 37*(2), 281–310.

Dudley, R. M. (2002). *Real Analysis and Probability* (2 ed.). Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Fleiss, J. L. (1971). On the distribution of a linear combination of independent chi squares. *Journal of the American Statistical Association 66*(333), 142–144.

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*(2), 211–255.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*(4), 1103–1124.

Lehmann, E. L. and J. P. Romano (2006). *Testing statistical hypotheses.* Springer Science & Business Media.

Lim, D., W. Wang, and Y. Zhang (2023). A conditional linear combination test with many weak instruments. *arXiv preprint arXiv:2207.11137*.

Marden, J. I. (1982). Combining independent noncentral chi squared or f tests. *The Annals of Statistics 10*(1), 266–277.

Mikusheva, A. and L. Sun (2022). Inference with many weak instruments. *Review of Economic Studies 89*(5), 2663–2686.

Phillips, G. D. A. and C. Hale (1977). The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review 18*(1), 219–228.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Sawa, T. (1969). The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical Association 64*, 923–937.

Van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Appendix

Contents

A Proofs for Main text

A.1 Proof of Theorem 1

For any vector $a, b \in \mathbb{R}^n$, we define $Q_{a,b} := \frac{\sum_{i\in[n]}\sum_{j\neq i} a_i P_{ij} b_j}{\sqrt{K}}$.

We will first prove the first part of Theorem 1. This is done in **Step 1**–**Step 4**. The proof of the second part of Theorem 1 is shown in **Step 5**.

Recall that $e = \widetilde{e} + P^W \widetilde{e}$ and $\mathcal{E} = \varepsilon + P^W \varepsilon$, so that we have

$$\begin{aligned} Q_{e,e} &= Q_{\widetilde{e},\widetilde{e}} + 2Q_{\widetilde{e},P^W\widetilde{e}} + Q_{P^W\widetilde{e},P^W\widetilde{e}} \\ Q_{\mathcal{E},\mathcal{E}} &= Q_{\varepsilon,\varepsilon} + 2Q_{\varepsilon,P^W\varepsilon} + Q_{P^W\varepsilon,P^W\varepsilon} \end{aligned} \tag{A.1}$$

We want to strongly approximate these two equations. It is instructive to first provide an outline for our proof before delving into it. To do so, consider a sequence of independent random variables $\{(\vartheta_i\}_{i=1}^n$ with the criteria that

$$\begin{aligned} &(i)\ \mathbb{E}\vartheta_i = 0 \\ &(ii)\ \mathbb{E}[\vartheta_i^2] = \mathbb{E}[\widetilde{e}_i^2] = \mathbb{E}[\varepsilon_i^2] \\ &(iii)\ \{(\vartheta_i\}_{i=1}^n \text{ is independent of } \{\widetilde{e}_i\}_{i=1}^n \text{ and } \{\varepsilon_i\}_{i=1}^n \end{aligned}$$

Such a sequence will always exist by the Kolmogorov-Extension-Theorem. This sequence will be used throughout the proof. We define $\vartheta := (\vartheta_1, ..., \vartheta_n)'$.

The idea of the proof is to express

$$Q_{e,e} - Q_{\mathcal{E},\mathcal{E}} = Remainder_n + O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{A.2}$$

The term '$Remainder_n$' collects all the difference in terms that cannot be collected as $O_p(\frac{p_n d_W^2}{K^{1/2}})$-terms. To be precise, **step 1** will imply that $Q_{P^W\widetilde{e},P^W\widetilde{e}} - Q_{P^W\varepsilon,P^W\varepsilon} = O_p(\frac{p_n d_W^2}{K^{1/2}})$, so that this term is collected in the last term of the right-hand-side of (A.2). In **step 2** we deal with the difference between the middle-term on the right-side of (A.1), which implies that

$$2Q_{(\widetilde{e},P^W\widetilde{e}} - 2Q_{\varepsilon,P^W\varepsilon} = \mathcal{H}_n + O_p(\frac{p_n d_W^2}{K^{1/2}})$$

where $\mathcal{H}_n := -\frac{1}{\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i} P_{ii}P_{ij}^W\{\widetilde{e}_i\widetilde{e}_j - \vartheta_i\vartheta_j\}$. Thus $\mathcal{H}_n$ goes into the '$Remainder_n$' term of (A.2), with the remaining terms collected as $O_p(\frac{p_n d_W^2}{K^{1/2}})$-terms. In **step 3** we deal with the first term on the right-side of (A.2) (i.e. $Q_{\widetilde{e},\widetilde{e}} - Q_{\varepsilon,\varepsilon}$) and note that this term goes into '$Remainder_n$'. We will then collect all the terms in '$Remainder_n$' and strongly approximate these terms. Specifically, we can express

$$Remainder_n = F_n - \mathcal{F}_n$$

where

$$F_n := Q_{\widetilde{e},\widetilde{e}} - \frac{2}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \widetilde{e}_i \widetilde{e}_j,$$
$$\mathcal{F}_n := Q_{\varepsilon,\varepsilon} - \frac{2}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \varepsilon_i \varepsilon_j$$

and we strongly-approximate these two terms. Note that F_n is the part of the terms in '$Remainder_n$' that belongs to $Q_{e,e}$, while $\mathcal{F}_n$ belongs to $Q_{\mathcal{E},\mathcal{E}}$. **Step 4** puts everything together and completes the proof for the first part of Theorem 1. **Step 5** completes the proof for the second part of Theorem 1.

Step 1: We show that for any

$$Q_{P^W\widetilde{e},P^W\widetilde{e}} - Q_{P^W\vartheta,P^W\vartheta} = O_p(\frac{p_n d_W^2}{K^{1/2}})$$
$$Q_{P^W\varepsilon,P^W\varepsilon} - Q_{P^W\vartheta,P^W\vartheta} = O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{A.3}$$

Consider first a sequence of independent random variables $\{U_i\}_{i=1}^n$ with bounded first and second moments. Furthermore, let $\{\widetilde{U}_i\}_{i=1}^n$ be independent random variables, as well as independent from $\{U_i\}_{i=1}^n$. Suppose that the $\mathbb{E}U_i = \mathbb{E}\widetilde{U}_i$ and $\mathbb{E}U_i^2 = \mathbb{E}\widetilde{U}_i^2$ for every $i \in [n]$. We will show that

$$Q_{P^W U,P^W U} - Q_{P^W\widetilde{U},P^W\widetilde{U}} = O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{A.4}$$

Note that $PP^W = 0$, so that

$$Q_{P^W U,P^W U} = \frac{1}{\sqrt{K}} U' P^W P P^W U - \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}\{(P_i^W)'U\}^2 = -\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}\{(P_i^W)'U\}^2$$

with $U := (U_1, ..., U_n)'$. Denoting $U_i^* := U_i - \mathbb{E}U_i, \widetilde{U}_i^* := \widetilde{U}_i - \mathbb{E}\widetilde{U}_i$, we have

$$(Q_{P^W U,P^W U} - Q_{P^W\widetilde{U},P^W\widetilde{U}}) = -\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} \left(\left[(P_i^W)'U^* + (P_i^W)'\mathbb{E}U\right]^2 - \left[(P_i^W)'\widetilde{U}^* + (P_i^W)'\mathbb{E}U\right]^2 \right)$$
$$= -\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}[(P_i^W)'U^*]^2 + \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}[(P_i^W)'\widetilde{U}^*]^2 - \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}(P_i^W)'U^*(P_i^W)'\mathbb{E}U$$
$$+ \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}(P_i^W)'\widetilde{U}^*(P_i^W)'\mathbb{E}U \equiv C_1 + C_2 + C_3 + C_4$$

By the fact that $\mathbb{E}U^* = 0$,

$$\mathbb{E}\left|\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}((P_i^W)'U^*)^2\right| = \frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\sum_{\ell\in[n]}(P_{i\ell}^W)^2 Var(U_i) \leq \frac{Cp_n}{\sqrt{K}}\sum_{i\in[n]}\sum_{\ell\in[n]}(P_{i\ell}^W)^2$$
$$= \frac{Cp_n}{\sqrt{K}}\sum_{i\in[n]} P_{ii}^W = \frac{Cp_n d_W}{K^{1/2}},$$

so that by Markov inequality, $C_1 = O_p(\frac{p_n d_W}{K^{1/2}})$. In a similar manner, we can show that $C_2 = O_p(\frac{p_n d_W}{K^{1/2}})$. Next,

$$\begin{aligned}
\mathbb{E}C_3^2 &\leq \frac{1}{K}\sum_{i,i'\in[n]} P_{ii}P_{i'i'}|(P_i^W)'\mathbb{E}U\cdot(P_{i'}^W)'\mathbb{E}U|\sum_{\ell\in[n]}|P_{i\ell}^W P_{i'\ell}^W|Var(U_i)\\
&\overset{(i)}{\leq} \frac{Cp_n^2}{K}\sum_{i,i'\in[n]}|(P_i^W)'\mathbb{E}U\cdot(P_{i'}^W)'\mathbb{E}U|\left\{\sum_{\ell\in[n]}(P_{i\ell}^W)^2\cdot\sum_{\ell\in[n]}P_{i'\ell}^W\right\}\\
&= \frac{Cp_n^2}{K}\sum_{i,i'}|(P_i^W)'\mathbb{E}U\cdot(P_{i'}^W)'\mathbb{E}U|\cdot P_{ii}^W P_{i'i'}^W\\
&\leq \frac{Cp_n^2}{K}\sum_{i,i'}\sum_{\ell,\ell'}|P_{i\ell}^W P_{i'\ell}^W|\cdot P_{ii}^W P_{i'i'}^W = \frac{Cp_n^2}{K}(\sum_{\ell\in[n]}\sum_{i\in[n]}|P_{i\ell}^W P_{ii}^W|)^2\\
&\overset{(ii)}{\leq} \frac{Cp_n^2}{K}\left(\sum_{\ell\in[n]}(\sum_{i\in[n]}(P_{i\ell}^W)^2\cdot\sum_{i\in[n]}(P_{ii}^W)^2)\right)^2 \leq \frac{Cp_n^2}{K}(\sum_{\ell\in[n]}P_{\ell\ell}^W d_W)^2 = \frac{Cp_n^2}{K}d_W^4
\end{aligned}$$

where (i) and (ii) follows from Cauchy-Schwartz inequality. Hence $C_3 = O_p(\frac{p_n d_W^2}{K^{1/2}})$. In a similar manner, $C_4 = O_p(\frac{p_n d_W^2}{K^{1/2}})$, so that (A.4) follows. An application of (A.4) with $(U, \widetilde{U})$ replaced by $(\widetilde{e}, \vartheta)$ and (ε, ϑ) yields the first and second equation of (A.3) respectively.

Step 2: We show that

$$2Q_{\widetilde{e},P^W\widetilde{e}} - 2Q_{\vartheta,P^W\vartheta} = \mathcal{H}_n^{(1)} - \frac{2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}P_{ii}^W(\widetilde{e}_i\widetilde{e}_j - \vartheta_i\vartheta_j) = \mathcal{H}_n^{(1)} + O_p(\frac{p_n d_W^2}{K^{1/2}})$$
$$2Q_{\varepsilon,P^W\varepsilon} - 2Q_{\vartheta,P^W\vartheta} = \mathcal{H}_n^{(2)} - \frac{2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}P_{ii}^W(\varepsilon_i\varepsilon_j - \vartheta_i\vartheta_j) = \mathcal{H}_n^{(2)} + O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{A.5}$$

where $\mathcal{H}_n^{(\ell)} := -\frac{2}{\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i}P_{ii}P_{ij}^W\left\{\zeta_i^{(\ell)}\zeta_j^{(\ell)} - \vartheta_i\vartheta_j\right\}$ and $\zeta_i^{(\ell)} := \widetilde{e}_i$ or ε_i for $\ell = 1$ or 2 respectively.

We first derive a general result: consider a sequence of independent random vectors $\{(U_i, T_i)'\}_{i=1}^n$. Suppose we have another sequence of independent random vectors $\{(\widetilde{U}_i, \widetilde{T}_i)'\}_{i=1}^n$ such that for every $i \in [n]$, $\mathbb{E}(U_i, T_i) = \mathbb{E}(\widetilde{U}_i, \widetilde{T}_i)$ and $\mathbb{E}[(U_i, T_i)(U_i, T_i)'] = \mathbb{E}[(\widetilde{U}_i, \widetilde{T}_i)(\widetilde{U}_i, \widetilde{T}_i)']$. We assume the two sequences are independent from each other, and that the first two moments are bounded. By noting $P^W P = 0$,

$$Q_{P^W U,T} = \frac{1}{\sqrt{K}} U' P^W P T - \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} (P_i^W)' U \cdot T_i = -\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} (P_i^W)' U \cdot T_i$$
$$= -\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} \sum_{j\neq i} P_{ij}^W U_j T_i - \frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} P_{ii}^W U_i T_i,$$

which implies that

$$Q_{P^W U,T} - Q_{P^W \widetilde{U},\widetilde{T}} = -\frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W U_j T_i + \frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \widetilde{U}_j \widetilde{T}_i + O_p\Big(\frac{p_n d_W^2}{K^{1/2}}\Big), \quad (A.6)$$

where the last equality follows from Markov inequality and

$$\mathbb{E}\left(\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii} P_{ii}^W (U_i T_i - \widetilde{U}_i \widetilde{T}_i)\right)^2 = \frac{1}{K} \sum_{i\in[n]} P_{ii}^2 (P_{ii}^W)^2 \mathbb{E}(U_i T_i - \widetilde{U}_i \widetilde{T}_i)^2 \leq \frac{C p_n^2}{K} \sum_{i\in[n]} P_{ii}^W = \frac{C p_n^2 d_W}{K}.$$

If replace (U_i, T_i) with $(\widetilde{e}_i, \widetilde{e}_i)$, as well as $(\widetilde{U}_i, \widetilde{T}_i)$ with $(\vartheta_i, \vartheta_i)$, then an application of (A.6) would yield the first equation of (A.5). The second equation of (A.5) follows by replacing (U_i, T_i) with $(\varepsilon_i, \varepsilon_i)$ and $(\widetilde{U}_i, \widetilde{T}_i)$ with $(\vartheta_i, \vartheta_i)$.

Step 3: Define

$$F_n := Q_{\widetilde{e},\widetilde{e}} - \frac{2}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \widetilde{e}_i \widetilde{e}_j \quad \text{and}$$
$$\mathcal{F}_n := Q_{\varepsilon,\varepsilon} - \frac{2}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \varepsilon_i \varepsilon_j$$

We will show that there exists a random variable $\mathcal{F}_n' \overset{d}{=} \mathcal{F}_n$ such that

$$F_n = \mathcal{F}_n' + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2} (p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3}\right) \quad (A.7)$$

Define $g_n(x) := \max\left(0, 1 - \frac{d(x, A^{3\delta_n})}{\delta_n}\right)$ and $f_n(x) := \mathbb{E} g_n(x + h_n \mathcal{N})$, where $\mathcal{N}$ has a standard normal distribution and $h_n := \frac{3\delta_n}{C_h}$ for some $C_h > 1$. By Pollard (2001)[Theorem 10.18], $f_n(\cdot)$ is twice-continuously differentiable such that for all x, y,

$$\left| f_n(x+y) - f_n(x) - y \partial f_n(x) - \frac{1}{2} y^2 \partial^2 f_n(x) \right| \leq \frac{|y|^3}{9 \delta_n h_n^2} \quad (A.8)$$

and

$$1 - B(C_h)\mathbb{1}\{x \in A\} \le f_n(x) \le B(C_h) + (1 - B(C_h))\mathbb{1}\{x \in A^{3\delta_n}\}, \tag{A.9}$$

where $C_h := \frac{3\delta_n}{h_n}$ and $B(C_h) := \left(\frac{C_h^2}{exp(C_h^2 - 1)}\right)^{1/2}$. Furthermore, define

$$\mathcal{G}_n(a_1, ..., a_n) := \frac{\sum_{i\in[n]}\sum_{j\neq i}\{a_i P_{ij} a_j - 2P_{ii}P_{ij}^W a_i a_j)\}}{\sqrt{K}}$$

so $F_n = \mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n)$ and $\mathcal{F}_n = \mathcal{G}_n(\varepsilon_1, ..., \varepsilon_n)$. By triangle inequality,

$$\begin{aligned}
&|\mathbb{E}f_n(F_n) - \mathbb{E}f_n(\mathcal{F}_n)| \\
&\le \sum_{i\in[n]} |\mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_i, \varepsilon_{i+1}, ..., \varepsilon_n)) - \mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{i-1}, \varepsilon_i, ..., \varepsilon_n))|\,,
\end{aligned} \tag{A.10}$$

where $\mathcal{G}_n(\varepsilon_1, ..., \varepsilon_n, \widetilde{e}_{n+1}) \equiv \mathcal{G}_n(\varepsilon_1, ..., \varepsilon_n)$ and $\mathcal{G}_n(\varepsilon_0, \widetilde{e}_1, ..., \widetilde{e}_n) \equiv \mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n)$. Then consider the last term of the telescoping sum. Define

$$\begin{aligned}
\lambda_{n-1} &:= \frac{\sum_{i\in[n-1]}\sum_{j\neq i, j\in[n-1]}\{\widetilde{e}_i P_{ij}\widetilde{e}_j - 2P_{ii}P_{ij}^W \widetilde{e}_i\widetilde{e}_j\}}{\sqrt{K}} \\
\Delta_n &:= \frac{2\widetilde{e}_n \sum_{i\in[n-1]}\widetilde{e}_i P_{in}}{\sqrt{K}} - \frac{2\widetilde{e}_n\sum_{i\in[n-1]}P_{ii}P_{in}^W\widetilde{e}_i}{\sqrt{K}} - \frac{2P_{nn}\widetilde{e}_n\sum_{i\in[n-1]}P_{in}^W\widetilde{e}_i}{\sqrt{K}} \\
\widetilde{\Delta}_n &:= \frac{2\varepsilon_n \sum_{i\in[n-1]}\widetilde{e}_i P_{in}}{\sqrt{K}} - \frac{2\varepsilon_n\sum_{i\in[n-1]}P_{ii}P_{in}^W\widetilde{e}_i}{\sqrt{K}} - \frac{2P_{nn}\varepsilon_n\sum_{i\in[n-1]}P_{in}^W\widetilde{e}_i}{\sqrt{K}}
\end{aligned}$$

so that $\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n) = \Delta_n + \lambda_{n-1}$ and $\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{n-1}, \varepsilon_n) = \widetilde{\Delta}_n + \lambda_{n-1}$. Further denote $\mathcal{I}_{n-1}$ as the σ-field generated by $\{\varepsilon_i, \widetilde{e}_i\}_{i\in[n-1]}$ and observe that

$$\begin{aligned}
&\mathbb{E}(\Delta_n|\mathcal{I}_{n-1}) = \mathbb{E}(\widetilde{\Delta}_n|\mathcal{I}_{n-1}) \quad \text{and} \\
&\mathbb{E}(\Delta_n^2|\mathcal{I}_{n-1}) = \mathbb{E}(\widetilde{\Delta}_n^2|\mathcal{I}_{n-1}),
\end{aligned}$$

so that together with (A.8), letting $x = \lambda_{n-1}$, $y = \Delta_n$ and $\widetilde{\Delta}_n$, we have

$$\begin{aligned}
&|\mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n)) - \mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{n-1}, \varepsilon_n))| \\
&\le |\mathbb{E}\partial f_n(\lambda_{n-1})(\widetilde{\Delta}_n - \Delta_n)| + \frac{1}{2}|\mathbb{E}\partial^2 f_n(\lambda_{n-1})(\widetilde{\Delta}_n^2 - \Delta_n^2)| + \frac{\mathbb{E}|\widetilde{\Delta}_n|^3 + \mathbb{E}|\Delta_n|^3}{9\delta_n h_n^2} \\
&= \frac{\mathbb{E}|\Delta_n|^3 + \mathbb{E}|\widetilde{\Delta}_n|^3}{9\delta_n h_n^2}.
\end{aligned} \tag{A.11}$$

We proceed to bound $\mathbb{E}|\Delta_n|^3$. Let $\{\xi_i\}_{i\in[n-1]}$ be a sequence of independent Rademacher random variables. Using the simple inequality that $|a+b|^3 \le 2(a^2+b^2)\cdot|a+b| \le 8(|a|^3+|b|^3)$, we have by independence of the errors across i that

$$\mathbb{E}|\Delta_n|^3 \leq \frac{C}{K^{3/2}}\mathbb{E}\left|\sum_{i\in[n]}(P_{in}+P_{ii}P_{in}^W+P_{nn}P_{in}^W)\widetilde{e}_i\right|^3 \tag{A.12}$$

Denoting θ_i as either $P_{in}\widetilde{e}_i$, $P_{ii}P_{in}^W\widetilde{e}_i$ or $P_{nn}P_{in}^W\widetilde{e}_i$, we have

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i\in[n-1]}\theta_i\right|^3 \overset{(i)}{\leq} 8\mathbb{E}\left|\sum_{i\in[n-1]}\theta_i\xi_i\right|^3 \overset{(ii)}{\leq} 8\int_0^\infty t^2\mathbb{P}\left(\left|\sum_{i\in[n-1]}\theta_i\xi_i\right|>t\right)dt \\
&= 8\mathbb{E}\int_0^\infty t^2\mathbb{P}\left(\left|\sum_{i\in[n-1]}\theta_i\xi_i\right|>t\middle|\mathcal{I}_{n-1}\right)dt \overset{(iii)}{\leq} 16\mathbb{E}\int_0^\infty t^2 exp(-\frac{1}{2}\frac{t^2}{\sum_{i\in[n-1]}\theta_i^2})dt \\
&\overset{(iv)}{\leq} C\mathbb{E}\left(\sum_{i\in[n-1]}\theta_i^2\right)^{3/2} \overset{(v)}{\leq} C\left(\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2\right)^{3/4}
\end{aligned} \tag{A.13}$$

where (i) follows from the Symmetrization Lemma of Van der Vaart and Wellner (1996)[Lemma 2.3.1]; (ii) follows from the integral identity; (iii) follows from Hoeffding's inequality (see Van der Vaart and Wellner (1996)[Lemma 2.2.7]); (iv) follows from the change of variable $s = t^2/\sum_{i\in[n-1]}\theta_i^2$; (v) follows from Holder's inequality. Note that for $\theta_i = P_{in}\widetilde{e}_i$,

$$\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2 = \sum_{i\in[n-1]}\sum_{j\in[n-1]}\mathbb{E}\theta_i^2\theta_j^2 \leq C\sum_{i\in[n]}\sum_{j\in[n]}P_{in}^2P_{jn}^2 = CP_{nn}^2,$$

so that

$$\left(\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2\right)^{3/4} \leq CP_{nn}^{3/2}$$

Similarly we can obtain

$$\begin{aligned}
&\left(\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2\right)^{3/4} \leq C(p_nP_{nn}^W)^{3/2} \quad \text{if } \theta_i = P_{ii}P_{in}^W\widetilde{e}_i \quad \text{and} \\
&\left(\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2\right)^{3/4} \leq C(P_{nn}P_{nn}^W)^{3/2} \quad \text{if } \theta_i = P_{nn}P_{in}^W\widetilde{e}_i
\end{aligned}$$

Hence, by (A.12) and (A.13), we have

$$\mathbb{E}|\widetilde{\Delta}_n|^3 \leq C\frac{P_{nn}^{3/2}+p_n^{3/2}(P_{nn}^W)^{3/2}+(P_{nn}P_{nn}^W)^{3/2}}{K^{3/2}}.$$

Similarly, we have

$$\mathbb{E}|\Delta_n|^3 \le C\frac{P_{nn}^{3/2} + p_n^{3/2}(P_{nn}^W)^{3/2} + (P_{nn}P_{nn}^W)^{3/2}}{K^{3/2}}.$$

In general, for any generic jth term, we can show that

$$|\mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n)) - \mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{n-1}, \varepsilon_n))| \le C\frac{P_{jj}^{3/2} + p_n^{3/2}(P_{jj}^W)^{3/2} + (P_{jj}P_{jj}^W)^{3/2}}{K^{3/2}\delta_n h_n^2}$$

where the constant C is independent of n. By (A.10), letting $h_n := \left[\frac{C_h(p_n^{1/2}+p_n^{3/2}(p_n^W)^{1/2}d_W)}{K^{1/2}}\right]^{1/3}$ and recalling $\delta_n = \frac{C_h h_n}{3}$, we have

$$|\mathbb{E}f_n(F_n) - \mathbb{E}f_n(\mathcal{F}_n)| \le C\frac{\sum_{i\in[n]} P_{ii}^{3/2} + p_n^{3/2}(P_{ii}^W)^{3/2}}{K^{3/2}\delta_n h_n^2} \le C\frac{p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2}d_W}{K^{1/2}\delta_n h_n^2} \le \frac{C}{C_h^2}.$$

Therefore, by (A.9) we have

$$\begin{aligned}
\mathbb{P}\{F_n \in A\} &\le \frac{\mathbb{E}f_n(F_n)}{1-B(C_h)} \le \frac{1}{1-B(C_h)}\left(\mathbb{E}f_n(\mathcal{F}_n) + \frac{C}{C_h^2}\right) \\
&\le \frac{1}{1-B(C_h)}\left(B(C_h) + (1-B(C_h))\mathbb{P}\left\{\mathcal{F}_n \in A^{3\delta_n}\right\} + \frac{C}{C_h^2}\right) \\
&= \mathbb{P}\left\{\mathcal{F}_n \in A^{3\delta_n}\right\} + \frac{B(C_h) + \frac{C}{C_h^2}}{1-B(C_h)}
\end{aligned}$$

By Strassen's Theorem (see Pollard (2001)[Theorem 10.8]),there exists a random variable $\mathcal{F}_n' \overset{d}{=} \mathcal{F}_n$ such that

$$\mathbb{P}\left\{|F_n - \mathcal{F}_n'| > C_h\left[\frac{C_h(p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2}d_W)}{K^{1/2}}\right]^{1/3}\right\} \le \frac{B(C_h) + \frac{C}{C_h^2}}{1-B(C_h)}$$

Fix any $\tau > 0$. Given that $B(C_h) \to 0$ whenever $C_h \to \infty$, we can find a sufficiently large C_h such that $\frac{B(C_h)+\frac{C}{C_h^2}}{1-B(C_h)} \le \tau$, implying

$$|F_n - \mathcal{F}_n'| = O_p\left(\left[\frac{(p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2}d_W)}{K^{1/2}}\right]^{1/3}\right),$$

so (A.7) is shown.

Step 4: We complete the proof. We can re-express

$$Q_{e,e} = F_n + R_n$$

and

$$Q_{\mathcal{E},\mathcal{E}} = \mathcal{F}_n + \mathcal{R}_n$$

where $F_n, \mathcal{F}_n$ were defined in **Step 3**, so clearly $R_n = Q_{e,e} - F_n$; similarly $\mathcal{R}_n = Q_{\mathcal{E},\mathcal{E}} - \mathcal{F}_n$. Define

$$\widetilde{\mathcal{R}}_n := -\frac{2}{\sqrt{K}} \sum_{i\in[n]} P_{ii} P^W_{ij} \vartheta_i \vartheta_j + Q_{P^W\vartheta, P^W\vartheta}$$

and note that by (A.3) and (A.5),

$$R_n - \widetilde{\mathcal{R}}_n = O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{A.14}$$

and

$$\mathcal{R}_n - \widetilde{\mathcal{R}}_n = O_p(\frac{p_n d_W^2}{K^{1/2}}). \tag{A.15}$$

Therefore, by noting that $F_n, \mathcal{F}_n, \widetilde{\mathcal{R}}_n$ are mutually independent, we have

$$\begin{aligned}
Q_{e,e} &= F_n + R_n = \mathcal{F}'_n + (F_n - \mathcal{F}'_n) + (R_n - \widetilde{\mathcal{R}}_n) + \widetilde{\mathcal{R}}_n \\
&= \mathcal{F}'_n + \widetilde{\mathcal{R}}_n + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3} + \frac{p_n d_W^2}{K^{1/2}}\right) \\
&\overset{d}{=} \mathcal{F}_n + \widetilde{\mathcal{R}}_n + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3} + \frac{p_n d_W^2}{K^{1/2}}\right) \\
&= \mathcal{F}_n + \mathcal{R}_n - (\mathcal{R}_n - \widetilde{\mathcal{R}}_n) + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3} + \frac{p_n d_W^2}{K^{1/2}}\right) \\
&= Q_{\mathcal{E},\mathcal{E}} + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2}(p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3} + \frac{p_n d_W^2}{K^{1/2}}\right).
\end{aligned}$$

where the second line of the preceding equation follows from (A.7) and (A.14); the last line follows from (A.15). This gives the first result of Theorem 1.

Step 5: We prove the second part of the Theorem here. Note that by $P^W P = 0$,

$$\frac{e'Pe}{K} = \frac{\widetilde{e}'P\widetilde{e}}{K} = \frac{1}{\sqrt{K}} Q_{\widetilde{e},\widetilde{e}} + \frac{\sum_{i\in[n]} P_{ii}\widetilde{e}_i^2}{K},$$

and similarly

$$\frac{\mathcal{E}'P\mathcal{E}}{K} = \frac{1}{\sqrt{K}} Q_{\varepsilon,\varepsilon} + \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{K}.$$

Then

$$\frac{\sum_{i\in[n]} P_{ii}\widetilde{e}_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} = O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right)$$
$$\frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} = O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right) \tag{A.16}$$

which follows from

$$\mathbb{E}\left(\frac{\sum_{i\in[n]} P_{ii}(\widetilde{e}_i^2 - \vartheta_i^2)}{K}\right)^2 = \frac{\sum_{i\in[n]} P_{ii}^2\mathbb{E}(\widetilde{e}_i^2 - \vartheta_i^2)^2}{K^2} \leq \frac{Cp_n \sum_{i\in[n]} P_{ii}}{K^2} = \frac{Cp_n}{K}$$

Then define $J_n := \frac{Q_{\widetilde{e},\widetilde{e}}}{\sqrt{K}}$ and $\mathcal{J}_n := \frac{Q_{\varepsilon,\varepsilon}}{\sqrt{K}}$. By repeating the proof of **step 3**, we can show that there exists a random variable $\mathcal{J}_n' \overset{d}{=} \mathcal{J}_n$ such that

$$J_n = \mathcal{J}_n' + O_p(\frac{p_n^{1/2}}{K}). \tag{A.17}$$

Putting everything together, we have

$$\begin{aligned}
\frac{e'Pe}{K} &= J_n + \left(\frac{\sum_{i\in[n]} P_{ii}\widetilde{e}_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K}\right) + \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} \\
&\overset{(i)}{=} \mathcal{J}_n' + \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} + O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right) \\
&\overset{d}{=} \mathcal{J}_n + \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} + O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right) \\
&= \frac{\mathcal{E}'P\mathcal{E}}{K} - \left(\frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{K}\right) + O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right) \\
&= \frac{\mathcal{E}'P\mathcal{E}}{K} + O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right)
\end{aligned}$$

where (i) follows from (A.16) and (A.17). This completes the proof of the second part of Theorem 1.

A.2 Proof of Theorem 2

Consider any sub-sequence $\lambda_{n_k} \in \Lambda_{n_k}$. We will show that for both fixed and diverging K,

$$\lim_{n_k\to\infty} \mathbb{P}_{\lambda_{n_k}}\left(\widehat{Q}(\beta_0) > q_{1-\alpha}(F_{\widetilde{w}})\right) = \alpha. \tag{A.18}$$
$$\lim_{n_k\to\infty} \mathbb{P}_{\lambda_{n_k}}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_\alpha^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) = \alpha \tag{A.19}$$

Then (A.18) and (A.19) satisfy **Assumption B*** of Andrews, Cheng, and Guggenberger (2020). By **Corollary 2.1(c)** of their paper, Theorem 2 follows. Without loss of generality, we implicitly consider the sequence $\lambda_n \in \Lambda_n$ and show that it satisfies (A.18) and (A.19). We break the proof into two parts, part I and II, which deals with (A.18) and (A.19) respectively. For each part, we deal with fixed and diverging instruments separately. We drop the dependence on β_0 for notational simplicity.

Part I:

Fixed K case: Consider first when K is fixed. We can write the rejection criteria (2.7) as

$$\widehat{Q}(\beta_0) > q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}} - 1\right) \tag{A.20}$$

We denote $Q(\beta_0)$ as $Q_n(\beta_0)$ to reflect its relationship to the sample size n. Under the null, by Theorem D.2.1 and Lemma B.3, we know that for any sub-sequence n_j, there exists a further sub-sequence n_{j_k} such that

$$\widehat{Q}_{n_{j_k}}(\beta_0) \rightsquigarrow \sum_{i\in[K]} w_i^* \chi_{1,i}^2 =: \overline{\chi}_{w^*}^2 \tag{A.21}$$

where the chi-squares are independent with one degree of freedom. Furthermore, $F_{\widetilde{w}_{n_{j_k}}} \rightsquigarrow \overline{\chi}_{w^*}^2$ since $\widetilde{w}_{n_{j_k}} \xrightarrow{p} w^*$ by Lemma B.3. By arguing along sub-sequences, we can assume without loss of generality that the above convergence is in terms of a full sequence, i.e. $\widetilde{w}_n \xrightarrow{p} w^*$ and $w_n \to w^*$. Furthermore, note that

$$(a)\quad ||w_n||_F^2 \cdot (\sum_{i\in[n]} P_{ii}\sigma_i^2)^2 = trace(U'\Lambda U U'\Lambda U) = \sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2 \sigma_i^2 \sigma_j^2$$

$$(b)\quad \sum_{i\in[n]} P_{ii}^2 \sigma_i^2 \le \overline{C} p_n K = o(1)$$

$$(c)\quad \widehat{\Phi}_1 \overset{(i)}{=} \Phi_1 + o_p(1) \overset{(ii)}{=} \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \widetilde{\sigma}_i^2 \widetilde{\sigma}_j^2 + o_p(1) \overset{(iii)}{=} \frac{2}{K}\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2 \sigma_i^2\sigma_j^2 + o_p(1)$$

$$(d)\quad \frac{1}{K}\sum_{i\in[n]} P_{ii} e_i^2 \overset{(iv)}{=} \frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2 + o_p(1)$$

where (i) follows from our assumption of consistent estimator; (ii) from the second part of Theorem C.0.1; (iii) follows from (b); (iv) follows from Lemma B.1. Then from (d) we have

$$(e)\quad \frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\sigma_i^2}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2} = \frac{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2}{\frac{1}{K}\sum_{i\in[n]} P_{ii} e_i^2} = \frac{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2}{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2 + o_p(1)} \xrightarrow{p} 1,$$

and from (c) we have

$$(f)\quad \frac{\sqrt{\widehat{\Phi}_1}}{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}} = \sqrt{\frac{\frac{2}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2 + o_p(1)}{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}} = \sqrt{2} + o_p(1)$$

Putting it together,

$$\begin{aligned}
\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2} &= \frac{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}\cdot\frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}\cdot\frac{\sqrt{\widehat{\Phi}_1}}{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}} \\
&\overset{(e),(f)}{=} \frac{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}(1+o_p(1))(\sqrt{2}+o_p(1)) = \sqrt{2}\frac{\sqrt{\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\sum_{i\in[n]}P_{ii}\sigma_i^2} + o_p(1) \\
&\overset{(a)}{=} \sqrt{2}||w_n|| + o_p(1) = \sqrt{2}||w^*|| + o_p(1),
\end{aligned}$$

so that since $\widetilde{w}_n \overset{p}{\to} w^*$ and $w_n \to w^*$,

$$\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} \overset{p}{\to} \frac{\sqrt{2}||w^*||}{\sqrt{2}||w^*||} = 1$$

Therefore,

$$(q_{1-\alpha}(F_{\widetilde{w}})-1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right) = (q_{1-\alpha}(F_{w^*}) - 1 + o_p(1))o_p(1) = o_p(1),$$

so we can write (A.20) as

$$q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n})-1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right) \rightsquigarrow q_{1-\alpha}(\overline{\chi}^2_{w^*})$$

By Van der Vaart and Wellner (1996)[Example 1.4.7],

$$\left(\widehat{Q}(\beta_0), q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n})-1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right)\right) \rightsquigarrow (\overline{\chi}^2_{w^*}, q_{1-\alpha}(\overline{\chi}^2_{w^*})),$$

from which an application of Theorem 1.3.6 from the same reference yields

$$\widehat{Q}(\beta_0) - q_{1-\alpha}(F_{\widetilde{w}_n}) - (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right) \rightsquigarrow \overline{\chi}^2_{w^*} - q_{1-\alpha}(\overline{\chi}^2_{w^*});$$

applying Theorem 1.3.4(vi) of the same reference yields

$$\lim_{n\to\infty}\mathbb{P}_{\lambda_n}\left(\widehat{Q}(\beta_0) - q_{1-\alpha}(F_{\widetilde{w}_n}) - (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right) > 0\right)$$
$$= \mathbb{P}\left(\overline{\chi}^2_{w^*} > q_{1-\alpha}(\overline{\chi}^2_{w^*})\right) = \alpha$$

We have therefore shown that for fixed K, (A.18) is satisfied.

Diverging K: assume now that $K \to \infty$. By Theorem D.1.2 we have

$$\frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2}{\sqrt{\widehat{\Phi}_1}}\left(\widehat{Q}(\beta_0) - 1\right) \rightsquigarrow \mathcal{N}(0,1) \tag{A.22}$$

Next, define $\mathcal{I} := \sigma\left(\{\widetilde{w}_{i,n}\}_{i=1}^n\right)_{n\geq 1}$ to be the sigma-field generated by the sequence of random variables $\widetilde{w}_{i,n}$ and $s_n^2 := 2\sum_{i\in[K]}\widetilde{w}_{i,n}^2$. Conditioning on $\mathcal{I}$, we have

$$Var(F_{\widetilde{w}_n} - 1 \mid \mathcal{I}) = \mathbb{E}\left(\sum_{i\in[K]}\widetilde{w}_{i,n}(\chi^2_{1,i} - 1)\right) = s_n^2. \tag{A.23}$$

Additionally, we have

$$\lim_{K\to\infty}\frac{C\max_i \widetilde{w}_{i,n}^2}{\sum_{i\in[n]}\widetilde{w}_{i,n}^2} = 0. \tag{A.24}$$

To see (A.24), note that $\max_i \widetilde{w}_{i,n} = o_p(1)$ by Lemma B.3. Furthermore, $\sum_{i\in[K]}\widetilde{w}_{i,n} = 1$ by construction. Let $\max_i \widetilde{w}_{i,n} = \theta_0$ for some $0 < \theta_0 < 1$. Denote i^* to be the index such that $\widetilde{w}_{i^*,n} = \max_i \widetilde{w}_{i,n}$. As $\sum_{i\neq i^*}\widetilde{w}_{i,n} = 1 - \theta_0$, we have

$$\sum_{i\in[n]}\widetilde{w}_{i,n}^2 = \sum_{i\neq i^*}\widetilde{w}_{i,n}^2 + \widetilde{w}_{i^*,n}^2 = \sum_{i\neq i^*}\widetilde{w}_{i,n}^2 + \theta_0^2 \geq \sum_{i\neq i^*}(\frac{1-\theta_0}{K-1})^2 + \theta_0^2 = \frac{(1-\theta_0)^2}{K-1} + \theta_0^2,$$

so that

$$\frac{\max_i \widetilde{w}_{i,n}^2}{\sum_{i\in[n]}\widetilde{w}_{i,n}^2} = \frac{\theta_0^2}{\sum_{i\in[n]}\widetilde{w}_{i,n}^2} \leq \frac{\theta_0^2}{\theta_0^2 + \frac{(1-\theta_0)^2}{K-1}} = \frac{1}{1 + \frac{(1-\theta_0)^2}{\theta_0^2(K-1)}} = o(1),$$

where the last equality follows from recalling Lemma B.3, i.e. $\theta_0^2 = \max_i \widetilde{w}_{i,n}^2 = o_p(K^{-1})$, so that

$$\frac{(1-\theta_0)^2}{\theta_0^2(K-1)} = \frac{1+o(1)}{\theta_0^2(K-1)} = \frac{1+o(1)}{o(1)} \to \infty$$

Thus, by (A.24) we can obtain

$$\begin{aligned}
\lim_{K\to\infty} \frac{1}{s_n^4} \sum_{i\in[K]} \mathbb{E}(\widetilde{w}_{i,n}(\chi_{1,i}^2 - 1))^4 &\leq \lim_{K\to\infty} \frac{C\sum_{i\in[n]} \widetilde{w}_{i,n}^4}{s_n^4} \leq \lim_{K\to\infty} \frac{C\max_i \widetilde{w}_{i,n}^2 \sum_{i\in[n]} \widetilde{w}_{i,n}^2}{(\sum_{i\in[K]} \widetilde{w}_{i,n}^2)^2} \\
&= \lim_{K\to\infty} \frac{C\max_i \widetilde{w}_{i,n}^2}{\sum_{i\in[K]} \widetilde{w}_{i,n}^2} = 0. \qquad (A.25)
\end{aligned}$$

Since the Lyapunov condition (A.23) and (A.25) is satisfied, by the Lyapunov Central Limit Theorem, conditional on $\mathcal{I}$ we have

$$\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}} \rightsquigarrow \mathcal{N}(0,1). \qquad (A.26)$$

Since the distributional convergence in (A.26) holds for any sequence $\widetilde{w}_{i,n}$, then it must hold unconditionally by Lemma B.4. Hence, asymptotically, by (A.22) we have exact α-level size control whenever

$$\frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2}{\sqrt{\widehat{\Phi}_1}} \left(\widehat{Q}(\beta_0) - 1\right) > q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}}\right)$$

We can rearrange this rejection criteria as

$$\widehat{Q}(\beta_0) > 1 + \frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2} \cdot q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}}\right) \equiv C_\alpha$$

implying that we have exact asymptotic size control for $K \to \infty$. By an application of Van der Vaart and Wellner (1996)[Example 1.4.7, Theorem 1.3.6, Theorem 1.3.4(vi)], as was done previously for the fixed K case, we have (A.18). The proof of part I is complete.

Part II:

Fixed K case: Consider first when K is fixed. As in part I, we assume without loss of generality that $\widetilde{w}_n \xrightarrow{p} w^*$ and $w_n \to w^*$ instead of over a sub-sequence. Note that

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) = \frac{\sum_{i\in[n]} P_{ii}e_i^2(\widehat{Q}_s(\beta_0) - 1)}{\sqrt{K\widehat{\Phi}_1}} = \frac{\widehat{Q}(\beta_0) - 1}{\sqrt{2}||w^*||} + o_p(1) \rightsquigarrow \sum_{i\in[K]} \frac{w_i^*}{\sqrt{2}||w^*||}(\chi_{1,i}^2 - 1) \quad (A.27)$$

where the last equality follows from recalling from Part I that

$$\frac{\sqrt{K}\widehat{\Phi}_1}{\sum_{i\in[n]} P_{ii}e_i^2} = \sqrt{2}||w^*|| + o_p(1)$$

for the fixed K case; the weak convergence follows from (A.21). Next, we will show that for any fixed $\ell \in [B]$, conditioning on data,

$$\widehat{J}^{BS,\ell} \rightsquigarrow \sum_{i\in[K]} \frac{w_i^{BS}}{\sqrt{2}||w^{BS}||}(\chi^2_{1,i} - 1) \tag{A.28}$$

where we drop the dependence of $\widehat{J}^{BS,\ell}$ on $(e(\beta_0), \mathcal{L}, \widehat{\Phi}_1(\beta_0))$ for notational simplicity, and w_i^{BS} are the eigenvalues of $\frac{(Z'\Lambda^{BS}Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda^{BS}Z)^{1/2}}{\sum_{i\in[n]} P_{ii}e_i^2}$. First observe that $\widehat{\Phi}_1^{BS,\ell} \xrightarrow{\hat{p}} \Phi_1^{BS,\ell} := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2 e_j^2$ by $Var(\eta_i) = e_i^2$ and the assumption of $\widehat{\Phi}_1(\beta_0)$ satisfying (2.9), where $\xrightarrow{\hat{p}}$ means convergence in probability conditioning on the data. Repeating the proof of (A.27) applied to

A.3 Proof of Theorem 3

Note that (A.26) holds for any sequence of $\Delta_n \to \overline{\Delta}$ not necessarily zero, i.e.

$$\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}} \rightsquigarrow \mathcal{N}(0,1) \tag{A.29}$$

Furthermore, our rejection criteria for the test under diverging K can be rewritten as

$$\frac{1}{\sqrt{K}} \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) > \sqrt{\widehat{\Phi}_1(\beta_0)} \cdot q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2}}\right) \tag{A.30}$$

By (2.9), noting that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_i^2(\beta_0)\sigma_j^2(\beta_0) \le \frac{C}{K}\sum_{i,j\in[n]} P_{ij}^2 = C = O(1),$$

the estimator $\widehat{\Phi}_1(\beta_0) = O_p(1)$. Therefore the right-hand-side of (A.30) is an $O_p(1)$ term. By Theorem D.1.2, the left-hand-side (A.30) diverges to infinity as $\overline{\mathcal{C}} \to \infty$ and $\Delta \neq 0$ is fixed. The result thus follow.

A.4 Proof of Theorem 4

By Theorem D.1.2,

$$\frac{1}{\sqrt{K\Phi_1(\beta_0)}} \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)(\widehat{Q}(\beta_0) - 1) \rightsquigarrow \mathcal{N}\left(\frac{\Delta^2\overline{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}, 1\right)$$

Therefore, by (A.29), for fixed Δ and any estimator $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$.

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \sum_{i\in[n]} P_{ii} e_i^2(\beta_0)(\widehat{Q}(\beta_0)-1) > q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n}-1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}}\right)\right) \\
&= 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\Delta^2\overline{\mathcal{C}}}{\sqrt{\widehat{\Phi}_1(\beta_0)}}\right) \\
&= 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\Delta^2\overline{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}\right)
\end{aligned}$$

Noting that $\Delta = \widetilde{\Delta}$ and $\overline{\mathcal{C}} = \widetilde{\mathcal{C}}$ completes the first part of the proof. Next, we show that

$$\widehat{\Phi}_1^{standard}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0), \tag{A.31}$$

$$\widehat{\Phi}_1^{cf}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0). \tag{A.32}$$

in order to complete the second part of the proof. Recall from section 2.4 that

$$\mathcal{D}^{standard}(\Delta) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(2\Delta^2\Pi_j^2\sigma_i^2(\beta_0) + \Delta^4\Pi_i^2\Pi_j^2) \to 0$$

by the assumption that $\frac{\Pi'\Pi}{K} \to 0$, $\sigma_i^2(\beta_0) < C$ and $\sum_{j\in[n]} P_{ij}^2 = P_{ii} \leq 1$. By (2.9) we have (A.31). Furthermore, by $\Pi' M\Pi \leq \frac{\Pi'\Pi}{K} \to 0$, (A.32) follows from Mikusheva and Sun (2022)[Theorem 3].

A.5 Proof of Theorem 5

Note that $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$ by (2.9) and $\Delta \to 0$. Furthermore, $\frac{\Delta^2\overline{\mathcal{C}}}{\sqrt{\widehat{\Phi}_1(\beta_0)}} = \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}} + o(1) = \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}$, so that by Theorem D.1.2 we have

$$\frac{1}{\sqrt{K\Phi_1(\beta_0)}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)(\widehat{Q}(\beta_0)-1) \rightsquigarrow \mathcal{N}\left(\frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\Phi^{1/2}(\beta_0)}, 1\right)$$

Finally, by (A.29) we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0))\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \sum_{i\in[n]} P_{ii} e_i^2(\beta_0)(\widehat{Q}(\beta_0)-1) > q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n}-1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}}\right)\right) \\
&= 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\Phi^{1/2}(\beta_0)}\right)
\end{aligned}$$

A.6 Proof of Lemma 4.1

The proof is similar to the proof of Theorem 2. For completeness we will include the proof here. Note that

$$(a)\quad ||w_n||_F^2 \cdot (\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2 = \sum_{i,j\in[n]} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$$

$$(b)\quad \sum_{i\in[n]} P_{ii}^2\sigma_i^2(\beta_0) \le Cp_nK = o(1)$$

$$(c)\quad \widehat{\Phi}_1(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + \mathcal{D}(\Delta) \text{ by assumption of (2.9)}$$

Hence

$$\begin{aligned}\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} &\overset{(i)}{=} \frac{\sqrt{\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + O_p(1)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + O_p(1)} + o_p(1) \\ &\overset{(a),(b)}{=} \sqrt{2}||w_n||_F + O_p(1) \le \sqrt{2}||D_n + \Lambda_H||_F + \sqrt{2}||\Lambda_H||_F + O_p(1) \\ &\overset{(ii)}{=} \sqrt{2}||D_n + \Lambda_H||_F + O_p(1)\end{aligned}$$

where (i) follows from (c) and Lemma B.1; Λ_H is defined in Lemma B.3; (ii) follows from $||\Lambda_H||_F^2 = ||\Omega_H(\beta_0)||_F^2 = \frac{\Delta^4\sum_{i,j\in[n]} P_{ij}^2\Pi_i^2\Pi_j^2}{\sum_{i\in[K]} P_{ii}\sigma_i^2(\beta_0)} \le \frac{\Delta^4 CK}{\underline{C}K} \le C$. Furthermore, we have by Lemma B.3

$$||D_{\widetilde{w}_n} - D_n - \Lambda_H||_F = o_p(1)$$

where $D_{\widetilde{w}_n} := diag(\widetilde{w}_{1,n}, ..., \widetilde{w}_{K,n})$, so that

$$||\widetilde{w}_n||_F = ||(D_{\widetilde{w}_n} - D_n - \Lambda_H) + \Lambda_H + D_n||_F = ||\Lambda_H + D_n||_F + o_p(1)$$

Putting it together we have

$$\begin{aligned}\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} &= \frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2}||\widetilde{w}_n||_F} \le \frac{\sqrt{2}||D_n + \Lambda_H||_F + O_p(1)}{\sqrt{2}||\widetilde{w}_n||_F} \\ &= \frac{\sqrt{2}||D_n + \Lambda_H||_F + O_p(1)}{\sqrt{2}||\Lambda_H + D_n||_F + o_p(1)} \xrightarrow{p} 1 + O_p(1) = O_p(1)\end{aligned}$$

which completes the proof.

A.7 Proof of Lemma 4.2

We require a Theorem by Fleiss (1971):

Theorem 9. *(Fleiss (1971)) Let $\{\chi^2_{n_i,i}\}_{i=1}^K$ be a sequence of mutually independent chi-squares with*

n_i-degrees of freedom. Define

$$T_i := \frac{\chi^2_{n_i,i}}{\sum_{i=1}^K \chi^2_{n_i,i}}$$

to be the ratio of chi-squares. Then for any non-negative constants $a_1, .., a_K$, conditional on $\{T_i\}_{i=1}^K$,

$$\sum_{i\in[p]} a_i \chi^2_{n_i,i} \overset{d}{=} c_1 \cdot \chi^2_{\sum_{i\in[K]} n_i}$$

where $c_1 := \sum_{i\in[K]} a_i T_i$

We denote $\mathcal{F}_\ell := \left\{w \in \Omega : T_\ell = min_{\ell\in[K]} T_\ell\right\}$ for every $\ell \in [K]$; furthermore $\mathbb{P}(\bigcup_{\ell\in[K]} \mathcal{F}_\ell) = 1$ and $\mathbb{P}(\bigcap_{\ell\in[K]} \mathcal{F}_\ell) = 0$. Then for any chosen non-negative $(a_1, ..., a_K)$ such that $\sum_{\ell\in[K]} a_\ell = 1$ and for any $x \in \mathbb{R}_+$, we have

$$\begin{aligned}
&\mathbb{P}\left(\chi^2_{1,1} \leq x \cap \mathcal{F}_1 | \{T_\ell\}_{\ell\in[K]}\right) = \mathbb{E}\left(\mathbb{1}_{\chi^2_{1,1}\leq x} \mathbb{1}_{\mathcal{F}_1} | \{T_\ell\}_{\ell\in[K]}\right) = \mathbb{1}_{\mathcal{F}_1} \mathbb{P}\left(\chi^2_{1,1} \leq x | \{T_\ell\}_{\ell\in[K]}\right) \\
&\overset{(i)}{=} \mathbb{1}_{\mathcal{F}_1} \mathbb{P}\left(T_1 \chi^2_K \leq x\right) \overset{(ii)}{\leq} \mathbb{1}_{\mathcal{F}_1} \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell T_\ell \cdot \chi^2_K \leq x\right) \\
&\overset{(iii)}{=} \mathbb{1}_{\mathcal{F}_1} \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell} \leq x | \{T_\ell\}_{\ell\in[K]}\right) = \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell} \leq x \cap \mathcal{F}_1 | \{T_\ell\}_{\ell\in[K]}\right)
\end{aligned}$$

where (i) and (iii) follows from Theorem 9; (ii) follows from the fact that whenever $\omega \in \mathcal{F}_1$, $T_1 \leq \sum_{\ell\in[K]} a_\ell T_\ell$ since $\sum_{\ell\in[K]} a_\ell = 1$. Taking expectation on both sides of the equation yield

$$\mathbb{P}\left(\chi^2_{1,1} \leq x \cap \mathcal{F}_1\right) \leq \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell} \leq x \cap \mathcal{F}_1\right).$$

Note that $\{\mathcal{F}_\ell\}_{\ell\in[K]}$ are mutually disjoint except on a null set. Therefore

$$\mathbb{P}(\chi^2_{1,1} \leq x) \overset{(iii)}{\leq} \sum_{i\in[K]} \mathbb{P}\left(\chi^2_{1,i} \leq x \cap \mathcal{F}_i\right) \leq \sum_{i\in[K]} \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell} \leq x \cap \mathcal{F}_i\right) = \mathbb{P}\left(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell} \leq x\right)$$

where (iii) follows from $\mathbb{1}_{\mathcal{F}_i} \chi^2_{1,i} \leq \mathbb{1}_{\mathcal{F}_i} \chi^2_{1,1}$ and

$$\mathbb{P}(\chi^2_{1,1} \leq x) = \sum_{i\in[K]} \mathbb{P}\left(\chi^2_{1,1} \leq x \cap \mathcal{F}_i\right) \leq \sum_{i\in[K]} \mathbb{P}\left(\chi^2_{1,i} \leq x \cap \mathcal{F}_i\right).$$

Hence we can conclude that the distribution function of a chi-square is smaller than that of a weighted-chi-square. This implies that

$$q_{1-\alpha}(\chi^2_1) \geq q_{1-\alpha}(\sum_{\ell\in[K]} a_\ell \chi^2_{1,\ell})$$

A.8 Proof of Theorem 6

We begin by establishing some results: we will show later on that for any sequence of $\Delta_n \to \Delta^\dagger$ with $\Delta^\dagger$ finite,

$$n^{-1/2}((Z'\widetilde{e})', (Z'\Delta_n\widetilde{v})')' \rightsquigarrow \mathcal{N}\left(0, \Sigma(\Delta^\dagger)\right) \tag{A.33}$$

where $\Sigma(\Delta^\dagger) := \lim_{n\to\infty} \frac{1}{n}\sum_{i\in[n]} \Lambda_{0,i}(\Delta_n) \otimes Z_i Z_i'$. Furthermore, $\beta_0 := \beta_{0,n}$ (since Δ_n is allowed to change) so that β_0 is allowed to change with n; however we drop the notational dependence on n and understand that this implicitly holds. Then we can obtain

$$\begin{aligned}
&e(\beta_0)'Pe(\beta_0)\\
&= (n^{-1/2}Z'\widetilde{e} + \Delta_n n^{-1/2}Z'\widetilde{v} + \Delta_n n^{-1/2}Z'\Pi)' \left(\frac{Z'Z}{n}\right)^{-1} (n^{-1/2}Z'\widetilde{e} + \Delta_n n^{-1/2}Z'\widetilde{v} + \Delta_n n^{-1/2}Z'\Pi)\\
&\rightsquigarrow ((I_K, I_K)\mathcal{N}(0,\Sigma(\Delta^\dagger)) + \Delta^\dagger\mu_K)'Q_{ZZ}^{-1}((I_K, I_K)\mathcal{N}(0,\Sigma(\Delta^\dagger)) + \Delta^\dagger\mu_K) \qquad \text{(A.34)}
\end{aligned}$$

Furthermore, note that

$$\left(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\right)^{-1} \geq C(1 + \Delta^\dagger + \Delta^{\dagger 2})^{-1} + o_p(1) \tag{A.35}$$

for some $C > 0$. To see (A.35), first denote $\sigma_i^2(\Delta^\dagger) := \sigma_i^2(\widetilde{\beta}_0)$, where $\Delta^\dagger = \beta - \widetilde{\beta}_0$. Then observe that

$$\begin{aligned}
\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)) &\overset{(i)}{=} \frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \frac{\Delta_n^2}{K}\sum_{i\in[n]} P_{ii}\Pi_i^2 + o_p(1+\Delta_n)\\
&\overset{(ii)}{\leq} \frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \Delta_n^2 \max_i \Pi_i^2 + o_p(1+\Delta_n)\\
&\overset{(iii)}{\leq} C(1+\Delta_n) + C\Delta_n^2 + o_p(1+\Delta_n)\\
&\leq C(1+\Delta_n+\Delta_n^2) + o_p(1+\Delta_n)\\
&\overset{(iv)}{=} C(1+\Delta^\dagger + \Delta^{\dagger 2}) + o_p(1)
\end{aligned}$$

where (i) follows from Lemma B.1; (ii) follows from $\sum_{i\in[n]} P_{ii} = K$; (iii) follows from $\max_i \sigma_i^2(\beta_0) \leq \max_i(\widetilde{\sigma}_i^2 + \Delta_n^2\widetilde{\zeta}_i^2 + 2\Delta_n\widetilde{\gamma}_i) \leq C(1+\Delta_n)$ and $\max_i \Pi_i^2 \leq \Pi'\Pi \leq \overline{C}$; for (iv), note that $o_p(1+\Delta_n) - o_p(1+\Delta^\dagger) = o_p(1)$; hence (A.35) is shown. To show (A.34), note that by assumption 4 we have

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{E}\left(((Z_i\widetilde{e}_i)', (\Delta_n Z_i\widetilde{v}_i)')'((Z_i\widetilde{e}_i)', (\Delta_n Z_i\widetilde{v}_i)')\right) = \frac{1}{n}\sum_{i\in[n]} \Lambda_{0,i}(\Delta_n) \otimes Z_iZ_i' \to \Sigma(\Delta^\dagger).$$

Furthermore, for every $\eta > 0$

$$\frac{1}{n}\sum_{i\in[n]}\mathbb{E}\left\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F^2 1\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F \geq \eta\sqrt{n}\}\right\} \to 0.$$

The preceding equation follows from

$$\begin{aligned}
&\left\{\mathbb{E}\left\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F^2 1\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F \geq \eta\sqrt{n}\}\right\}\right\}^2 \\
&\overset{(i)}{\leq} \mathbb{E}||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F^4 \cdot \mathbb{P}\left(n^{-1/2}||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)|| \geq \eta\right) \\
&\overset{(ii)}{\leq} C(1+\Delta^{\dagger 2})\mathbb{P}\left(n^{-1/2}||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F \geq \eta\right) + o(1) \\
&\overset{(iii)}{\leq} C(1+\Delta^{\dagger 2})\frac{||Z_i||_F^2\mathbb{E}(\widetilde{e}_i^2 + \Delta_n\widetilde{v}_i^2)}{\eta^2 n} \leq \frac{C(1+\Delta_n)^2}{n} = \frac{C(1+\Delta^{\dagger})^2}{n} + o(1)
\end{aligned}$$

where (i) follows from Cauchy-Schwartz inequality and (ii) follows from $\sup_i \mathbb{E}||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F^4 \leq 2\sup_i ||Z_i||_F^4 \cdot \mathbb{E}(\widetilde{e}_i^4 + \Delta_n^2\widetilde{v}_i^4) \leq C(1+\Delta_n^2) \leq C(1+\Delta^{\dagger 2}) + o(1) < \infty$, by assumption 2 and 4; (iii) follows from Markov-inequality. We can then apply the Lindeberg-Feller Central-Limit-Theorem to obtain (A.34). We are now ready to prove our result.

Let $\Delta_n = \Delta^{\dagger} = \Delta$. Then

$$(I_K, I_K)\mathcal{N}(0,\Sigma) + \Delta\mu_K = d_n^{-1}\left(d_n(I_K, I_K)\mathcal{N}(0,\Sigma) + \Delta d_n\mu_K\right) = d_n^{-1}\left(o_p(1) + \Delta d_n\mu_K\right),$$

so that WPA1,

$$\begin{aligned}
&(o_p(1) + \Delta d_n\mu_K)' Q_{ZZ}^{-1}(o_p(1) + \Delta d_n\mu_K) \geq mineig(Q_{ZZ}^{-1}) \cdot \Delta^2 d_n^2 \mu_K'\mu_K \\
&= mineig(Q_{ZZ}^{-1}) \cdot \Delta^2 d_n^2 \widetilde{\mu}_n^2 = mineig(Q_{ZZ}^{-1}) \cdot \Delta^2 \widetilde{\mu}^2 > 0.
\end{aligned}$$

Therefore, WPA1, the last line of (A.34) diverges to ∞, as $d_n^{-1} \to \infty$. By (A.34) and (A.35) we have

$$\widehat{Q}(\beta_0) \geq Ce(\beta_0)'Pe(\beta_0) + o_p(1) \to \infty.$$

Furthermore, by lemma 4.2 we know that $q_{1-\alpha}(F_{\widetilde{w}_n}) = O_p(1)$; by lemma 4.1 and (A.20), we have

$$\begin{aligned}
\mathbb{P}\left(\widehat{Q}(\beta_0) > C_\alpha\right) &= \mathbb{P}\left(\widehat{Q}(\beta_0) > q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} - 1\right)\right) \\
&= \mathbb{P}\left(\widehat{Q}(\beta_0) > O_p(1)\right) = 1
\end{aligned}$$

A.9 Proof of Theorem 7

Note that we have $d_n \mu_K = \widetilde{\mu}$ and $\Delta = \Delta_n = d_n \widetilde{\Delta} \to 0$. Then by (A.33), $\Delta_n n^{-1/2} Z' \widetilde{v} = o_p(1)$, whence

$$e(\beta_0)' P e(\beta_0) = (n^{-1/2} Z' \widetilde{e} + \Delta_n n^{-1/2} Z' \Pi)' \left(\frac{Z'Z}{n} \right)^{-1} (n^{-1/2} Z' \widetilde{e} + \Delta_n n^{-1/2} Z' \Pi) + o_p(1)$$
$$= (n^{-1/2} Z' \widetilde{e} + \widetilde{\Delta} \widetilde{\mu})' \left(\frac{Z'Z}{n} \right)^{-1} (n^{-1/2} Z' \widetilde{e} + \widetilde{\Delta} \widetilde{\mu}) + o_p(1)$$

Furthermore, by Lemma B.1, $p_n \frac{\Pi' \Pi}{K} = O(1)$ and $\Delta \to 0$, we have

$$\frac{1}{K} \sum_{i \in [n]} P_{ii} e_i^2(\beta_0) = \frac{1}{K} \sum_{i \in [n]} P_{ii} \sigma_i^2(\beta) + o_p(1) = \frac{1}{K} \sum_{i \in [n]} P_{ii} \widetilde{\sigma}_i^2 + o_p(1)$$

where β is the true parameter. Therefore we have

$$\widehat{Q}(\beta_0) = \frac{(n^{-1/2} Z' \widetilde{e} + \widetilde{\Delta} \widetilde{\mu})' \left(\frac{Z'Z}{n} \right)^{-1} (n^{-1/2} Z' \widetilde{e} + \widetilde{\Delta} \widetilde{\mu})}{\sum_{i \in [n]} P_{ii} \widetilde{\sigma}_i^2} + o_p(1)$$
$$= \left((Z' \Lambda_0 Z)^{-1/2} Z' \widetilde{e} + (n^{-1} Z' \Lambda_0 Z)^{-1/2} \widetilde{\Delta} \widetilde{\mu} \right)' \Omega(\beta) \left((Z' \Lambda_0 Z)^{-1/2} Z' \widetilde{e} + (n^{-1} Z' \Lambda_0 Z)^{-1/2} \widetilde{\Delta} \widetilde{\mu} \right) + o_p(1)$$
$$\rightsquigarrow \left(\mathcal{N}(0, I_K) + \Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right)' \Omega^*(\beta) \left(\mathcal{N}(0, I_K) + \Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right) = \mathcal{Z}_K \left(\Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right)' \Omega^*(\beta) \mathcal{Z}_K \left(\Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right) \tag{A.36}$$

where $\Omega(\beta)$ is defined in (2.6) and the convergence follows from (A.33) and $\Omega^*(\beta) := \lim_{n \to \infty} \Omega(\beta)$. Next, we deal with the critical value. If we show

$$\widetilde{w}_n \xrightarrow{p} w^* \quad \textit{and} \quad \frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}} \sum_{i \in [n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2}} \xrightarrow{p} 1, \tag{A.37}$$

then by (A.36) and (A.20) we can obtain

$$\lim_{n \to \infty} \mathbb{P} \left(\widehat{Q}(\beta_0) > C_\alpha(\widehat{\Phi}_1(\beta_0)) \right) = \mathbb{P} \left(\mathcal{Z}_K \left(\Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right)' \Omega^*(\beta) \mathcal{Z}_K \left(\Sigma(0) \widetilde{\Delta} \widetilde{\mu} \right) > q_{1-\alpha}(F_{w^*}) \right),$$

which completes the proof. Note that by Lemma B.1, since $\Delta \to 0$, we have

$$\widehat{\Phi}_1(\beta_0) = \frac{2}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \widetilde{\sigma}_i^2 \widetilde{\sigma}_j^2 + o_p(1)$$

Repeating the proof of Lemma 4.1 yields

$$\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} = \sqrt{2}||w_n||_F + o_p(1)$$

By Lemma B.3 we have that

$$\max_{i\in[K]}(\widetilde{w}_{i,n} - w_n)^2 = o_p(1)$$

Finally,

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} = \frac{\sqrt{2}||w_n||_F}{\sqrt{2}||\widetilde{w}_n||_F} + o_p(1) \xrightarrow{p} 1,$$

so that together with the assumption that $w_n \to w^*$ (which holds as $\lim_{n\to\infty}\Omega(\beta) \to \Omega^*(\beta)$), (A.37) is shown.

A.10 Proof of Corollary 4.1

The result is a straightforward application of Marden (1982)[Theorem 2.1], by observing that the acceptance region $\mathcal{A} := \{(a_1, ..., a_K) \in \mathbb{R}_+^K : \sum_{i\in[K]} a_i w_i^* \leq q_{1-\alpha}(\sum_{i\in[K]} w_i^* \chi_{1,i}^2)\}$ is convex and monotone decreasing in the sense that if $(a_1, ..., a_K) \in \mathcal{A}$ and $b_i \leq a_i$ for all i, then $b \in \mathcal{A}$

A.11 Proof of Theorem 8:

We begin by noting that $\Delta = \widetilde{\Delta}$ and $\mu_K = \widetilde{\mu}$. Defining $\mathbb{A}_n := n^{-1/2}Z'\widetilde{e} + \widetilde{\Delta}n^{-1/2}Z'\widetilde{v}$, $\mathbb{V}_n := \mathbb{E}\mathbb{A}_n\mathbb{A}_n'$ and $\mathcal{Y}_n := \frac{\widetilde{\Delta}^2\sum_{i\in[n]} P_{ii}\Pi_i^2}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}$, we have

$$\begin{aligned}
\widehat{Q}(\beta_0) &\overset{(i)}{=} \frac{(\mathbb{A}_n + \widetilde{\mu})'(\frac{Z'Z}{n})^{-1}(\mathbb{A}_n + \widetilde{\mu})}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \widetilde{\Delta}^2\sum_{i\in[n]} P_{ii}\Pi_i^2 + o_p(1)} \\
&\overset{(ii)}{=} (\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})'\frac{Z'\Lambda(\beta_0)P\Lambda(\beta_0)Z}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \widetilde{\Delta}^2\sum_{i\in[n]} P_{ii}\Pi_i^2}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1) \\
&= (1+\mathcal{Y}_n)^{-1}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})'\frac{Z'\Lambda(\beta_0)P\Lambda(\beta_0)Z}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1) \\
&\overset{(iii)}{=} (1+\mathcal{Y}_n)^{-1}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})'\Omega(\beta_0)(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1) \\
&\overset{(iv)}{\rightsquigarrow} (1+\mathcal{Y}_n)^{-1}\left(\mathcal{N}(0, I_K) + \Sigma(\widetilde{\Delta})\widetilde{\mu}\right)'\Omega^*(\beta_0)\left(\mathcal{N}(0, I_K) + \Sigma(\widetilde{\Delta})\widetilde{\mu}\right) \qquad \text{(A.38)}
\end{aligned}$$

where (i) follows from Lemma B.1; (ii) follows by recalling that

$$\Lambda(\beta_0) := diag\left((\widetilde{\sigma}_1^2 + 2\widetilde{\Delta}\widetilde{\gamma}_1 + \widetilde{\Delta}^2\widetilde{\zeta}_i^2), ..., (\widetilde{\sigma}_n^2 + 2\widetilde{\Delta}\widetilde{\gamma}_n + \widetilde{\Delta}^2\widetilde{\zeta}_n^2)\right);$$

(iii) follows from definition (2.6); (iv) follows from (A.33). To deal with the critical-value, note that by Lemma B.3 we have that

$$\max_{i\in[K]}(\widetilde{w}_{i,n} - w_n - \lambda_{i,n}^H)^2 = o_p(1)$$

so that

$$\begin{aligned}||\widetilde{w}_n||_F^2 &= ||w_n + \Lambda^H||_F^2 + o_p(1) = ||w_n||_F^2 + \frac{\widetilde{\Delta}^2\sum_{i\in[n]}P_{ii}\Pi_i^2}{\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)} + 2w_n'\Lambda^H + o_p(1)\\ &= ||w_n||_F^2 + \mathcal{Y}_n + 2w_n'\Lambda^H + o_p(1)\end{aligned} \tag{A.39}$$

where $\Lambda^H = (\lambda_{1,n}^H, ..., \lambda_{K,n}^H)$ is defined in Lemma B.3. Furthermore,

$$\begin{aligned}\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)} &\overset{(i)}{=} \frac{\sqrt{\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)} + \mathcal{D}(\widetilde{\Delta})}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0) + \frac{\widetilde{\Delta}^2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\Pi_i^2} + o_p(1)\\ &\overset{(ii)}{=} \frac{\sqrt{\frac{2}{K}\sum_{i,j\in[n]}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)} + \mathcal{D}(\widetilde{\Delta})}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0) + \frac{\widetilde{\Delta}^2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\Pi_i^2} + o_p(1)\\ &= \frac{\frac{\sqrt{\frac{2}{K}\sum_{i,j\in[n]}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)} + \frac{\mathcal{D}(\widetilde{\Delta})}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)}}{1 + \frac{\widetilde{\Delta}^2\sum_{i\in[n]}P_{ii}\Pi_i^2}{\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)}} + o_p(1)\\ &\overset{(iii)}{=} \frac{\sqrt{2}||w_n||_F + \frac{\mathcal{D}(\widetilde{\Delta})}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)}}{1 + \frac{\widetilde{\Delta}^2\sum_{i\in[n]}P_{ii}\Pi_i^2}{\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)}} + o_p(1)\end{aligned}$$

where (i) follows from Lemma B.1 and (c) in the proof of Lemma 4.1; (ii) follows from (b) in the proof of Lemma 4.1; (iii) follows from (a) in the proof of Lemma 4.1. Combining the preceding equation with (A.39) yields

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} = \frac{||w_n||_F + \sqrt{K}\frac{\mathcal{D}(\widetilde{\Delta})\mathcal{Y}_n}{\sum_{i\in[n]}P_{ii}\Pi_i^2}}{(1+\mathcal{Y}_n)\sqrt{||w_n||_F^2 + \mathcal{Y}_n + 2w_n'\Lambda^H}} + o_p(1) \overset{(i)}{=} \frac{||w^*||_F}{\sqrt{||w^*||_F + 2w^{*\prime}\Lambda_H}} + o_p(1). \tag{A.40}$$

where (i) follows from $||w_n - w^*||_F = o(1)$ and

$$\mathcal{Y}_n := \frac{\widetilde{\Delta}^2\sum_{i\in[n]}P_{ii}\Pi_i^2}{\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)} \overset{(ii)}{\leq} p_n\frac{\widetilde{\Delta}^2\sum_{i\in[n]}\Pi_i^2}{\sum_{i\in[n]}P_{ii}} = \frac{\widetilde{\Delta}^2 p_n\Pi'\Pi}{K} \overset{(iii)}{=} o(1),$$

(ii) follows from $\sigma_i^2(\beta_0) \geq \underline{C} > 0$ by assumption 2, (iii) follows from $\Pi'\Pi = O(1)$ and $\frac{p_n}{K} = o(1)$ by assumption 2. Furthermore, we can show that

$$\Lambda_H = (n^{-1}Z'Z)^{-1/2}\frac{Z'H_nZ}{n}(n^{-1}Z'Z)^{-1/2} \to 0, \tag{A.41}$$

which follows from

$$\begin{aligned}\lambda_{\max}(\frac{Z'H_nZ}{n}) &= \widetilde{\Delta}^2\lambda_{\max}(\frac{1}{n}\sum_{i\in[n]} Z_iZ_i'\Pi_i^2) \leq \frac{\widetilde{\Delta}^2}{n}\sum_{i\in[n]}\lambda_{\max}(Z_iZ_i'\Pi_i^2)\\ &\leq \frac{\widetilde{\Delta}^2}{n}\sum_{i\in[n]}\Pi_i^2||Z_i||_F^2 \overset{(i)}{\leq} C\widetilde{\Delta}^2\frac{\Pi'\Pi}{n} = o(1)\end{aligned}$$

where (i) follows from $\sup_i ||Z_i||_F < \infty$ by assumption 4. Therefore, combining (A.40) and (A.41) yields

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2}} \xrightarrow{p} 1 \tag{A.42}$$

Finally, since $\lambda_{i,n}^H \to 0$ and $\max_{i\in[K]}(\widetilde{w}_{i,n} - w_n - \lambda_{i,n}^H)^2 = o_p(1)$, we have $||\widetilde{w}_n - w_n||_F^2 = o_p(1)$. This implies

$$q_{1-\alpha}(F_{\widetilde{w}_n}) = q_{1-\alpha}(F_{w_n}) + o_p(1) \xrightarrow{p} q_{1-\alpha}(F_{w^*})$$

In view of the preceding equation, (A.38), (A.42) and (2.8), we have Theorem 8.

A.12 Proof of Corollary 4.2

Repeat the proof of corollary 4.1 and replace $\mathbb{M}_i$ by $\overline{\mathbb{M}}_i$ for each i

B Auxiliary Lemmas

Lemma B.1. *Under Assumption 1 and 2, for any fixed $\Delta := \beta - \beta_0$ not necessarily zero,*

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \frac{\Delta^2}{K}\sum_{i\in[n]} P_{ii}\Pi_i^2 + o_p(1),$$

where $\frac{\Delta^2}{K}\sum_{i\in[n]} P_{ii}\Pi_i^2 = O_p(\Delta^2 p_n\frac{\Pi'\Pi}{K})$

Proof of Lemma B.1:
To begin, recall

$$\sigma_i^2(\beta_0) = \widetilde{\sigma}_i{}^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i \tag{B.1}$$

Furthermore,

$$\begin{aligned}
e_i^2(\beta_0) &= (e_i + \Delta X_i)^2 = ((M_i^W)'\widetilde{e} + \Delta\Pi_i + \Delta v_i)^2 \\
&= ((M_i^W)'\widetilde{e})^2 + 2\Delta\Pi_i(M_i^W)'\widetilde{e} + 2\Delta v_i(M_i^W)'\widetilde{e} + \Delta^2\Pi_i^2 + 2\Delta^2\Pi_i v_i + \Delta^2 v_i^2 \\
&= A_{i,1} + 2\Delta A_{i,2} + 2\Delta A_{i,3} + \Delta^2 A_{i,4} + 2\Delta^2 A_{i,5} + \Delta^2 A_{i,6} \qquad \text{(B.2)}
\end{aligned}$$

We will show that

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,1} - \widetilde{\sigma}_i^2) = O_p\left(\sqrt{\frac{p_n}{K}} + \sqrt{p_n^W}\right) \qquad \text{(B.3)}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,2} = O_p(\sqrt{\frac{p_n}{K}}), \qquad \text{(B.4)}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,3} - \widetilde{\gamma}_i) = O_p(\sqrt{\frac{p_n}{K}} + \sqrt{p_n^W})), \qquad \text{(B.5)}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,4} = O_p(\Delta^2 p_n \frac{\Pi'\Pi}{K}) \qquad \text{(B.6)}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,5} = O_p(\sqrt{\frac{p_n}{K} + p_n^W}). \quad \text{and} \qquad \text{(B.7)}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,6} - \widetilde{\varsigma}_i^2) = O_p(\sqrt{\frac{p_n}{K}} + \sqrt{p_n^W}) \qquad \text{(B.8)}$$

Observe that

$$\begin{aligned}
\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,1} - \widetilde{\sigma}_i^2) &= \frac{1}{K}\sum_{i\in[n]} P_{ii}(\widetilde{e}_i^2 - \widetilde{\sigma}_i^2) - \frac{2}{K}\sum_{i\in[n]} P_{ii}\sum_{j\in[n]} P_{ij}^W \widetilde{e}_j\widetilde{e}_i + \frac{1}{K}\sum_{i\in[n]} P_{ii}(\sum_{j\in[n]} P_{ij}^W\widetilde{e}_j)^2 \\
&= B_1 + B_2 + B_3
\end{aligned}$$

By Markov inequality and

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]} P_{ii}(\widetilde{e}_i^2 - \widetilde{\sigma}_i^2)\right)^2 \leq \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2 = O(\frac{p_n}{K})$$

we have that $B_1 = O_p(\sqrt{\frac{p_n}{K}})$. Since

$$\begin{aligned}
\mathbb{E}(B_2)^2 &\leq \frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\in[n]} P_{ii}P_{i'i'}\sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{i'j'}^W \mathbb{E}(\widetilde{e}_i\widetilde{e}_j\widetilde{e}_{i'}\widetilde{e}_{j'}) \\
&= \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2 \sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{ij'}^W \mathbb{E}(\widetilde{e}_i^2\widetilde{e}_j\widetilde{e}_{j'}) + \frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\neq i} P_{ii}P_{i'i'}\sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{i'j'}^W \mathbb{E}(\widetilde{e}_i\widetilde{e}_j\widetilde{e}_{i'}\widetilde{e}_{j'})
\end{aligned}$$

$$\leq \frac{C}{K^2}\sum_{i\in[n]}P_{ii}^2\sum_{j\in[n]}(P_{ij}^W)^2+\frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\neq i}P_{ii}P_{i'i'}(P_{ii}^W P_{i'i'}^W+(P_{ii'}^W)^2)$$

$$\leq Cp_n^W \tag{B.9}$$

we have $B_2 = O_p(\sqrt{p_n^W})$. Also,

$$\mathbb{E}B_3=\frac{1}{K}\sum_{i\in[n]}P_{ii}\sum_{j\in[n]}(P_{ij}^W)^2\widetilde{\sigma}_i^2\leq\frac{C}{K}\sum_{i\in[n]}P_{ii}P_{ii}^W\leq Cp_n^W=O(p_n^W)$$

so that putting it all together yields (B.3). Next, we can express $A_{i,2}=\Pi_i\widetilde{e}_i-\Pi_i(P_i^W)'\widetilde{e}\equiv A_{i,2,1}+A_{i,2,2}$. By Markov inequality,

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}P_{ii}\Pi_i\widetilde{e}_i\right)^2\leq\frac{C}{K^2}\sum_{i\in[n]}P_{ii}^2\leq\frac{Cp_n}{K}=O(\frac{p_n}{K})$$

and

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}P_{ii}A_{i,2,2}\right)^2\leq\frac{C}{K^2}\sum_{i,j\in[n]}P_{ii}P_{jj}|\Pi_i||\Pi_j|\sum_{\ell\in[n]}|P_{i\ell}^W P_{j\ell}^W|\leq Cp_n^W,$$

we obtain (B.4). For (B.5), observe that $v_i=\widetilde{v}_i-\sum_{j\in[n]}P_{ij}^W\widetilde{v}_j$ and $M_i'\widetilde{e}=\widetilde{e}_i-\sum_{j\in[n]}P_{ij}^W\widetilde{e}_j$, so that

$$\begin{aligned}\frac{1}{K}\sum_{i\in[n]}P_{ii}(A_{i,3}-\widetilde{\gamma}_i)^2&=\frac{1}{K}\sum_{i\in[n]}P_{ii}(\widetilde{e}_i\widetilde{v}_i-\widetilde{\gamma}_i)-\frac{1}{K}\sum_{i\in[n]}P_{ii}\widetilde{v}_i\sum_{j\in[n]}P_{ij}^W\widetilde{e}_j\\&-\frac{1}{K}\sum_{i\in[n]}P_{ii}\widetilde{e}_i\sum_{j\in[n]}P_{ij}^W\widetilde{v}_j+\frac{1}{K}\sum_{i\in[n]}P_{ii}(\sum_{j\in[n]}P_{ij}^W\widetilde{e}_j)(\sum_{j\in[n]}P_{ij}^W\widetilde{v}_j)\\&\equiv B_5+B_6+B_7+B_8\end{aligned}$$

Note $B_5=O_p(\sqrt{\frac{p_n}{K}})$ and $B_6=O_p(\sqrt{p_n^W})$ by

$$\mathbb{E}B_5^2\leq\frac{C}{K^2}\sum_{i\in[n]}P_{ii}^2=O(\frac{p_n}{K}),$$

and

$$\mathbb{E}B_6^2\leq Cp_n^W$$

as in (B.9); the argument for $B_7=O_p(\sqrt{p_n^W})$ is analogous to B_6. Furthermore, by

$$\mathbb{E}B_8^2\leq\frac{C}{K^2}\sum_{i,i'\in[n]}P_{ii}P_{i'i'}(\sum_{j\in[n]}\sum_{j'\in[n]}(P_{ij}^W)^2(P_{ij'}^W)^2+\sum_{j\in[n]}(P_{ij}^W)^4)\leq\frac{C(p_n^W)^2}{K^2}(\sum_{i\in[n]}P_{ii})^2=O((p_n^W)^2)$$

we have (B.5). Next, (B.6) is obvious. For (B.7), noting that $v_i v_{i'} = \widetilde{v}_i\widetilde{v}_{i'} + \sum_{\ell\in[n]} P^W_{i\ell}\widetilde{v}_\ell \sum_{\ell\in[n]} P^W_{i'\ell}\widetilde{v}_\ell - \sum_{\ell\in[n]} P^W_{i'\ell}\widetilde{v}_\ell\widetilde{v}_i - \sum_{\ell\in[n]} P^W_{i\ell}\widetilde{v}_\ell\widetilde{v}_{i'}$, we have

$$\begin{aligned}
\mathbb{E}(\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,5})^2 &= \frac{C}{K^2}\sum_{i,i'\in[n]} P_{ii}\Pi_i P_{i'i'}\Pi_{i'}\mathbb{E}(v_i v_{i'}) \\
&\le \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2\Pi_i^2 + \frac{C}{K^2}\sum_{i,i'\in[n]} P_{ii}|\Pi_i|P_{i'i'}|\Pi_{i'}|\sum_{\ell\in[n]}|P^W_{i\ell}P^W_{i'\ell}| + \frac{C}{K^2}\sum_{i,i'\in[n]} P_{ii}|\Pi_i|P_{i'i'}|\Pi_{i'}||P^W_{i'i}| \\
&\le C\frac{p_n}{K^2}\sum_{i\in[n]} P_{ii} + \frac{C}{K^2}\sum_{i,i'\in[n]} P_{ii}P_{i'i'}\sqrt{\sum_{\ell\in[n]}(P^W_{i\ell})^2}\sqrt{\sum_{\ell\in[n]}(P^W_{i'\ell})^2} + Cp^W_n \\
&\le C\frac{p_n}{K} + Cp^W_n + Cp^W_n = O(\frac{p_n}{K} + p^W_n)
\end{aligned}$$

Finally we deal with (B.8). Since $v_i^2 = \widetilde{v}_i^2 - 2\sum_{j\in[n]} P^W_{ij}\widetilde{v}_i\widetilde{v}_j + (\sum_{j\in[n]} P^W_{ij}\widetilde{v}_i)^2$, we have

$$\begin{aligned}
\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,6} - \widetilde{\varsigma}_i^2) &= \frac{1}{K}\sum_{i\in[n]} P_{ii}(\widetilde{v}_i^2 - \widetilde{\varsigma}_i^2) - \frac{2}{K}\sum_{i\in[n]} P_{ii}\sum_{j\in[n]} P^W_{ij}\widetilde{v}_i\widetilde{v}_j + \frac{1}{K}\sum_{i\in[n]} P_{ii}(\sum_{j\in[n]} P^W_{ij}\widetilde{v}_i)^2 \\
&= B_9 + B_{10} + B_{11}
\end{aligned}$$

Observe $B_9 = O_p(\sqrt{\frac{p_n}{K}})$ by

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]} P_{ii}(\widetilde{v}_i^2 - \widetilde{\varsigma}_i^2)\right)^2 \le \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2 = O(\frac{p_n}{K}).$$

Furthermore, similar to (B.9) we have

$$\mathbb{E}B_{10}^2 \le Cp^W_n = O(p^W_n)$$

and

$$\mathbb{E}B_{11} \le \frac{C}{K}\sum_{i\in[n]} P_{ii}\sum_{j\in[n]}(P^W_{ij})^2 \le Cp^W_n = O(p^W_n)$$

This completes the proof of (B.8). By the assumption of $\frac{p_n}{K} = o(1)$ and $p^W_n = o(1)$, each term from (B.3)-(B.8) except (B.6) is $o_p(1)$. Hence Lemma B.1 is shown.

□

Lemma B.2. *Suppose Assumption 1 and 2 holds. Then for fixed* Δ *not necessarily zero,*

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2(\beta_0)\sigma_j^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + \frac{\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i^2\sigma_j^2(\beta_0) + o_p(1)$$

Proof of Lemma B.2:

Step 1: We first show that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 e_i^2\sigma_j^2(\beta_0)=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2\sigma_j^2(\beta_0)+o_p(1) \tag{B.10}$$

Note $\sigma_i^2=\widetilde{\sigma}_i^2$, so we can express

$$\begin{aligned}e_i^2-\sigma_i^2&=\left(\widetilde{e}_i^2-\widetilde{\sigma}_i^2\right)-2\sum_{j\in[n]}P_{ij}^W\widetilde{e}_j\widetilde{e}_i+(\sum_{j\in[n]}P_{ij}^W\widetilde{e}_j)^2\\&=C_{i,1}+C_{i,2}+C_{i,3}.\end{aligned}$$

Therefore

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_j^2(\beta_0)(C_{i,1}+C_{i,2}+C_{i,3})\right)^2\\&=\frac{1}{K^2}\sum_{\ell=1}^{3}\sum_{\ell'=1}^{3}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{i'j'}^2\sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}(C_{i,\ell}C_{i',\ell'})\\&\equiv\frac{1}{K^2}\sum_{\ell=1}^{3}\sum_{\ell'=1}^{3}B_{\ell,\ell'}\end{aligned}$$

We will show that $\frac{1}{K^2}B_{\ell,\ell'}=o(1)$ for each $\ell,\ell'\in\{1,2,3\}$, which will complete the proof by Markov inequality. First,

$$\begin{aligned}\frac{1}{K^2}B_{1,1}&=\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{i'j'}^2\sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}(C_{i,1}C_{i',1})\\&=\frac{1}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{ij'}^2\sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}C_{i,1}^2\le\frac{C}{K^2}p_nK=o(1)\end{aligned}$$

where the inequality is from

$$\mathbb{E}C_{i,1}^2=\mathbb{E}(\widetilde{e}_i^2-\widetilde{\sigma}_i^2)^2\le\mathbb{E}\widetilde{e}_i^4+\widetilde{\sigma}_i^4\le C$$

Second,

$$\begin{aligned}\frac{1}{K^2}B_{1,2}&=\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{i'j'}^2\sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}(\widetilde{e}_i^2-\widetilde{\sigma}_i^2)(\sum_{k\in[n]}P_{i'k}^W\widetilde{e}_k\widetilde{e}_{i'}))\\&\le\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{ij'}^2\sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)P_{ii}^W\le\frac{Cp_n^W}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{j'\neq i}P_{ij}^2P_{ij'}^2\le Cp_n^W=o(1),\end{aligned}$$

Third, note that

$$C_{i,3}=\sum_{j\neq i}(P_{ij}^W)^2\widetilde{e}_j^2+\sum_{j\neq i}\sum_{k\neq i,j}P_{ij}^WP_{kj}^W\widetilde{e}_j\widetilde{e}_k \tag{B.11}$$

so

$$\frac{1}{K^2}B_{1,3} = \frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\widetilde{e}_i^2-\widetilde{\sigma}_i^2)(\sum_{k\neq i'}(P_{i'k}^W)^2\widetilde{e}_k^2)\right)$$
$$+\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\widetilde{e}_i^2-\widetilde{\sigma}_i^2)(\sum_{k\neq i'}\sum_{k'\neq i',k} P_{i'k}^W P_{k'k}^W\widetilde{e}_k\widetilde{e}_{k'})\right)$$
$$=\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\widetilde{e}_i^2-\widetilde{\sigma}_i^2)(\sum_{k\neq i'}(P_{i'k}^W)^2\widetilde{e}_k^2)\right)$$
$$\leq \frac{Cp_n^W)}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \leq Cp_n^W = o(1).$$

Fourth, the proof that $\frac{1}{K}B_{2,1} = o_p(1)$ is analogous to that of $\frac{1}{K}B_{1,2} = o_p(1)$. Fifth, using the simple inequality of $|ab| \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$

$$\frac{1}{K^2}B_{2,2} = \frac{4}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\sum_{k\in[n]} P_{ik}^W\widetilde{e}_k\widetilde{e}_i)(\sum_{k\in[n]} P_{i'k}^W\widetilde{e}_k\widetilde{e}_{i'})\right)$$
$$\leq \frac{4}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\sum_{k\in[n]} P_{ik}^W\widetilde{e}_k\widetilde{e}_i)\right)^2$$
$$\leq \frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 (\sum_{k\neq i}(P_{ik}^W)^2 \leq Cp_n^W = o(1).$$

Sixth,

$$\frac{1}{K^2}B_{2,3} \overset{(B.11)}{=} \frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\sum_{k\neq i} P_{ik}^W\widetilde{e}_k\widetilde{e}_i)(\sum_{k\neq i'}(P_{i'k}^W)^2\widetilde{e}_k^2\right)$$
$$+\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\mathbb{E}\left((\sum_{\ell\neq i} P_{i\ell}^W\widetilde{e}_\ell\widetilde{e}_i)(\sum_{k\neq i'}\sum_{k'\neq i',k} P_{i'k}^W P_{k'k}^W\widetilde{e}_k\widetilde{e}_{k'})\right)$$
$$\leq \frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)P_{ii}^W$$
$$+\frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0)\sigma_{j'}^2(\beta_0)\sum_{\ell\neq i}(|P_{i\ell}^W P_{i'\ell}^W P_{i\ell}^W| + (P_{i\ell}^W)^2|P_{ii'}^W|)$$
$$\leq \frac{Cp_n^W}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \leq Cp_n^W = o(1).$$

Seventh, the proof that $\frac{1}{K}B_{3,1} = o_p(1)$ is analogous to that of $\frac{1}{K}B_{1,3} = o_p(1)$. Eighth, that

$\frac{1}{K}B_{3,2} = o_p(1)$ is analogous to that of $\frac{1}{K}B_{2,3} = o_p(1)$. Finally, using $2|ab| \leq a^2 + b^2$,

$$\begin{aligned}
\frac{1}{K^2}B_{3,3} &\leq \frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \mathbb{E}\left((\sum_{k\in[n]} P_{ik}^W \widetilde{e}_k)^2 (\sum_{k\in[n]} P_{i'k}^W \widetilde{e}_k)^2 \right) \\
&\leq \frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \left(\sum_{k\in[n]}\sum_{k'\in[n]} (P_{ik}^W)^2 (P_{i'k'}^W)^2 + \sum_{k\in[n]}\sum_{k'\in[n]} |P_{ik}^W P_{i'k}^W P_{ik'}^W P_{i'k'}^W| \right) \\
&\leq \frac{C(p_n^W)^2}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i} P_{ij}^2 P_{i'j'}^2 \leq C(p_n^W)^2 = o(1)
\end{aligned}$$

The proof of (B.10) is complete.

Step 2: We complete the proof.
Note that we can write $e_i(\beta_0) = e_i^2 + \Delta^2(\Pi_i^2 + v_i^2 + 2\Pi_i v_i) + 2\Delta v_i e_i + 2\Delta\Pi_i e_i$, so

$$e_i^2(\beta_0) - \sigma_i^2(\beta_0) = (e_i^2 - \widetilde{\sigma}_i^2) + \Delta^2(v_i^2 - \widetilde{\varsigma}_i^2) + 2\Delta\Pi_i v_i + 2\Delta\Pi_i e_i + 2\Delta(v_i e_i - \widetilde{\gamma}_i) + \Delta^2\Pi_i^2$$

Note that by the same proof as step 1, we have

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 v_i^2 \sigma_j^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \widetilde{\varsigma}_i^2 \sigma_j^2(\beta_0) + o_p(1) \tag{B.12}$$

and

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 v_i e_i \sigma_j^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \widetilde{\gamma}_i \sigma_j^2(\beta_0) + o_p(1) \tag{B.13}$$

Finally, we will show that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_j^2(\beta_0)\Pi_i e_i = o_p(1) \tag{B.14}$$

and

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_j^2(\beta_0)\Pi_i v_i = o_p(1) \tag{B.15}$$

We will only show (B.14) since (B.15) follows the same proof. By the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and $e_i = \widetilde{e}_i - (P_i^W)'\widetilde{e}$, we have

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_j^2(\beta_0)\Pi_i e_i \right)^2$$

$$\leq 2\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_j^2(\beta_0)\Pi_i\widetilde{e}_i\right)^2+2\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_j^2(\beta_0)\Pi_i(P_i^W)'\widetilde{e}\right)^2\equiv A_1+A_2\overset{(i)}{=}o(1),$$

where (i) follows from

$$A_1\leq\frac{C}{K^2}\sum_{i,j,j'\in[n]}P_{ij}^2P_{ij'}^2\leq\frac{Cp_n}{K}=o(1)$$

and

$$A_2\leq\frac{C}{K^2}\sum_{i,i',j,j'}P_{ij}^2P_{i'j'}^2\sum_{\ell\in[n]}|P_{i\ell}^WP_{i'\ell}^W|\overset{(ii)}{\leq}\frac{Cp_n^W}{K^2}\sum_{i,i',j,j'}P_{ij}^2P_{i'j'}^2=Cp_n^W=o(1)$$

where (ii) follows from Cauchy-Schwartz inequality. Therefore, by Markov inequality we have (B.14). Combining (B.10)-(B.15) yields Lemma B.2

□

Lemma B.3. *Suppose Assumption 1, 2 and 3 holds. Fix any Δ not necessarily zero. For either fixed or diverging K, consider any sub-sequence $n_j\subset n$. Then there exists a further sub-sequence $n_{j_k}\subset n_j$ such that*

$$\max_{i\in[K]}(\widetilde{w}_{i,n_{j_k}}-w_{i,n_{j_k}}-\lambda_{i,n_{j_k}}^H)^2=o_p(1)$$

where $\Lambda_H=(\lambda_{1,n}^H,...,\lambda_{K,n}^H)$ are the eigenvalues of $\Omega_H(\beta_0):=\frac{U'H_nU}{\sum_{i\in[n]}P_{ii}\sigma_i^2(\beta_0)}$, $H_n:=diag(T_{1,n},...,T_{n,n})$ and $T_{i,n}:=\Delta^2\Pi_i^2$. Furthermore,
(i) for $K\to\infty$, $\max_i\widetilde{w}_{i,n}=o(K^{-1/2})$;
(ii) for fixed K, if w_n converges to a limit under the full-sequence (i.e. $||w_n-w^*||_F=o(1)$), then

$$\max_{i\in[K]}(\widetilde{w}_{i,n}-w_{i,n}-\lambda_{i,n}^H)^2=o_p(1)$$

Proof of Lemma B.3:
For notational simplicity, we abuse notation and write $T_i\equiv T_{i,n}$. Furthermore, we write $\widehat{\Lambda}(\beta_0)$ and $\Lambda(\beta_0)$ as $\widehat{\Lambda}$ and Λ respectively. Note that for both fixed and diverging K, we have

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2(\beta_0)-\sigma_i^2(\beta_0)-T_i)(e_j^2(\beta_0)-\sigma_j^2(\beta_0)-T_j)=o_p(1)\qquad\text{(B.16)}$$

where the last equality follows from

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2(\beta_0)-\sigma_i^2(\beta_0)-T_i)(e_j^2(\beta_0)-\sigma_j^2(\beta_0)-T_i)=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2(\beta_0)-T_i)(e_j^2(\beta_0)-T_j)$$
$$+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)-\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2(\beta_0)-T_i)\sigma_j^2(\beta_0)-\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_j^2(\beta_0)-T_j)\sigma_i^2(\beta_0)$$

$$\overset{(i)}{=} 2\Phi_1 - \frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(e_i^2(\beta_0) - T_i)\sigma_j^2(\beta_0) + o_p(1) \overset{(ii)}{=} 2\Phi_1 - 2\Phi_1 + o_p(1) = o_p(1)$$

where (i) follows from noting that by repeating the proof of Theorem C.0.1 will show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(e_i^2(\beta_0) - T_i)(e_j^2(\beta_0) - T_j) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1) = \Phi_1 + o_p(1);$$

(ii) follows from noting that by repeating the proof of **Step 2** in Lemma B.2, we can show in a similar manner that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(e_i^2(\beta_0) - T_i)\sigma_j^2(\beta_0) = \Phi_1 + o_p(1).$$

Fixed K case: Assume first that K is fixed. Then we have

$$\begin{aligned}
&\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)(e_j^2(\beta_0) - \sigma_j^2(\beta_0) - T_j) \\
&= \frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)(e_j^2(\beta_0) - \sigma_j^2(\beta_0) - T_j) \\
&+ \frac{1}{K}\sum_{i\in[n]} P_{ii}^2\mathbb{E}(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)^2 = o_p(1)
\end{aligned}$$

where the last equality follows from (B.16) and

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}^2\mathbb{E}(e_i^2(\beta_0) - \sigma_i^2(\beta_0))^2 \leq \frac{C}{K}\sum_{i\in[n]} P_{ii}^2 \leq Cp_n = \frac{p_n}{K}K = o(1)$$

for fixed K. Therefore

$$\begin{aligned}
&||U'\hat{\Lambda}U - U'\Lambda U - U'H_nU||_F^2 = \mathbb{E}||U'(\hat{\Lambda} - \Lambda - H_n)U||_F^2 \\
&= \mathbb{E}trace(U'(\hat{\Lambda} - \Lambda - H_n)UU'(\hat{\Lambda} - \Lambda - H_n)U) \\
&= trace\left((Z'Z)^{-1/2}\sum_{i\in[n]} Z_iZ_i'(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)(Z'Z)^{-1}\sum_{j\in[n]} Z_iZ_i'(e_j^2(\beta_0) - \sigma_j^2(\beta_0) - T_j)(Z'Z)^{-1/2}\right) \\
&= \sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)(e_j^2(\beta_0) - \sigma_j^2(\beta_0) - T_j) = o_p(1),
\end{aligned}$$

which gives us

$$||U'\hat{\Lambda}U - U'\Lambda U - U'H_nU||_F = o_p(1) \tag{B.17}$$

Then we have

$$||\widehat{\Omega}_{s,n}(\beta_0) - \Omega_{s,n}(\beta_0) - \Omega_H(\beta_0)||_F^2 = \left|\left|\frac{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)U'\Lambda U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\cdot\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_F^2$$

$$= \frac{1/K^2}{\left(\frac{1}{K}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\cdot\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\right)^2}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0))U'\Lambda U\right|\right|_F^2$$

$$\overset{(i)}{=} \frac{1/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4 + o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)U'\Lambda U\right|\right|_F^2$$

$$\overset{(ii)}{\leq} \frac{2/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4 + o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-\Lambda-H_n)U\right|\right|_F^2$$

$$+ \frac{2/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4 + o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}(e_i^2(\beta_0)-\sigma_i^2(\beta_0))\cdot U'\Lambda U\right|\right|_F^2$$

$$\leq \frac{2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4 + o_p(1)}\left|\left|\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\right|\right|_F^2\cdot\left|\left|U'(\hat{\Lambda}-\Lambda-H_n)U\right|\right|_F^2$$

$$+ \frac{2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4 + o_p(1)}\left|\left|\frac{1}{K}\sum_{i\in[n]} P_{ii}(e_i^2(\beta_0)-\sigma_i^2(\beta_0))\right|\right|_F^2\cdot\left|\left|U'\Lambda U\right|\right|_F^2 \overset{(iii)}{=} o_p(1)$$

where (i) follows from Lemma B.1; (ii) follows from $(a+b)^2 \leq 2a^2 + 2b^2$; (iii) follows from

(a) $$\left|\left|\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\right|\right|_F^2 \leq \left|\left|\max_i \sigma_i^2(\beta_0)\right|\right|_F^2 \leq \max_i(\sigma_i^2 + \Delta^2\varsigma_i^2 + 2\Delta\gamma_i) = O(1)$$

(b) $$\left|\left|\frac{1}{K}\sum_{i\in[n]} P_{ii}\{e_i^2(\beta_0)-\sigma_i^2(\beta_0)\}\right|\right|_F^2 = ||o_p(1)||_F^2 = o_p(1) \text{ by Lemma } B.1$$

(c) $$\left|\left|U'(\hat{\Lambda}-\Lambda-H_n)U\right|\right|_F^2 = o_p(1) \text{ by } (B.17)$$

(d) $$\left|\left|U'\Lambda U\right|\right|_F^2 = \sum_{i\in[n]} P_{ii}\sigma_i^2 = O(K) = O(1)$$

(e) $$\frac{1}{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \leq \frac{1}{\frac{\underline{C}}{K}\sum_{i\in[n]} P_{ii}} = \frac{1}{\underline{C}} = O(1).$$

Note that

$$||\Omega_{s,n}(\beta_0)||_F^2 = \frac{1}{(\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2}||U'\Lambda U||_F^2 = \frac{1}{(\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2}\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$$

$$\leq \frac{1}{C_1} \sum_{i\in[n]} \sum_{j\in[n]} P_{ij}^2 \sigma_i^2(\beta_0)\sigma_j^2(\beta_0) = O(1).$$

therefore, by Bolzano-Weierstrass Theorem, for every sub-sequence n_j there exists a further sub-sequence n_{j_k} such that $\Omega_{s,n_{j_k}}(\beta_0) \to \Omega^*(\beta_0)$. Let w^* to be the eigenvalues of $\Omega^*(\beta_0)$, so that $w_i^* \geq 0$ and $\sum_{i\in K} w_i^* = 1$. By continuous mapping theorem, $w_{i,n_{j_k}} \to w_i^*$ for each $i \in [K]$. By $||\widehat{\Omega}_{s,n}(\beta_0) - \Omega_{s,n}(\beta_0) - \Omega_H(\beta_0)||_F^2 = o_p(1)$ and $||\Omega_{s,n_{j_k}}(\beta_0) - \Omega^*(\beta_0)||_F^2 = o(1)$, we know

$$||\widehat{\Omega}_{s,n_{j_k}}(\beta_0) - \Omega^*(\beta_0) - \Omega_H(\beta_0)||_F^2 = o_p(1)$$

Given that $\widetilde{w}_n$ are the eigenvalues of $\widehat{\Omega}_{s,n}(\beta_0)$, by continuous mapping theorem $\widetilde{w}_{n_{j_k}} - \lambda^H_{n_{j_k}} \xrightarrow{p} w^*$. Clearly this means that $\max_{i\in[K]}(\widetilde{w}_{i,n_{j_k}} - w_{i,n_{j_k}} - \lambda^H_{i,n_{j_k}})^2 = o_p(1)$. This concludes the proof for fixed K.

Diverging K case: Assume now that $K \to \infty$.
Note first that

$$\frac{1}{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \leq \frac{1}{\frac{C}{K}\sum_{i\in[n]} P_{ii}} = \frac{1}{\underline{C}} \leq C.$$

We will show that[22]

$$\max_i \widetilde{w}_{i,n} = o_p(K^{-1/2}) = o_p(1) \tag{B.18}$$

To this end, denote $||\cdot||_S$ as the spectral-norm. Observe that

$$\begin{aligned} \max_i w_{i,n} = ||\Omega_s(\beta_0)||_S &= \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||U'\Lambda U||_S \leq \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||U||_S^2||\Lambda||_S \\ &\overset{(i)}{=} \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||\Lambda||_S = \frac{\max_i \sigma_i^2(\beta_0)}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \overset{(ii)}{\leq} \frac{C/K}{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} = o(K^{-1/2}) \end{aligned} \tag{B.19}$$

where (i) follows by $U'U = I_K$; (ii) follows from expression (B.1). Furthermore, we have

$$\max_i \lambda^H_{i,n} = ||\Omega_H(\beta_0)||_S = \frac{||U'H_nU||_S}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \leq \frac{||H_n||_S}{K\underline{C}} = \frac{\max_i \Delta^2\Pi_i^2}{K\underline{C}} \leq \frac{C}{K} = o(K^{-1/2}) \tag{B.20}$$

Next, we can orthogonally diagonalize $\Omega_s(\beta_0) = Q_1'D_wQ_1$, $\widehat{\Omega}_s(\beta_0) = Q_2'D_{\widetilde{w}}Q_2$ and $\Omega_H(\beta_0) = Q_3'\Lambda_H Q_3$, where $D_{\widetilde{w}} = diag(\widetilde{w}_{1,n}, ..., \widetilde{w}_{K,n})$, $D_w = diag(w_{1,n}, ..., w_{K,n})$; $Q_1'Q_1 = Q_1'Q_1 = I_K = Q_2'Q_2 = Q_2Q_2' = Q_3'Q_3 = Q_3Q_3'$. Then

$$\max_{i\in[n]}(\widetilde{w}_{i,n} - w_{i,n} - \lambda^H_{i,n})^2 = ||D_{\widetilde{w}} - D_w - \Lambda_H||_S^2 \overset{(i)}{=} ||\widehat{\Omega}_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A} - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S^2$$

[22]The reason we show that $\max_i \widetilde{w}_{i,n} = o_p(K^{-1/2})$ instead of showing $o_p(1)$ immediately is that we will be using this property in the proof of Theorem 2 later on

$$\leq \Big(||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S + ||\Omega_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A} + \Omega_H(\beta_0) - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S\Big)^2$$
$$\overset{(ii)}{\leq} 4||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 + 4||\Omega_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A}||_S^2 + 4||\Omega_H(\beta_0) - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S^2$$
$$\overset{(iii)}{\leq} 4||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 + o(K^{-1}) \tag{B.21}$$

where (i) follows from $\mathcal{A}' := Q_1'Q_2$ and $\mathcal{B}' := Q_1'Q_3$; (ii) follows from the simple inequality $(a+b)^2 \leq 2a^2 + 2b^2$; the first part of (iii) follows from

$$4||\Omega_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A}||_S^2 \leq 8||\Omega_s(\beta_0)||_S^2 + 8||\mathcal{A}'\Omega_s(\beta_0)\mathcal{A}||_S^2 \overset{(iv)}{\leq} 16||\Omega_s(\beta_0)||_S^2 \overset{(v)}{=} o(K^{-1})$$

with (iv) following from $\mathcal{A}'\mathcal{A} = I_K$ and (v) following in the same manner as (B.19). The second part of (iii) follows from

$$4||\Omega_H(\beta_0) - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S^2 \leq 16||\Omega_H(\beta_0)||_S^2 \leq \frac{||U||_S^2||H_n||_S^2}{(\sum_{i\in[K]} P_{ii}\sigma_i^2(\beta_0))^2} \leq \frac{||H_n||_S^2}{K^2\underline{C}^2} \leq \frac{C}{K^2} = o(K^{-1}).$$

Next, we can express

$$||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 = \left|\left|\frac{U'\hat{\Lambda}U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} - \frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_S^2$$
$$\leq 2\left|\left|\frac{U'(\hat{\Lambda} - \Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}\right|\right|_S^2 + 2\left|\left|\frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} - \frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_S^2$$
$$\leq 2\left|\left|\frac{U'(\hat{\Lambda} - \Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}\right|\right|_S^2 + \frac{2(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0) - \sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2 \cdot ||U'(\Lambda - H_n)U||_S^2}{\Big(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0) \cdot \sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\Big)^2}$$
$$\overset{(i)}{=} \frac{2||U'(\hat{\Lambda} - \Lambda - H_n)U||_S^2}{(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0))^2} + o(K^{-2}) \tag{B.22}$$

where (i) follows from Lemma B.1 and $||U'(\Lambda - H_n)U||_S^2 \leq ||\Lambda - H_n||_S^2 = \max_i(\sigma_i^2(\beta_0) - \Delta^2\Pi_i^2)^2 \leq C$, in the same manner as in (B.19). We now separate the problem into two cases now to consider: **(A)** $\frac{K}{n} = o(1)$ and **(B)** $\frac{K}{n} \to c^* > 0$[23]. Suppose for the moment that we are under case **(A)**. Then

$$\left|\left|U'(\hat{\Lambda} - \Lambda - H_n)U\right|\right|_S^2 \leq \left|\left|U'(\hat{\Lambda} - \Lambda - H_n)U\right|\right|_F^2$$
$$= \sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)(e_j^2(\beta_0) - \sigma_j^2(\beta_0) - T_j) + \sum_{i\in[n]} P_{ii}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)^2$$
$$\overset{(ii)}{=} o(K) + \sum_{i\in[n]} P_{ii}^2(e_i^2(\beta_0) - \sigma_i^2(\beta_0) - T_i)^2 \overset{(iii)}{=} o(K)$$

[23] Note that **(B)** should really be for some sub-sequence $\frac{K}{n}$ rather than the full sequence. However, we can always assume W.L.O.G that **(B)** holds for the full sequence since the result of Lemma B.3 is provided for some sub-sequence.

where (ii) follows from (B.16) and (iii) follows from

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}P_{ii}^2(e_i^2(\beta_0)-\sigma_i^2(\beta_0)-T_i)^2\right)\leq C\frac{1}{K}\sum_{i\in[n]}P_{ii}^2\leq Cp_n\frac{1}{K}\sum_{i\in[n]}P_{ii}=Cp_n=o(1)$$

since $p_n\leq \overline{C}\frac{K}{n}=o(1)$ under case **(A)**, together with assumption 3. Therefore, by Lemma B.1 we have

$$\frac{2||U'(\hat{\Lambda}-\Lambda-H_n)U||_S^2}{(\sum_{i\in[n]}P_{ii}e_i^2(\beta_0))^2}=o(K^{-1}) \tag{B.23}$$

so that combining (B.19), (B.20), (B.21),(B.22) and (B.23) yields

$$\max_i \widetilde{w}_{i,n}^2\leq 4\max_i(\widetilde{w}_{i,n}-w_{i,n}-\lambda_{i,n}^H)^2+4\max_i w_{i,n}^2+4\max_i(\lambda_{i,n}^H)^2=o(K^{-1})$$

which proves (B.18).

Next, suppose we are now under case **(B)**. Denote $\widehat{\Lambda}:=diag(e_1^2+\Delta^2v_1^2+2\Delta e_1v_1,...,e_n^2+\Delta^2v_n^2+2\Delta e_nv_n)$ and $\Lambda^\dagger:=2diag(\Delta\Pi_1e_1+\Delta^2\Pi_1v_1,...,\Delta\pi_ne_n+\Delta^2\Pi_nv_n)$. Then

$$||U'(\hat{\Lambda}-\Lambda-H_n)U||_S^2=||U'(\widehat{\Lambda}-\Lambda+\Lambda^\dagger)U||_2^s\leq 2||U'(\widehat{\Lambda}-\Lambda)U||_S^2+2||U'\Lambda^\dagger U||_S^2 \tag{B.24}$$

We first show that the preceding equation is $o(K)$. To begin, observe that

$$\begin{aligned}||U'\Lambda^\dagger U||_S^2\leq||U'\Lambda^\dagger U||_F^2&=4\sum_{i,j\in[n]}P_{ij}^2(\Delta\Pi_ie_i+\Delta^2\Pi_iv_i)(\Delta\Pi_je_j+\Delta^2\Pi_jv_j)\\&=4\sum_{i,j\in[n]}P_{ij}^2(\Delta^2\Pi_i\Pi_je_ie_j+2\Delta^3\Pi_i\Pi_je_iv_j+\Delta^4\Pi_i\Pi_jv_iv_j)\end{aligned} \tag{B.25}$$

Furthermore,

$$\sum_{i,j\in[n]}P_{ij}^2\Pi_i\Pi_je_ie_j=\sum_{i,j\in[n]}P_{ij}^2\Pi_i\Pi_j\left(\widetilde{e}_i\widetilde{e}_j-2\widetilde{e}_j(P_i^W)'\widetilde{e}+(P_i^W)'\widetilde{e}(P_j^W)'\widetilde{e}\right)=o(K) \tag{B.26}$$

where the last equality follows from

$$(a)\quad \mathbb{E}\left(\frac{1}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i\Pi_j\widetilde{e}_i\widetilde{e}_j\right)^2\leq\frac{C}{K^2}\sum_{i,j\in[n]}P_{ij}^4+\frac{C}{K^2}\sum_{i\in[n]}P_{ii}^4\leq C\frac{p_n}{K}=o(1)$$

$$(b)\quad \mathbb{E}\left(\frac{1}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i\Pi_j\widetilde{e}_j(P_i^W)'\widetilde{e}\right)^2\leq\frac{C}{K^2}\sum_{i,j,i',j'\in[n]}P_{ij}^2P_{i'j'}^2|P_{ij}^WP_{i'j'}^W+P_{ij'}^WP_{i'j}^W|\leq Cp_n^W=o(1)$$

$$(c)\quad \mathbb{E}\left|\frac{1}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i\Pi_j(P_i^W)'\widetilde{e}(P_j^W)'\widetilde{e}\right|\overset{(i)}{\leq}\frac{1}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i^2\mathbb{E}((P_i^W)'\widetilde{e})^2\leq\frac{C}{K}\sum_{i,j\in[n]}P_{ij}^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2$$

$$\leq Cp_n = o(1)$$

where (i) follows from $2|ab| \leq a^2 + b^2$. In the same way as we have shown (B.26), we can show that

$$\sum_{i,j \in [n]} P_{ij}^2 \Pi_i \Pi_j e_i v_j = o(K)$$

and

$$\sum_{i,j \in [n]} P_{ij}^2 \Pi_i \Pi_j v_i v_j = o(K),$$

so that by (B.25) we can conclude

$$||U' \Lambda^{\dagger} U||_S^2 = o(K). \tag{B.27}$$

Next, we will show that

$$||U'(\widehat{\Lambda} - \Lambda)U||_S^2 = o(K) \tag{B.28}$$

We can express

$$\widehat{\Lambda} = diag(e_1^2, ..., e_n^2) + \Delta^2 diag(v_1^2, ..., v_n^2) + 2\Delta diag(e_1 v_1, ..., e_n v_n) \equiv \widehat{\Lambda}_1 + \widehat{\Lambda}_2 + \widehat{\Lambda}_3$$

and

$$\Lambda = diag(\widetilde{\sigma}_1^2, ..., \widetilde{\sigma}_n^2) + \Delta^2 diag(\widetilde{\varsigma}_1^2, ..., \widetilde{\varsigma}_n^2) + 2\Delta diag(\widetilde{\gamma}_1, ..., \widetilde{\gamma}_n) \equiv \Lambda_1 + \Lambda_2 + \Lambda_3$$

Then by using $2|ab| \leq a^2 + b^2$ we have

$$||U'(\widehat{\Lambda} - \Lambda)U||_S^2 \leq 4||U'(\widehat{\Lambda}_1 - \Lambda_1)U||_S^2 + 4||U'(\widehat{\Lambda}_2 - \Lambda_2)U||_S^2 + 4||U'(\widehat{\Lambda}_3 - \Lambda_3)U||_S^2.$$

Therefore, to show (B.28) it suffices to show

$$||U'(\widehat{\Lambda}_1 - \Lambda_1)U||_S^2 = o(K), \tag{B.29}$$

since the other terms can be shown in the same way. To this end, recall that $e_i^2 = \widetilde{e}_i^2 + ((P_i^W)'\widetilde{e})^2 - 2\widetilde{e}_i(P_i^W)'\widetilde{e}$. Then define $\widehat{\Lambda}_{1,1} := diag(\widetilde{e}_1^2, ..., \widetilde{e}_n^2)$ so that

$$\begin{aligned} &||U'(\widehat{\Lambda}_1 - \Lambda_1)U||_S^2 \leq 2||\widehat{\Lambda}_{1,1} - \Lambda_1||_S^2 + 2||U'(\widehat{\Lambda}_1 - \widehat{\Lambda}_{1,1})U||_S^2 \\ &\leq 2||\widehat{\Lambda}_{1,1} - \Lambda_1||_S^2 + 2||U'(\widehat{\Lambda}_1 - \widehat{\Lambda}_{1,1})U||_F^2 = \max_i (e_i^2 - \widetilde{\sigma}_i^2)^2 + \sum_{i,j \in [n]} P_{ij}^2((P_i^W)'\widetilde{e})^2((P_j^W)'\widetilde{e})^2 \\ &+ 4 \sum_{i,j \in [n]} P_{ij}^2(\widetilde{e}_i(P_i^W)'\widetilde{e})(\widetilde{e}_j(P_j^W)'\widetilde{e}) - 4 \sum_{i,j \in [n]} P_{ij}^2 \widetilde{e}_i(P_i^W)'\widetilde{e}((P_j^W)'\widetilde{e})^2 \end{aligned} \tag{B.30}$$

By Van der Vaart and Wellner (1996)[Lemma 2.2.2] and noting the l_p-norm inequality $||f||_1 \leq ||f||_2$, defining $f := \max_i(\widetilde{e}_i^2 - \widetilde{\sigma}_i^2)^2$ we have

$$\mathbb{E}\left(\frac{1}{K}\max_i(e_i^2 - \widetilde{\sigma}_i^2)^2\right) = \frac{1}{K}||f||_1 \le \frac{1}{K}||f||_2 \le \frac{n^{1/2}}{K}\max_i\left(\mathbb{E}(e_i^2 - \widetilde{\sigma}_i^2)^4\right)^{1/2}$$
$$\le C\frac{n^{1/2}}{K} = C\frac{n^{1/2}}{K^{1/2}}\frac{1}{K^{1/2}} \le C\frac{1}{K^{1/2}} = o(1).$$

under case **(B)**. Furthermore,

$$(a)\quad \mathbb{E}\left(\sum_{i,j\in[n]} P_{ij}^2((P_i^W)'\widetilde{e})^2((P_j^W)'\widetilde{e})^2\right) \le \sum_{i,j\in[n]} P_{ij}^2\mathbb{E}((P_i^W)'\widetilde{e})^4$$
$$\le \sum_{i,j\in[n]} P_{ij}^2(\sum_{\ell\in[n]}(P_{i\ell}^W)^4 + \sum_{\ell\in[n]}\sum_{\ell'\in[n]}(P_{i\ell}^W)^2(P_{i\ell'}^W)^2) \le (p_n^W)^2K = o(K)$$

$$(b)\quad \mathbb{E}\left(\sum_{i,j\in[n]} P_{ij}^2|(\widetilde{e}_i(P_i^W)'\widetilde{e})(\widetilde{e}_j(P_j^W)'\widetilde{e})|\right) \le \sum_{i,j\in[n]} P_{ij}^2\mathbb{E}\widetilde{e}_i^2((P_i^W)'\widetilde{e})^2$$
$$\le C\sum_{i,j\in[n]} P_{ij}^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2 \le p_n^W\sum_{i,j\in[n]} P_{ij}^2 = o(K)$$

$$(c)\quad 2\mathbb{E}\left|\sum_{i,j\in[n]} P_{ij}^2\widetilde{e}_i(P_i^W)'\widetilde{e}((P_j^W)'\widetilde{e})^2\right| \le \sum_{i,j\in[n]} P_{ij}^2\mathbb{E}(\widetilde{e}_i(P_i^W)'\widetilde{e})^2 + \sum_{i,j\in[n]} P_{ij}^2\mathbb{E}((P_j^W)'\widetilde{e})^4$$

Putting everything together into (B.30) yields (B.29), which in turn yields (B.28). Combining (B.24), (B.27) and (B.28) yields

$$||U'(\hat{\Lambda} - \Lambda - H_n)U||_S^2 = o(K)$$

Combining the preceding equation with Lemma B.1, (B.19), (B.20), (B.21) and (B.22) yields

$$\max_i \widetilde{w}_{i,n}^2 \le 4\max_i(\widetilde{w}_{i,n} - w_{i,n} - \lambda_{i,n}^H)^2 + 4\max_i w_{i,n}^2 + 4\max_i(\lambda_{i,n}^H)^2 = o(K^{-1})$$

which proves (B.18) for **Case (B)**. The proof for diverging K case is complete.

□

Lemma B.4. *(Conditional distributional convergence implies unconditional distributional convergence) Suppose we have real random variables $X, X_1, X_2, X_3, ...$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider any sub-sigma-field $\mathcal{A} \subset \mathcal{F}$ such that $\mathbb{P}$-almost everywhere, for any Borel set $B \in \mathcal{B}(\mathbb{R})$ we have $\mathbb{P}(X_i \in B|\mathcal{A})(\omega) \rightsquigarrow \mathbb{P}(X \in B|\mathcal{A})(\omega)$. Then $X_i \rightsquigarrow X$.*

Proof of Lemma B.4:
We need to show that for any function $f \in C_b(\mathbb{R})$, where $C_b(\mathbb{R})$ is the set of continuous and bounded functions on $\mathbb{R}$, we can obtain

$$\mathbb{E}f(X_i) \to \mathbb{E}f(X) \tag{B.31}$$

By Dudley (2002)[Theorem 10.2.5], we can express

$$\mathbb{E}\left(f(X_i)|\mathcal{A}\right)(\omega) = \int_{\mathbb{R}} f(x)\mathbb{P}_{X_i|\mathcal{A}}(dx,\omega) \quad \forall \omega \in N_i^c \tag{B.32}$$

where N_i is the negligible set for each $i \in [n]$. Define $N := \cup_{i\in\mathbb{Z}_+} N_i$ where $\mathbb{Z}_+ := \{0,1,2,...\}$, so that ($B.32$) holds for any $\omega \in N^c$, with $\mathbb{P}N^c = 1$. For any $w \in N^c$, by our assumption we know $\mathbb{P}(X_i \in B|\mathcal{A})(\omega)$ weakly converges to $\mathbb{P}(X \in B|\mathcal{A})(\omega)$. Therefore, for every ω,

$$\int_{\mathbb{R}} f(x)\mathbb{P}_{X_i|\mathcal{A}}(dx,\omega) \to \int_{\mathbb{R}} f(x)\mathbb{P}_{X|\mathcal{A}}(dx,\omega).$$

By Dudley (2002)[Theorem 10.2.2], for every fixed ω, $\mathbb{P}_{X_i|\mathcal{A}}(dx,\omega)$ is probability measure over $x \in \mathbb{R}$. Hence, by dominated convergence Theorem and (B.32)

$$\begin{aligned}\mathbb{E}f(X_i) = \mathbb{E}\left(\mathbb{E}\left(f(X_i)|\mathcal{A}\right)(\omega)\right) &= \int_{\omega\in N^c}\int_{\mathbb{R}} f(x)\mathbb{P}_{X_i|\mathcal{A}}(dx,\omega)\mathbb{P}(d\omega)\\ &\to \int_{\omega\in N^c}\int_{\mathbb{R}} f(x)\mathbb{P}_{X|\mathcal{A}}(dx,\omega)\mathbb{P}(d\omega) = \mathbb{E}f(X)\end{aligned}$$

which proves (B.31)

□

Lemma B.5. *Assume that we do not have controls W in the data-generating process of (2.1). Fix any $\Delta \neq 0$ and let $\frac{Z'\Lambda_\Pi}{\sqrt{n}} = \Theta_K \in \mathbb{R}^{K\times n}$ such that $\Theta_K \mathbb{1}_n = \widetilde{\theta}_K \in \mathbb{R}^K$ is fixed for every fixed K, where $\Lambda_\Pi := diag(\Pi_1, ..., \Pi_n)$ and $\mathbb{1}_n \in \mathbb{R}^n$ is a vector of ones. Suppose that for every fixed K, $||Z'(\xi\xi' - \mathbb{E}\xi\xi')Z||_F = o_p(1)$ and assumption 4 holds, where $\xi_i := e_i + \Delta v_i$. Furthermore, assume that $\lambda_{min}(\Theta_K'\Theta_K) \geq C_1 > 0$, $\lambda_{max}(\Sigma_{1,K}(\Delta)) \leq C_2 < \infty$, and $||\widetilde{\theta}_K||_F^2/K < \frac{C_1}{C_2}$, where C_1, C_2 does not depend on K. Then*

$$\lim_{K\to\infty}\lim_{n\to\infty}\mathbb{P}\left((Z'e(\beta_0))'(Z'\widehat{\Lambda}(\beta_0)Z)^{-1}(Z'e(\beta_0)) > q_{1-\alpha}(\chi_K^2)\right) = 0$$

where $\widehat{\Lambda}(\beta_0) := diag(e_1^2(\beta_0), ..., e_n^2(\beta_0))$

Proof of Lemma B.5:
Fix some K. Define $J_{n,K} := (Z'e(\beta_0))'(Z'\widehat{\Lambda}(\beta_0)Z)^{-1}(Z'e(\beta_0))$ and $\Sigma_{1,K}(\Delta) := \mathbb{I}_{2K}'\Sigma(\Delta)\mathbb{I}_{2K} \in \mathbb{R}^{K\times K}$, where $\mathbb{I}_{2K} = (I_K, I_K)'$. Then $e_i(\beta_0)^2 = \xi_i^2 + \Delta^2\Pi_i^2 + 2\Delta\Pi_i\xi_i$ and $Z'e(\beta_0) = Z'\xi + \Delta\sqrt{n}\widetilde{\theta}_K$.

$$n^{-1/2}Z'e(\beta_0) \rightsquigarrow \mathcal{N}\left(\Delta\Sigma_{1,K}^{1/2}(\Delta)\widetilde{\theta}_K, \Sigma_1(\Delta)\right) \tag{B.33}$$

where the convergence follows from the Lindeberg-Feller Central-Limit-Theorem, assumption 4, $\frac{\Pi'\Pi}{n^2} = o(1)$ and $||Z'(\xi\xi' - \mathbb{E}\xi\xi')Z||_F = o_p(1)$. The Lindeberg-Feller condition can be verified by fixing any $\eta > 0$ and observing that

$$\frac{1}{n}\sum_{i\in[n]}\mathbb{E}\{||Z_i\xi||_F^2 1(||Z_i\xi||_F > \eta\sqrt{n})\} \overset{(i)}{\leq} \frac{1}{n}\sum_{i\in[n]}\sqrt{\mathbb{E}||Z_i\xi||_F^4\mathbb{P}(||Z_i\xi||_F > \eta\sqrt{n})}$$

$$\overset{(iii)}{\leq} \frac{C}{n} \sum_{i \in [n]} \frac{\mathbb{E}||Z_i \xi_i||_F^2}{\eta n} \leq \frac{C}{n} \sum_{i \in [n]} \frac{1}{\eta n} = \frac{C}{\eta n} \to 0$$

where (i) follows from the Cauchy-Schwartz inequality; (ii) follows from $\mathbb{E}||Z_i\xi_i||_F^4 \leq \max_i ||Z_i||_F^4 \mathbb{E}\xi_i^4 \leq C$; (iii) follows from Markov-inequality. Furthermore, we have

$$\frac{Z'\widehat{\Lambda}(\beta_0)Z}{n} = \Sigma_{1,K}(\Delta) + \Delta^2 \Theta_K' \Theta_K + o_p(1) \tag{B.34}$$

where the equality in the preceding equation follows from Markov inequality and

$$\mathbb{E}\left|\left|\frac{\sum_{i\in[n]} Z_i Z_i' \Pi_i \xi_i}{n}\right|\right|_F^2 = \frac{\sum_{i\in[n]} \mathbb{E}\xi_i^2 \Pi_i^2 trace(Z_i Z_i' Z_i Z_i')}{n^2} \leq \frac{C \sum_{i\in[n]} \Pi_i^2 \sup_i ||Z_i||_F^4}{n^2} \leq \frac{\Pi'\Pi}{n^2} = o(1)$$

Therefore, by (B.33) and (B.34), we have

$$\begin{aligned}
J_{n,K} &\rightsquigarrow \mathcal{Z}(\Delta\widetilde{\theta}_K)'(I_K + \Delta^2 \Sigma_1(\Delta)^{-1/2} \Theta_K' \Theta \Sigma_{1,K}(\Delta)^{-1/2})^{-1} \mathcal{Z}(\Delta\widetilde{\theta}_K) \\
&\leq \frac{\chi_K^2(\Delta^2 ||\widetilde{\theta}_K||_F^2)}{\lambda_{min}(I_K + \Delta^2 \Sigma_{1,K}(\Delta)^{-1/2} \Theta_K' \Theta_K \Sigma_{1,K}(\Delta)^{-1/2})} \\
&= \frac{\chi_K^2(\Delta^2 ||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 \lambda_{min}(\Sigma_{1,K}(\Delta)^{-1/2} \Theta_K' \Theta_K \Sigma_{1,K}(\Delta)^{-1/2})} \\
&\leq \frac{\chi_K^2(\Delta^2 ||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 \lambda_{min}(\Sigma_{1,K}(\Delta)^{-1}) \lambda_{min}(\Theta_K' \Theta_K)} \\
&= \frac{\chi_K^2(\Delta^2 ||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 \frac{\lambda_{min}(\Theta_K'\Theta_K)}{\lambda_{max}(\Sigma_{1,K}(\Delta))}} \leq \frac{\chi_K^2(\Delta^2 ||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 C_3},
\end{aligned} \tag{B.35}$$

where $C_3 > 0$ is some chosen constant such that it does not depend on K and $\frac{\lambda_{min}(\Theta_K'\Theta_K)}{\lambda_{max}(\Sigma_{1,K}(\Delta))} \geq \frac{C_1}{C_2} \geq C_3 > 0$ by assumption. Finally, note that

$$\frac{\frac{\chi_K^2(\Delta^2||\widetilde{\theta}_K||_F^2)}{K}}{1 + \Delta^2 C_3} = \frac{1 + \frac{\Delta^2||\widetilde{\theta}_K||_F^2}{K}}{1 + \Delta^2 C_3} < 1 \tag{B.36}$$

whenever $C_3 > \frac{||\widetilde{\theta}_K||_F^2}{K}$. Since $||\widetilde{\theta}_K||_F^2/K < \frac{C_1}{C_2}$, we can always find such a C_3, so that by noting $q_{1-\alpha}(\frac{\chi_K^2}{K}) \to 1$, combining with (B.35) and (B.36) yields

$$\lim_{K\to\infty} \lim_{n\to\infty} \mathbb{P}\left(J_{n,K} > q_{1-\alpha}(\chi_K^2)\right) \leq \lim_{K\to\infty} \mathbb{P}\left(\frac{\chi_K^2(\Delta^2||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 C_3} > q_{1-\alpha}(\frac{\chi_K^2}{K})\right) = \mathbb{P}(1 - \eta_1 > 1) = 0$$

for some $\eta_1 > 0$.

C Two estimators satisfying criteria (2.9)

This section provides proof for the consistency of Crudu et al. (2021) and Mikusheva and Sun (2022)'s estimators under the null, for both fixed and diverging instruments. The diverging instruments case is discussed in the aforementioned papers. We show that under some regularity conditions, consistency under the null still holds for fixed instruments.

Theorem C.0.1 (Standard estimator)**.** *Suppose Assumption 1 and 2 holds. If* $\frac{p_n \Pi'\Pi}{K} = O(1)$*, then for fixed* Δ*,*

$$\begin{aligned}\widehat{\Phi}_1^{standard}(\beta_0) &:= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2(\beta_0)e_j^2(\beta_0)\\ &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + 2\Delta^2\Pi_j^2\sigma_i^2(\beta_0) + \Delta^4\Pi_i^2\Pi_j^2) + o_p(1+\sum_{i\in[4]}\Delta^i)\\ &= \Phi_1(\beta_0) + \mathcal{D}^{standard}(\Delta) + o_p(1+\sum_{i\in[4]}\Delta^i)\end{aligned}$$

where $\Phi_1(\beta_0) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$

Theorem C.0.2 (Cross-fit estimator)**.** *Suppose Assumption 1 and 2 holds. Furthermore, assume* $p_n\frac{\Pi'\Pi}{K}$*. Then*

$$\widehat{\Phi}_1^{cf}(\beta) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2[e_i(\beta_0)M_i'e(\beta_0)][e_j(\beta_0)M_j'e(\beta_0)] = \Phi_1(\beta) + o_p(1)$$

where $M := I_n - Z(Z'Z)^{-1}Z'$ *and* $\widetilde{P}_{ij}^2 := \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$*. For fixed* $\Delta \neq 0$*, if* $p_n\frac{\Pi'M\Pi}{K} = O(1)$*, then*

$$\widehat{\Phi}_1^{cf}(\beta_0) = \Phi_1(\beta_0) + \mathcal{D}^{cf}(\Delta) + o_p(1+\sum_{i\in[4]}\Delta^i)$$

where

$$\begin{aligned}\mathcal{D}^{cf}(\Delta) = \mathbb{E}\Bigg(&\frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi V_j(\Delta)M_j'\Pi\\ &+\frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\Pi_i M_i'e(\beta_0)\Pi_j M_j'e(\beta_0) + \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)V_j(\Delta)M_j'\Pi\\ &+\frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)\Pi_j M_j'e(\beta_0) + \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi\Pi_j M_j'e(\beta_0)\Bigg)\end{aligned}$$

with $V(\Delta) := e + \Delta v$*.*

C.1 Proof of Theorem C.0.1

Noting that $e_i(\beta_0) = V_i(\Delta) + \Delta\Pi_i$ where $V_i(\Delta) := e_i + \Delta v_i$, we have

$$
\begin{aligned}
\widehat{\Phi}_1^{standard}(\beta_0) &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 (V_i^2(\Delta) + \Delta^2\Pi_i^2 + 2\Delta\Pi_i V_i(\Delta))(V_j^2(\Delta) + \Delta^2\Pi_j^2 + 2\Delta\Pi_j V_j(\Delta)) \\
&= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 V_i^2(\Delta)V_j^2(\Delta) + \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 V_i^2(\Delta)\Pi_j^2 \\
&+ \frac{8\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_j V_j(\Delta) V_i^2(\Delta) + \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i^2\Pi_j^2 \\
&+ \frac{8\Delta^3}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i^2\Pi_j V_j(\Delta) + \frac{8\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i\Pi_j V_i(\Delta)V_j(\Delta) \\
&\equiv \sum_{\ell=0}^{5} T_\ell
\end{aligned}
$$

The proof entails showing that

$$T_0 = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1 + \sum_{i\in[4]}\Delta^i) \tag{C.1}$$

$$T_1 = \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_j^2(\widetilde{\sigma}_i^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i) + o_p(1 + \Delta^3 + \Delta^4) \tag{C.2}$$

$$T_2 = o_p(1 + \Delta^2 + \Delta^3 \tag{C.3}$$

$$T_3 = \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i^2\Pi_j^2 \tag{C.4}$$

$$T_4 = o_p(1 + \Delta^3 + \Delta^4) \tag{C.5}$$

$$T_5 = o_p(1 + \Delta^2 + \Delta^3 + \Delta^4) \tag{C.6}$$

Combining (C.1)–(C.6) yields the second equation of Theorem C.0.1. By recalling that $\sigma_i^2(\beta_0) = \widetilde{\sigma}_i^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i$. Combining with

$$\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_j^2(\widetilde{\sigma}_i^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i) \leq \frac{C(\Delta^2 + \Delta^3 + \Delta^4)}{K}\sum_{i,j\in[n]} P_{ij}^2 = C(\Delta^2 + \Delta^3 + \Delta^4)$$

and

$$\frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i^2\Pi_j^2 \leq \frac{C\Delta^4}{K}\sum_{i,j\in[n]} P_{ij}^2 = C\Delta^4$$

yields the last equation of Theorem C.0.1.

Step 1: We show

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2e_j^2=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2\sigma_j^2+o_p(1) \tag{C.7}$$

By noting $e_i=(\widetilde{e}_i-\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_\ell)$, we observe

$$\begin{aligned}\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2e_j^2&=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_i^2\widetilde{e}_j^2-\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_i^2\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell\widetilde{e}_j+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_i^2(\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell)^2\\&+\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_j^2\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_\ell\widetilde{e}_i+\frac{8}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_i\widetilde{e}_\ell)(\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_j\widetilde{e}_\ell)\\&-\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_\ell\widetilde{e}_i)(\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell)^2+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_j^2(\sum_{\ell\in[ni}P_{i\ell}^W\widetilde{e}_\ell)^2\\&-\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\sum_{\ell\in[n]}P_{\ell j}^W\widetilde{e}_\ell\widetilde{e}_j)(\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_\ell)^2+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\sum_{\ell\in[n]}P_{i\ell}^W\widetilde{e}_\ell)^2(\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell)^2\\&\equiv\sum_{m=1}^{9}A_m\end{aligned}$$

We will show that $A_m=o_p(1)$ for $m=2,3,...,9$. First,

$$\begin{aligned}&\mathbb{E}\left(\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\widetilde{e}_i^2-\widetilde{\sigma}_i^2)\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell\widetilde{e}_j\right)^2\\&=\frac{16}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}\sum_{j'\neq i'}P_{ij}^2P_{i'j'}^2\sum_{\ell\in[n]}\sum_{\ell'\in[n]}P_{j\ell}^WP_{j'\ell'}^W\mathbb{E}((\widetilde{e}_i^2-\widetilde{\sigma}_i^2)(\widetilde{e}_{i'}^2-\widetilde{\sigma}_{i'}^2))\widetilde{e}_\ell\widetilde{e}_j\widetilde{e}_{\ell'}\widetilde{e}_{j'})\\&\leq\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{\ell\in[n]}P_{ij}^4(P_{j\ell}^W)^2+\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{\ell\in[n]}P_{ij}^2P_{\ell i}^2|P_{j\ell}^WP_{ij}^W|+\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{\ell\in[n]}P_{ij}^2P_{\ell j}^2|P_{j\ell}^WP_{ji}^W|\\&+\frac{C}{K^2}\sum_{i\in[n]}\sum_{\ell\in[n]}P_{ii}^2P_{\ell i}^2\leq\frac{Cp_n^Wp_n}{K}=o(1)\end{aligned}$$

implying that

$$A_2=\frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell\widetilde{e}_j+o_p(1)$$

Furthermore,

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\varsigma_i^2\sum_{\ell\in[n]}P_{j\ell}^W\widetilde{e}_\ell\widetilde{e}_j\right)^2$$

$$= \frac{1}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i'} P_{ij}^2 P_{i'j'}^2 \varsigma_i^2 \varsigma_{i'}^2 \sum_{\ell \in [n]} \sum_{\ell' \neq j'} P_{j\ell}^W P_{j'\ell'}^W \mathbb{E}(\widetilde{e}_\ell \widetilde{e}_j \widetilde{e}_{\ell'} \widetilde{e}_{j'})$$
$$\leq \frac{C}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} P_{ij}^2 P_{i'j}^2 \sum_{\ell \in [n]} (P_{j\ell}^W)^2 + \frac{C}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i'} P_{ij}^2 P_{i'j'}^2 P_{jj}^W |P_{j'j}^W|$$
$$\leq \frac{C}{K^2} p_n^W K + \frac{C}{K^2} (p_n^W)^2 K^2 = O(p_n^W) = o(1)$$

so that $A_2 = o_p(1)$. We can show that $A_4 = o_p(1)$ analogously. Next,

$$\mathbb{E} A_3 \leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \sum_{\ell \in [n]} (P_{j\ell}^W)^2 \leq C p_n^W = o(1)$$

so $A_3 = o_p(1)$. Note that $A_7 = o_p(1)$ by the same argument. Next,

$$\mathbb{E} A_9 \leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \left(\sum_{\ell, k \in [n]} \left((P_{i\ell}^W)^2 (P_{ik}^W)^2 + |P_{i\ell}^W P_{ik}^W P_{jk}^W P_{j\ell}^W| \right) \right) \leq C (p_n^W)^2 = o(1)$$

so $A_9 = o_p(1)$. By the simple inequality of $|ab| \leq \frac{1}{2} a^2 + \frac{1}{2} b^2$,

$$\mathbb{E} \left| \frac{8}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 (\sum_{\ell \in [n]} P_{i\ell}^W \widetilde{e}_i \widetilde{e}_\ell)(\sum_{\ell \in [n]} P_{j\ell}^W \widetilde{e}_j \widetilde{e}_\ell) \right|$$
$$\leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} (\sum_{\ell \in [n]} P_{i\ell}^W \widetilde{e}_i \widetilde{e}_\ell)^2 + \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} (\sum_{\ell \in [n]} P_{j\ell}^W \widetilde{e}_j \widetilde{e}_\ell)^2$$
$$\leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \widetilde{\sigma}_i^2 \mathbb{E} \left(\sum_{\ell \in [n]} P_{i\ell}^W \widetilde{e}_\ell \right)^2 \leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \sum_{\ell \in [n]} (P_{i\ell}^W)^2 \leq C p_n^W = o(1)$$

so $A_5 = o_p(1)$. Next, observe that

$$\frac{C}{K^2} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} (\sum_{\ell \in [n]} P_{j\ell}^W \widetilde{e}_j \widetilde{e}_\ell)^4 = \frac{C}{K^2} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} \widetilde{e}_j^4 \mathbb{E} (\sum_{\ell \in [n]} P_{j\ell}^W \widetilde{e}_\ell)^4$$
$$\leq \frac{C}{K^2} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \left(\sum_{\ell \in [n]} \sum_{k \in [n]} (P_{j\ell}^W)^2 (P_{jk}^W)^2 + \sum_{\ell \in [n]} (P_{j\ell}^W)^4 \right)$$
$$\leq C (p_n^W)^2$$

implying that

$$\mathbb{E} A_6^2 \leq \frac{C}{K^2} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} (\sum_{\ell \in [n]} P_{i\ell}^W \widetilde{e}_i \widetilde{e}_\ell)^2 + \frac{C}{K^2} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \mathbb{E} (\sum_{\ell \in [n]} P_{j\ell}^W \widetilde{e}_j \widetilde{e}_\ell)^4$$
$$\leq C p_n^W + C (p_n^W)^2 = o_p(1)$$

Hence $A_6 = o_p(1)$. The proof of $A_8 = o_p(1)$ is analogous. Therefore we have shown that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2 e_j^2 = A_1 + o_p(1)$$

It remains to show that

$$A_1 = \Phi_1 + o_p(1) \tag{C.8}$$

By defining $\widehat{\gamma}_e := (W'W)^{-1}W'\widetilde{e}$, we can write $e = \widetilde{e} - W\widehat{\gamma}_e$, so

$$Q_{e,e} = Q_{\widetilde{e},\widetilde{e}} - 2Q_{\widetilde{e},W\widehat{\gamma}_e} + Q_{W\widehat{\gamma}_e,W\widehat{\gamma}_e}$$

By the fact that $\lambda_{\min}(W'W/n) \geq \underline{C} > 0$, we have that $\widehat{\gamma}_e = O_p(n^{-1/2})$. We can express

$$\begin{aligned}
|Q_{W\widehat{\gamma}_e,W\widehat{\gamma}_e}| &= \left|\frac{1}{\sqrt{K}}\widehat{\gamma}_e' WPW'\widehat{\gamma}_e - \frac{1}{\sqrt{K}}\widehat{\gamma}_e'\sum_{i\in[n]} P_{ii}W_iW_i'\widehat{\gamma}_e\right| = \left|-\frac{1}{\sqrt{K}}\widehat{\gamma}_e'\sum_{i\in[n]} P_{ii}W_iW_i'\widehat{\gamma}_e\right| \\
&\leq \frac{1}{\sqrt{K}}||\widehat{\gamma}_e||_F^2\lambda_{\max}\left(\sum_{i\in[n]} P_{ii}W_iW_i'\right) \leq \frac{p_n}{\sqrt{K}}||\widehat{\gamma}_e||_F^2\lambda_{\max}\left(W'W\right) \\
&= \frac{p_n}{\sqrt{K}}O_p(n^{-1})O_p(n) = O_p(\frac{p_n}{\sqrt{K}}) = o_p(1)
\end{aligned}$$

so $Q_{W\widehat{\gamma}_e,W\widehat{\gamma}_e} = o_p(1)$. Furthermore,

$$\begin{aligned}
\mathbb{E}||\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\widetilde{e}_iW_i'||_F^2 &= \frac{1}{K}\mathbb{E}\left(\sum_{i\in[n]}\sum_{j\in[n]} P_{ii}P_{jj}\widetilde{e}_i\widetilde{e}_jW_iW_i'\right) = \frac{1}{K}trace(\sum_{i\in[n]} P_{ii}^2\widetilde{\sigma}_i^2W_iW_i') \\
&\leq C\frac{p_n^2}{K}trace(W'W) = O(\frac{p_n^2}{K}n)
\end{aligned}$$

so that

$$\begin{aligned}
Q_{\widetilde{e},W\widehat{\gamma}_e} &= \frac{1}{\sqrt{K}}\widetilde{e}'PW\widehat{\gamma}_e - \frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\widetilde{e}_iW_i'\widehat{\gamma}_e = \left(\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\widetilde{e}_iW_i'\right)\widehat{\gamma}_e \\
&= O_p(\frac{p_n}{\sqrt{K}}n^{1/2})O_p(n^{-1/2}) = o_p(1).
\end{aligned}$$

Therefore $Q_{e,e} = Q_{\widetilde{e},\widetilde{e}} + o_p(1)$, implying that $\Phi_1 = Avar(Q_{\widetilde{e},\widetilde{e}}) = \frac{2}{K}\sum_{i\in n}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2$, so we can express our requirement of showing (C.8) as

$$A_1 = \frac{2}{K}\sum_{i\in n}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2 + o_p(1) \tag{C.9}$$

instead. Express

$$
\begin{aligned}
A_1 - \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2 &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\widetilde{e}_i^2\widetilde{e}_j^2 - \widetilde{e}_i^2\widetilde{\sigma}_j^2 + \widetilde{e}_i^2\widetilde{\sigma}_j^2 - \widetilde{\sigma}_i^2\widetilde{\sigma}_j^2) \\
&= \frac{2}{K}\sum_{i\in n}\sum_{j\neq i} P_{ij}^2\widetilde{e}_i^2(\widetilde{e}_j^2 - \widetilde{\sigma}_j^2) + \frac{2}{K}\sum_{i\in n}\sum_{j\neq i} P_{ij}^2(\widetilde{e}_i^2\widetilde{\sigma}_j^2 - \widetilde{\sigma}_i^2\widetilde{\sigma}_j^2) = B_1 + B_2
\end{aligned}
$$

and note that

$$
B_1 \overset{(i)}{=} \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2(\widetilde{e}_j^2 - \widetilde{\sigma}_j^2) + o_p(1) \overset{(ii)}{=} o_p(1)
$$

where (i) follows from

$$
\begin{aligned}
&\mathbb{E}\left(B_1 - \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2(\widetilde{e}_j^2 - \widetilde{\sigma}_j^2)\right)^2 = \frac{2}{K^2}\sum_{i,i'\in[n]}\sum_{\substack{j\neq i\\ j'\neq i'}} P_{ij}^2P_{i'j'}^2\mathbb{E}\left((\widetilde{e}_i^2 - \widetilde{\sigma}_i^2)(\widetilde{e}_j^2 - \widetilde{\sigma}_j^2)(\widetilde{e}_{i'}^2 - \widetilde{\sigma}_{i'}^2)(\widetilde{e}_{j'}^2 - \widetilde{\sigma}_{j'}^2)\right) \\
&\leq \frac{C}{K^2}\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^4 \leq \frac{Cp_n^2}{K} = o(1)
\end{aligned}
$$

and (ii) follows from

$$
\mathbb{E}\left(\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2(\widetilde{e}_j^2 - \widetilde{\sigma}_j^2)\right)^2 \leq \frac{C}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i} P_{ij}^2P_{i'j}^2 \leq \frac{Cp_n}{K} = o(1).
$$

The proof of $B_2 = o_p(1)$ is analogous to (ii). Hence (C.9) is shown, which proves (C.7).

Step 2: We show (C.1) In a similar way to showing (C.7) we have

$$
\begin{aligned}
\frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2v_i^2v_j^2 &= \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\varsigma_i^2\varsigma_j^2 + o_p(1+\Delta^4), \\
\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2v_ie_iv_je_j &= \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\gamma}_i\widetilde{\gamma}_j + o_p(1+\Delta^2) \\
\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2e_i^2v_j^2 &= \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\varsigma_j^2 + o_p(1+\Delta^2) \\
\frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2e_i^2v_je_j &= \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\gamma}_j + o_p(1+\Delta) \\
\frac{4\Delta^3}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2v_i^2v_je_j &= \frac{4\Delta^3}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\varsigma_i^2\widetilde{\gamma}_j + o_p(1+\Delta^3)
\end{aligned}
$$

Therefore by expression (B.1),

$$
\begin{aligned}
\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2V_i^2(\Delta)V_j^2(\Delta) &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2e_j^2 + \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2v_j^2 + \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_ie_iv_je_j \\
&+ \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2v_j^2 + \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2v_je_j + \frac{4\Delta^3}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2v_je_j \\
&= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1+\sum_{i\in[4]}\Delta^i) \qquad \text{(C.10)}
\end{aligned}
$$

Therefore (C.1) is shown

Step 3: We show (C.2). Note that we have

$$
\begin{aligned}
\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2\Pi_j^2 &= \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\Pi_j^2 + o_p(1+\Delta^2) \\
\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2\Pi_j^2 &= \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\varsigma}_i^2\Pi_j^2 + o_p(1+\Delta^2) \\
\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_iv_i\Pi_j^2 &= \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\gamma}_i\Pi_j^2 + o_p(1+\Delta^2) \qquad \text{(C.11)}
\end{aligned}
$$

To see this, for the first equation, observe that $\mathbb{E}\widetilde{e}_i\widetilde{e}_\ell\widetilde{e}_{i'}\widetilde{e}_{\ell'} \neq 0$ only if $i=\ell=i'=\ell'$ or two pairs are equal (e.g. $i=\ell$ and $i'=\ell'$). Therefore

$$
\begin{aligned}
&\mathbb{E}\left(\frac{8\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_i(P_i^W)'\widetilde{e}\Pi_j^2\right)^2 = \frac{64\Delta^4}{K^2}\sum_{i,j\neq i,i',j'\neq i',\ell,\ell'}P_{ij}^2P_{i'j'}^2\Pi_j^2\Pi_{j'}^2P_{i\ell}^WP_{i'\ell'}^W\mathbb{E}\widetilde{e}_i\widetilde{e}_\ell\widetilde{e}_{i'}\widetilde{e}_{\ell'} \\
&\leq \frac{C\Delta^4}{K^2}\sum_{i,j,j'}P_{ij}^2P_{ij'}^2\Pi_j^2\Pi_{j'}^2(P_{ii}^W)^2 + \frac{C\Delta^4}{K^2}\sum_{i,i',j,j'}P_{ij}^2P_{i'j'}^2\Pi_j^2\Pi_{j'}^2P_{ii}^WP_{i'i'}^W \\
&\leq C\Delta^4(p_n^W)^2\frac{p_n\Pi'\Pi}{K^2} + C(p_n^W)^2\Delta^4\frac{p_n\Pi'\Pi}{K^2} = o_p(\Delta^4)
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
\mathbb{E}\left(\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\widetilde{e}_i^2-\widetilde{\sigma}_i^2)\Pi_j^2\right)^2 &\leq \frac{C\Delta^4}{K^2}\sum_{i,j,\ell}P_{ij}^2\Pi_j^2P_{i\ell}^2\Pi_\ell^2 \leq \frac{Cp_n\Delta^4}{K^2}\sum_{i,\ell}P_{i\ell}^2\Pi_\ell^2 \\
&= \frac{Cp_n\Delta^4}{K^2}\sum_{\ell}\Pi_\ell^2P_{\ell\ell} \leq C\Delta^4\frac{p_n}{K}\frac{p_n\Pi'\Pi}{K} = \Delta^4o(1)O(1) = o(\Delta^4),
\end{aligned}
$$

and

$$\mathbb{E}\left(\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(P_i^W)'\widetilde{e}\widetilde{e}'P_i^W\Pi_j^2\right)\leq\frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2\leq C\Delta^2p_n^W\frac{p_n\Pi'\Pi}{K}=o(\Delta^2),$$

so that by expressing $e_i=\widetilde{e}_i+(P_i^W)'\widetilde{e}$ and using Markov inequality,

$$\begin{aligned}\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2-\widetilde{\sigma}_i^2)\Pi_j^2&=\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\widetilde{e}_i^2-\widetilde{\sigma}_i^2)\Pi_j^2-\frac{8\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{e}_i(P_i^W)'\widetilde{e}\Pi_j^2\\&\quad+\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(P_i^W)'\widetilde{e}\widetilde{e}'P_i^W\Pi_j^2=o_p(1+\Delta^2).\end{aligned}$$

The second and third equation of (C.11) is shown similarly. Expressing $V_i^2(\Delta)=e_i^2+\Delta^2v_i^2+2\Delta v_ie_i$ and combining with what we just showed, we have (C.2).

Step 4: We show (C.3). We can express

$$\Pi_jV_j(\Delta)V_i^2(\Delta)=\Pi_je_jV_i^2(\Delta)+\Delta\Pi_jv_jV_i^2(\Delta)$$

Notice then that to show $T_2=o_p(1+\Delta^2+\Delta^3)$, it suffices to show $\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_je_jV_i^2(\Delta)=o_p(1+\Delta^2+\Delta^3$. However, since $V_i^2(\Delta)=e_i^2+\Delta^2v_i^2+2\Delta v_ie_i$, showing $T_2=o_p(1+\Delta^2+\Delta^3$ can be reduced to showing

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_je_je_i^2=o_p(1),\tag{C.12}$$

since the other terms are dealt is a similar manner. To begin, express $e_i^2=\widetilde{e}_i^2+(\sum_{m\in[n]}P_{im}^W\widetilde{e}_m)^2-2\widetilde{e}_i\sum_{m\in[n]}P_{im}^W\widetilde{e}_m$ so that

$$\begin{aligned}\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_je_je_i^2&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\widetilde{e}_j\widetilde{e}_i^2+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\widetilde{e}_j(\sum_{m\in[n]}P_{im}^W\widetilde{e}_m)^2\\&\quad-\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\widetilde{e}_j\sum_{m\in[n]}P_{im}^W\widetilde{e}_m\widetilde{e}_i+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\sum_{m\in[n]}P_{jm}^W\widetilde{e}_m\widetilde{e}_i^2\\&\quad+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\sum_{m\in[n]}P_{jm}^W\widetilde{e}_m(\sum_{m\in[n]}P_{im}^W\widetilde{e}_m)^2\\&\quad+\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j\sum_{m\in[n]}P_{jm}^W\widetilde{e}_m\sum_{m\in[n]}P_{im}^W\widetilde{e}_m\widetilde{e}_i\equiv\sum_{\ell=1}^{6}T_{2,\ell}\end{aligned}$$

Then $T_{2,1}=o_p(1)$ by

$$\mathbb{E}(T_{2,1})^2\leq\frac{1}{K^2}\sum_{i,i'\in[n]}\sum_{j\neq i}P_{ij}^2P_{i'j}^2\Pi_j^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_{i'}^2\widetilde{e}_j^2+\frac{1}{K^2}\sum_{i,i'\in[n]}P_{ii'}^4|\Pi_i\Pi_{i'}|\mathbb{E}\widetilde{e}_{i'}^2\widetilde{e}_i^4$$

$$\leq \frac{C}{K^2}\sum_{j\in[n]}P_{jj}^2+\frac{Cp_n^2}{K^2}\sum_{i,i'\in[n]}P_{ii'}^2\leq C\frac{p_n}{K}+C\frac{p_n^2}{K}=o(1)$$

Next, $T_{2,2}=o_p(1)$ by

$$\mathbb{E}|T_{2,2}|\leq \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2|\Pi_j|\sum_{m\in[n]}(P_{im}^W)^2\mathbb{E}|\widetilde{e}_j|\widetilde{e}_m^2\leq \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2P_{ii}^W\leq Cp_n^W=o(1).$$

Furthermore,

$$\mathbb{E}T_{2,3}^2\leq \frac{C}{K^2}\sum_{i,j,i',j'\in[n]}P_{ij}^2P_{i'j'}^2(\sum_{m\in[n]}(P_{im}^W)^2+|P_{ij}P_{i'j'}|)\leq \frac{Cp_n^W}{K^2}\sum_{i,j,i',j'\in[n]}P_{ij}^2P_{i'j'}^2=Cp_n^W=o(1)$$

so $T_{2,3}=o_p(1)$. We can repeat a similar proof to show $T_{2,4}=o_p(1)$. Next,

$$\begin{aligned}\mathbb{E}|T_{2,5}|&\leq \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j^2\mathbb{E}(\sum_{m\in[n]}P_{jm}^W\widetilde{e}_m)^2+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(\sum_{m\in[n]}P_{im}^W\widetilde{e}_m)^4\\&\leq Cp_n^W=o(1)\end{aligned}$$

so $T_{2,5}=o_p(1)$. We can show in a similar manner that $T_{2,6}=o_p(1)$. Therefore we have shown (C.12), which proves (C.3)

Step 5: We prove (C.5). Since $V_i(\Delta)=e_i+\Delta v_i$, it suffices to prove

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\Pi_je_j=o_p(1),$$

which follows from $e_j=\widetilde{e}_j-(P_j^W)'\widetilde{e}$, together with

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\Pi_j\widetilde{e}_j\right)^2\leq \frac{C}{K^2}\sum_{i,i',j\in[n]}P_{ij}^2P_{i'j}^2\leq \frac{Cp_n}{K}=o(1)$$

and

$$\begin{aligned}\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\Pi_j(P_j^W)'\widetilde{e}\right)^2&\leq \frac{C}{K^2}\sum_{i,j,i',j'}P_{ij}^2P_{i'j'}^2\sum_{\ell\in[n]}|P_{j\ell}^WP_{j'\ell}^W|\\&\leq \frac{C}{K^2}\sum_{i,j,i',j'}P_{ij}^2P_{i'j'}^2\sum_{\ell\in[n]}(P_{j\ell}^W)^2\sum_{\ell\in[n]}(P_{j'\ell}^W)^2\\&=\frac{C}{K^2}\sum_{i,j,i',j'}P_{ij}^2P_{i'j'}^2P_{jj}^WP_{j'j'}^W\leq C(p_n^W)^2=o(1)\end{aligned}$$

Step 6: We prove (C.6). Since $V_i(\Delta)V_j(\Delta) = e_ie_j + \Delta e_iv_j + \Delta v_ie_j + \Delta^2 v_iv_j$, it suffices to prove

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i\Pi_j e_ie_j = o_p(1)$$

We can express $e_ie_j = \widetilde{e}_i\widetilde{e}_j - \widetilde{e}_i(P_j^W)'\widetilde{e} - \widetilde{e}_j(P_i^W)'\widetilde{e} + (P_i^W)'\widetilde{e}(P_j^W)'\widetilde{e}$ and note that

$$\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i\Pi_j\widetilde{e}_i\widetilde{e}_j\right)^2 \leq \frac{C}{K^2}\sum_{i,j\in[n]} P_{ij}^4 \leq \frac{Cp_n^2}{K} = o(1)$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i\Pi_j\widetilde{e}_i(P_j^W)'\widetilde{e}\right)^2 &\leq \frac{C}{K^2}\sum_{i,j,i',j'\in[n]} P_{ij}^2P_{i'j'}^2\Big(\sum_{m\in[n]}|P_{jm}^WP_{j'm}^W| + |P_{ji'}^WP_{ij'}^W|\Big)\\
&\leq \frac{C}{K^2}\sum_{i,j,i',j'\in[n]} P_{ij}^2P_{i'j'}^2\Big(\sqrt{\sum_{m\in[n]}(P_{jm}^W)^2\sum_{m\in[n]}(P_{j'm}^W)^2} + (p_n^W)^2\Big)\\
&= \frac{C}{K^2}\sum_{i,j,i',j'\in[n]} P_{ij}^2P_{i'j'}^2\Big(\sqrt{P_{jj}^WP_{j'j'}^W} + (p_n^W)^2\Big) \leq C(p_n^W)^2 = o(1)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i\Pi_j(P_i^W)'\widetilde{e}(P_j^W)'\widetilde{e}\right)^2\\
&\leq \frac{C}{K^2}\sum_{i,j,i',j'\in[n]} P_{ij}^2P_{i'j'}^2\Big(\sum_{m\in[n]}|P_{im}^WP_{i'm}^WP_{jm}^WP_{j'm}^W| + \sum_{m,m'}|P_{im}^WP_{i'm}^WP_{im'}^WP_{i'm'}^W|\Big)\\
&\leq \frac{C(p_n^W)^2}{K^2}\sum_{i,j,i',j'\in[n]} P_{ij}^2P_{i'j'}^2 \leq C(p_n^W)^2 = o(1)
\end{aligned}$$

We have shown (C.6), and the proof is complete.

C.2 Proof of Theorem C.0.2

Observe that we can express

$$\begin{aligned}
\widehat{\Phi}_1^{cf}(\beta_0) &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(V_i(\Delta)+\Delta\Pi_i)M_i'(V(\Delta)+\Delta\Pi)(V_j(\Delta)+\Delta\Pi_j)M_j'(V(\Delta)+\Delta\Pi)\\
&= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2V_i(\Delta)M_i'V(\Delta)V_j(\Delta)M_j'V(\Delta) + \frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2V_i(\Delta)M_i'\Pi V_j(\Delta)M_j'\Pi\\
&+ \frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\Pi_iM_i'e(\beta_0)\Pi_jM_j'e(\beta_0) + \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2V_i(\Delta)M_i'V(\Delta)V_j(\Delta)M_j'\Pi
\end{aligned}$$

$$+\frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)\Pi_j M_j'e(\beta_0)+\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi\Pi_j M_j'e(\beta_0)$$

$$\equiv\sum_{\ell=0}^{5}T_\ell$$

where $V(\Delta):=e+\Delta v$. The proof entails showing

$$T_0=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)+o_p(1+\sum_{i\in[4]}\Delta^i)\tag{C.13}$$

as well as

$$T_\ell=\mathbb{E}T_\ell+o_p(1+\sum_{i\in[4]}\Delta^i)\quad\text{for}\quad\ell\in\{1,...,5\}\quad\text{and}$$
$$\sum_{\ell\in[n]}\mathbb{E}T_\ell=\mathcal{D}^{cf}(\Delta)\tag{C.14}$$

When $\Delta=0$, it is clear that $T_1=T_2=...=T_5=0$, so that the case of Theorem C.0.2 for $\Delta=0$ is shown immediately upon proving (C.13); this is shown in **Step 1** below. We can therefore focus on the case of $\Delta\neq 0$.

Step 1: We prove (C.13):

Sub-step 1: We show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2[e_iM_i'e][e_jM_j'e]=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2+o_p(1)\tag{C.15}$$

Express

$$e_iM_i'e=\widetilde{e}_iM_i'\widetilde{e}-\widetilde{e}_i(P_i^W)'\widetilde{e}-(P_i^W)'\widetilde{e}M_i'\widetilde{e}+((P_i^W)'\widetilde{e})^2\equiv\sum_{\ell=1}^{4}A_{i,\ell}$$

Therefore

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2[e_iM_i'e][e_jM_j'e]=\frac{2}{K}\sum_{\ell=1}^{4}\sum_{\ell'=1}^{4}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,\ell}A_{j,\ell'}$$

We first show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,1}A_{j,1}=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2+o_p(1)\tag{C.16}$$

Define the random variable $\xi_{ij}:=\widetilde{e}_iM_i'\widetilde{e}\widetilde{e}_jM_j'\widetilde{e}-\mathbb{E}(\widetilde{e}_iM_i'\widetilde{e}\widetilde{e}_jM_j'\widetilde{e})$ so that the mean of $\xi_{ij}=0$.

Then

$$\mathbb{E}\left(\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,1} - \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(M_{ii}M_{jj}+M_{ij}^2)\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2\right)^2 = \mathbb{E}\left(\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\xi_{ij}\right)^2$$
$$= \frac{4}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^4\mathbb{E}\xi_{ij}^2 + \frac{4}{K^2}\sum_{I_3}\widetilde{P}_{ij}^2\widetilde{P}_{ik}^2\mathbb{E}\xi_{ij}\xi_{ik} + \frac{4}{K^2}\sum_{I_4}\widetilde{P}_{ij}^2\widetilde{P}_{k\ell}^2\mathbb{E}\xi_{ij}\xi_{k\ell}$$

where I_3 is the distinct index of $\{i,j,k\}\in[n]$ and I_4 is the distinct index of $\{i,j,k,\ell\}\in[n]$. We first note that $\max_{i,j\neq i}\mathbb{E}\xi_{ij}^2\leq C$, which follows from the proof of Lemma 2 in Mikusheva and Sun (2022). Furthermore, noting that $\widetilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}\leq CP_{ij}^2$ by $M_{ii}=1-P_{ii}\geq 1-\delta>0$, we have

$$(a)\quad \frac{4}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^4\mathbb{E}\xi_{ij}^2 \leq \frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^4 \leq \frac{Cp_n^2}{K^2}\sum_{i\in[n]}P_{ii} = \frac{Cp_n^2}{K} = o(1),$$
$$(b)\quad |\frac{4}{K^2}\sum_{I_3}\widetilde{P}_{ij}^2\widetilde{P}_{ik}^2\mathbb{E}\xi_{ij}\xi_{ik}| \leq \frac{8}{K^2}\sum_{I_3}\widetilde{P}_{ij}^2\widetilde{P}_{ik}^2\mathbb{E}\xi_{ij}^2\mathbb{E}\xi_{ij}^2$$
$$\leq \frac{C}{K^2}\sum_{I_3}P_{ij}^2P_{ik}^2 \leq \frac{C}{K^2}\sum_{I_2}P_{ij}^2\sum_{k\in[n]}P_{ik}^2 \leq \frac{Cp_n}{K^2}\sum_{I_2}P_{ij}^2 \leq \frac{Cp_n}{K} = o(1)\quad \text{and}$$
$$(c)\quad \frac{4}{K^2}\sum_{I_4}\widetilde{P}_{ij}^2\widetilde{P}_{k\ell}^2\mathbb{E}\xi_{ij}\xi_{k\ell} \leq \frac{C}{K^2}\sum_{I_4}P_{ij}^2P_{k\ell}^2|\mathbb{E}\xi_{ij}\xi_{\ell k}| \leq \frac{Cp_n}{K} = o(1),$$

where the first inequality of (c) follows from the fact that since i,j,k,ℓ are distinct in I_4, the non-zero terms of $\mathbb{E}(\xi_{ij}\xi_{k\ell})$ are given in the proof of Mikusheva and Sun (2022)[Lemma 2] as

$$|\mathbb{E}\xi_{ij}\xi_{\ell k}|$$
$$\leq C|M_{ii}M_{jk}+M_{ij}M_{ik})(M_{\ell\ell}M_{jk}+M_{\ell j}M_{\ell k})| + C|(M_{jj}M_{i\ell}+M_{ij}M_{\ell j})(M_{kk}M_{i\ell}+M_{k\ell}M_{i\ell})|$$
$$+ C(M_{i\ell}M_{jk}+M_{ik}M_{\ell j})^2 + C(P_{ij}P_{k\ell}+P_{i\ell}P_{jk})^2$$

The second inequality of (c) follows from Mikusheva and Sun (2022)[Lemma S1.2]. Specifically, we have

$$\frac{1}{K^2}\sum_{i,j,k,\ell}P_{ij}^2P_{k\ell}^2|M_{ii}M_{jk}M_{\ell\ell}M_{jk}| \leq \frac{1}{K^2}\sum_{i,j,k,\ell}P_{ij}^2P_{k\ell}^2M_{jk}^2 = \frac{1}{K^2}\sum_{j,k,\ell}P_{ii}P_{k\ell}^2M_{jk}^2 \leq \frac{p_n}{K^2}\sum_{k,\ell}P_{k\ell}^2M_{kk}$$
$$\leq \frac{p_n}{K^2}\sum_{k,\ell}P_{k\ell}^2 = \frac{p_n}{K},$$

with the rest of the terms in $|\mathbb{E}\xi_{ij}\xi_{\ell k}|$ dealt in a similar manner. Therefore (C.16) is shown. It remains to show that $\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,\ell}A_{j,\ell'} = o_p(1)$ for $(\ell,\ell')\in\{1,2,3,4\}\times\{1,2,3,4\}\backslash(1,1)$. Note that

$$\mathbb{E}\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,2}^2 = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(P_i^W)'\mathbb{E}(\widetilde{e}_i^2\widetilde{e}\widetilde{e}')P_i^W = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\sum_{k\in[n]}(P_{ik}^W)^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_j^2$$

$$\leq \frac{Cp_n^W}{K} \sum_{i,j\in[n]} P_{ij}^2 = Cp_n^W = o(1)$$

so that by Markov inequality,

$$\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,2}^2 = o_p(1) \tag{C.17}$$

Next,

$$\begin{aligned}
\mathbb{E}\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,3}^2 &= \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 \sum_{k,\ell,m,p\in[n]} P_{ik}^W M_{i\ell} P_{im}^W M_{ip} \mathbb{E}(\widetilde{e}_k \widetilde{e}_\ell \widetilde{e}_m \widetilde{e}_p) \\
&\overset{(i)}{\leq} \frac{C}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \left(\sum_{k,\ell} (|P_{ik}^W M_{i\ell} P_{i\ell}^W M_{ik}| + (P_{ik}^W)^2 M_{i\ell}^2) + \sum_k (P_{ik}^W)^2 M_{ik}^2 \right) \\
&\overset{(ii)}{\leq} \frac{Cp_n^W}{K} \sum_{i,j\in[n]} P_{ij}^2 = Cp_n^W = o(1)
\end{aligned}$$

where (i) follows from the fact that the non-zero terms in $\mathbb{E}(\widetilde{e}_k \widetilde{e}_\ell \widetilde{e}_m \widetilde{e}_p)$ are when the indexes $k = \ell = m = p$, or when we have two sets of indexes such that the first two indexes equal the first set, and the next two indexes equal the second set, e.g. $k = \ell$ and $m = p$; (ii) follows from

$$\sum_{k,\ell} |P_{ik}^W M_{i\ell} P_{i\ell}^W M_{ik}| = (\sum_k |P_{ik}^W M_{ik}|)^2 \leq \sum_k (P_{ik}^W)^2 \sum_k M_{ik}^2 = P_{ii}^W M_{ii}^W \leq p_n^W.$$

Hence

$$\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,3}^2 = o_p(1) \tag{C.18}$$

Furthermore,

$$\mathbb{E}\left((P_i^W)'\widetilde{e}\right)^4 \leq C \sum_{\ell,k\in[n]} (P_{i\ell}^W)^2 (P_{ik}^W)^2 + C \sum_{\ell\in[n]} (P_{i\ell}^W)^4 \leq C(P_{ii}^W)^2 + C(p_n^W)^2 P_{ii}^W \leq Cp_n^W$$

so that

$$\mathbb{E}\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,4}^2 = \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 \mathbb{E}\left((P_i^W)'\widetilde{e}\right)^4 \leq \frac{Cp_n^W}{K} \sum_{i,j\in[n]} P_{ij}^2 = Cp_n^W = o(1),$$

implying

$$\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,4}^2 = o_p(1) \tag{C.19}$$

By the simple inequality $|ab| \le \frac{1}{2}a^2 + \frac{1}{2}b^2$,

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}A_{j,\ell'} \le \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}^2 + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{j,\ell'}^2 \tag{C.20}$$

Restricting $(\ell,\ell') \in \{2,3,4\}\times\{2,3,4\}$, by (C.17)-(C.19), using (C.20) we have

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}A_{j,\ell'} = o_p(1) \tag{C.21}$$

It remains to show that $\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}A_{j,\ell'} = o_p(1)$ for $(\ell,\ell') \in \{(1,2),(1,3),(1,4)\}$. To this end, we can repeat the argument in the proof of (C.16) to show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,2} = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 \mathbb{E}(A_{i,1}A_{j,2}) + o_p(1) = o_p(1) \tag{C.22}$$

where the last equality follows from Markov inequality and

$$\begin{aligned}
\left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 \mathbb{E}(A_{i,1}A_{j,2})\right| &= \left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 \sum_{\ell\in[n]} M_{i\ell}P_{i\ell}^W\,\mathbb{E}(\widetilde{e}_i^2\widetilde{e}_\ell^2)\right| \le \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sum_{\ell\in[n]}|M_{i\ell}P_{i\ell}^W| \\
&\overset{(i)}{\le} \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sum_{\ell\in[n]}M_{i\ell}^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2 = \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 M_{ii}P_{ii}^W \\
&\le \frac{Cp_n^W}{K}\sum_{i,j\in[n]}P_{ij}^2 = Cp_n^W = o(1)
\end{aligned}$$

where (i) follows from Cauchy-Schwartz inequality. Next, we will show

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,3} = o_p(1) \tag{C.23}$$

Fix any i. For indexes $(k,k',\ell,\ell',m,m') \in [n]^6$, define $\mathcal{J}_1$ to be the set where $k = k' = ... = m'$, so $|\mathcal{J}_1| = 1$. Define $\mathcal{J}_2$ to be the set where three indexes are equal, e.g. $k = k' = \ell$ and $\ell' = m = m'$. Define $\mathcal{J}_3$ to be the set where two indexes are equal, e.g. $k = k'$, $\ell = \ell'$, $m = m'$. Define $\mathcal{J}_4$ to be the set where three indexes and two indexes are equal, and one index equal i, e.g. $k = k' = \ell$, $\ell' = m$, $m' = i$. Note that $\{\mathcal{J}_s\}_{s=1}^4$ are not necessarily mutually exclusive in that there may be overlap. For any $i \in [n]$, the non-zero terms in $\mathbb{E}(\widetilde{e}_i^2\widetilde{e}_k\widetilde{e}_{k'}\widetilde{e}_\ell\widetilde{e}_{\ell'}\widetilde{e}_m\widetilde{e}_{m'})$ are in $\{\mathcal{J}_s\}_{s=1}^4$. Therefore, for any i,j,

$$\begin{aligned}
\mathbb{E}\widetilde{e}_i^2((M_i'\widetilde{e})((P_i^W)'\widetilde{e})(M_j'\widetilde{e}))^2 &= \sum_{k,k',\ell,\ell',m,m'} M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}\mathbb{E}(\widetilde{e}_i^2\widetilde{e}_k\widetilde{e}_{k'}\widetilde{e}_\ell\widetilde{e}_{\ell'}\widetilde{e}_m\widetilde{e}_{m'}) \\
&\le C\sum_{s=1}^4\sum_{\mathcal{J}_s}|M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}|
\end{aligned}$$

Then

$$
\begin{aligned}
(a) \quad & \sum_{\mathcal{J}_1} |M_{ik}P^W_{ik'}M_{j\ell}M_{i\ell'}P^W_{im}M_{jm'}| = \sum_k M_{ik}^2 M_{jk}^2 (P^W_{ik})^2 \le M_{ii}(p^W_n)^2 \le p^W_n \\
(b) \quad & \sum_{\mathcal{J}_2} |M_{ik}P^W_{ik'}M_{j\ell}M_{i\ell'}P^W_{im}M_{jm'}| \le C \sum_{k,\ell'} |M_{ik}P^W_{ik}M_{jk}||M_{i\ell'}P^W_{i\ell'}M_{j\ell'}| \\
& \le C(p^W_n)^2 \sum_{k,\ell'} |M_{ik}M_{jk}||M_{i\ell'}M_{j\ell'}| = Cp^W_n (\sum_k |M_{ik}M_{jk}|)^2 \\
& \overset{(i)}{\le} Cp^W_n \sum_k M_{ik}^2 \sum_k M_{jk}^2 = Cp^W_n M_{jj}M_{jj} \le Cp^W_n \\
(c) \quad & \sum_{\mathcal{J}_3} |M_{ik}P^W_{ik'}M_{j\ell}M_{i\ell'}P^W_{im}M_{jm'}| \le C \sum_{k,\ell,m} |M_{ik}P^W_{ik}M_{j\ell}M_{i\ell}P^W_{im}M_{jm}| \\
& \overset{(ii)}{\le} CM_{ii}P^W_{ii}M_{jj}M_{ii}P^W_{ii}M_{jj} \le Cp^W_n \\
(d) \quad & \sum_{\mathcal{J}_4} |M_{ik}P^W_{ik'}M_{j\ell}M_{i\ell'}P^W_{im}M_{jm'}| \le C \sum_{k,\ell'} |M_{ik}P^W_{ik}M_{jk}M_{i\ell'}P^W_{i\ell'}M_{ji}| \\
& \le C \sum_{k,\ell'} |M_{ik}P^W_{ik}M_{jk}M_{i\ell'}P^W_{i\ell'}| \le Cp^W_n \sum_k |M_{ik}M_{jk}| \sum_{\ell'} |M_{i\ell'}P^W_{i\ell'}| \\
& \overset{(iii)}{\le} Cp^W_n M_{ii}M_{jj}M_{ii}P^W_{ii} \le Cp^W_n
\end{aligned}
$$

where (i),(ii) and (iii) follows by Cauchy-Schwartz inequality. Putting (a)-(d) together we have

$$
\mathbb{E}\widetilde{e}_i^2((M_i'\widetilde{e})((P^W_i)'\widetilde{e})(M_j'\widetilde{e}))^2 \le Cp^W_n. \tag{C.24}
$$

Hence

$$
\begin{aligned}
& \mathbb{E}\left(\frac{2}{K}\sum_{i\in[n]}\sum_{j\ne i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,3}\right)^2 = \frac{4}{K^2}\sum_{i,i'}\sum_{j\ne i}\sum_{j'\ne i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2 \mathbb{E}[\widetilde{e}_i M_i'\widetilde{e}((P^W_j)'\widetilde{e}M_j'\widetilde{e})][\widetilde{e}_{i'}M_{i'}'\widetilde{e}((P^W_{j'})'\widetilde{e}M_{j'}'\widetilde{e}] \\
& \overset{(i)}{\le} \frac{2}{K^2}\sum_{i,i'}\sum_{j\ne i}\sum_{j'\ne i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2 \mathbb{E}[\widetilde{e}_i M_i'\widetilde{e}((P^W_j)'\widetilde{e}M_j'\widetilde{e})]^2 + \frac{2}{K^2}\sum_{i,i'}\sum_{j\ne i}\sum_{j'\ne i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2 \mathbb{E}[\widetilde{e}_{i'}M_{i'}'\widetilde{e}((P^W_{j'})'\widetilde{e}M_{j'}'\widetilde{e}]^2 \\
& \overset{(ii)}{\le} \frac{Cp^W_n}{K^2}\sum_{i,i'}\sum_{j\ne i}\sum_{j'\ne i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2 \le \frac{Cp^W_n}{K^2}\sum_{i,i',j,j'} P_{ij}^2 P_{i'j'}^2 = Cp^W_n = o(1)
\end{aligned}
$$

where (i) follows from $2|ab| \le a^2 + b^2$ and (ii) follows from (C.24). By Markov inequality, (C.23) is shown. Finally,

$$
\begin{aligned}
\mathbb{E}\left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\ne i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,4}\right| & \overset{(i)}{\le} \frac{1}{K}\sum_{i\in[n]}\sum_{j\ne i}\widetilde{P}_{ij}^2 \left(\mathbb{E}(\widetilde{e}_i(P^W_j)'\widetilde{e})^2 + \mathbb{E}(M_i'\widetilde{e}(P^W_j)'\widetilde{e})^2\right) \\
& = \frac{1}{K}\sum_{i\in[n]}\sum_{j\ne i}\widetilde{P}_{ij}^2 \left(\sum_{\ell\in[n]}(P^W_{j\ell})^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_\ell^2 + \mathbb{E}(M_i'\widetilde{e}(P^W_j)'\widetilde{e})^2\right) \overset{(ii)}{=} o(1)
\end{aligned}
$$

where (i) follows from $2|ab| \le a^2 + b^2$ and (ii) follows from

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\sum_{\ell\in[n]}(P_{j\ell}^W)^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_\ell^2 \le \frac{C}{K}\sum_{i,j\in[n]}P_{ij}^2P_{jj}^W \le Cp_n^W = o(1)$$

and

$$\begin{aligned}\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}(M_i'\widetilde{e}(P_j^W)'\widetilde{e})^2 &\le \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\left(\sum_{k,\ell}(M_{ik})^2(P_{j\ell}^W)^2 + \sum_k (M_{ik})^2(P_{jk}^W)^2\right)\\ &\le \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\left(M_{ii}P_{jj}^W + M_{ii}(p_n^W)^2\right)\\ &\le \frac{Cp_n^W}{K}\sum_{i,j\in[n]}P_{ij}^2 = Cp_n^W = o(1)\end{aligned}$$

Therefore

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,1}A_{j,4} = o_p(1). \tag{C.25}$$

Putting (C.16)-(C.25) yields (C.15).

Sub-step 2: In a similar way to **sub-step 1**, we can show that

$$\begin{aligned}\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2e_iM_i'ee_jM_j'v &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\gamma}_j + o_p(1)\\ \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2v_iM_i'vv_jM_j'v &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\varsigma}_i^2\widetilde{\varsigma}_j^2 + o_p(1)\\ \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2v_iM_i'ev_jM_j'e &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\gamma}_i\widetilde{\gamma}_j + o_p(1)\end{aligned} \tag{C.26}$$

By expression (B.1) we have

$$\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) = (\widetilde{\sigma}_i^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i)(\widetilde{\sigma}_j^2 + \Delta^2\widetilde{\varsigma}_j^2 2\Delta\widetilde{\gamma}_j)$$

Combining with (C.15) and (C.26) yields (C.13).

Step 2: In a similar way to **step 1**, we can show that $T_\ell = \mathbb{E}T_\ell + o_p(1 + \sum_{i\in[4]}\Delta^i)$ for $\ell \in [5]$. It remains to show that $\sum_{\ell\in[5]}\mathbb{E}T_\ell = \mathcal{D}^{cf}(\Delta)$, which reduces to showing $\mathbb{E}T_\ell$ satisfies the property of $\mathcal{D}(\Delta)$ in (2.9) for $\ell \in \{1, ..., 5\}$, in order to complete the proof of (C.14). Note first that

$$\mathbb{E}e_i^2 = \mathbb{E}(\widetilde{e}_i - (P_i^W)'\widetilde{e})^2 = \widetilde{\sigma}_i^2 + \sum_{\ell\in[n]}(P_{i\ell}^W)^2\widetilde{\sigma}_i^2 - 2P_{ii}^W\widetilde{\sigma}_i^2 \le C$$

since $\sum_{\ell\in[n]}(P_{i\ell}^W)^2 = P_{ii}^W \leq 1$, by property of a projection matrix. Similarly,

$$\mathbb{E}v_i^2 \leq C \quad \text{and} \quad \mathbb{E}v_ie_i \leq C,$$

so that

$$\mathbb{E}V_i^2(\Delta) = \mathbb{E}e_i^2 + \Delta^2\mathbb{E}v_i^2 + 2\Delta\mathbb{E}v_ie_i \leq C(1+\Delta+\Delta^2) \tag{C.27}$$

By the inequality $(a+b)^2 \leq 2a^2+2b^2$ and noting that $\widetilde{P}_{ij}^2 \leq CP_{ij}^2$, we have

$$\begin{aligned}
\mathbb{E}|T_1| &\leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}V_i^2(\Delta)(M_i'\Pi)^2 \leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}V_i^2(\Delta)(M_i'\Pi)^2 \\
&\leq \frac{C\Delta^2(1+\Delta+\Delta^2)}{K}\sum_{i\in[n]}P_{ii}(M_i'\Pi)^2 \leq \frac{C\Delta^2(1+\Delta+\Delta^2)p_n}{K}\sum_{i\in[n]}(M_i'\Pi)^2 \\
&= \frac{C\Delta^2(1+\Delta+\Delta^2)p_n}{K}\Pi'M\Pi = O\left(\Delta^2+\Delta^3+\Delta^4\right)
\end{aligned}$$

For T_2, note that

$$\mathbb{E}(M_i'V(\Delta))^2 \leq C(1+\Delta+\Delta^2) \tag{C.28}$$

To see this, it suffices to show $\mathbb{E}(M_i'e)^2 \leq C$, since the other terms in $V(\Delta)$ are dealt in a similar manner. Now, $MM^W = M^W - P$, where we recall $M = I_n - P$, $P := Z(Z'Z)^{-1}Z'$ and $M^W = I_n - W(W'W)^{-1}W'$. Hence

$$\begin{aligned}
\mathbb{E}(M_i'e)^2 &= \mathbb{E}(M_i'M^W\widetilde{e})^2 = \mathbb{E}((M_i^W)'\widetilde{e} - P_i'\widetilde{e})^2 \leq 2\mathbb{E}((M_i^W)'\widetilde{e})^2 + 2\mathbb{E}(P_i'\widetilde{e})^2) \\
&= 2\sum_{\ell\in[n]}(M_{i\ell}^W)^2\widetilde{\sigma}_\ell^2 + 2\sum_{\ell\in[n]}P_{i\ell}^2\widetilde{\sigma}_\ell^2 \leq CM_{ii}^W + CP_{ii} \leq C
\end{aligned}$$

since $M_{ii}^W, P_{ii} \leq 1$. This implies (C.28). Expressing $M_i'e(\beta_0) = M_i'V(\Delta) + \Delta M_i'\Pi$, we have

$$\begin{aligned}
\mathbb{E}|T_2| &\leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\mathbb{E}(M_i'e(\beta_0))^2 \leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\mathbb{E}((M_i'V(\Delta))^2 + \Delta^2(M_i'\Pi)^2) \\
&\leq \frac{C\Delta^2(1+\Delta+\Delta^2)}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i^2 + \frac{C\Delta^4}{K}\sum_{i,j\in[n]}P_{ij}^2(M_i'\Pi)^2 \\
&\leq \frac{C\Delta^2(1+\Delta+\Delta^2)p_n\Pi'\Pi}{K} + \frac{C\Delta^4}{K}\sum_{i\in[n]}P_{ii}(M_i'\Pi)^2 \\
&\leq \frac{C\Delta^2(1+\Delta+\Delta^2)p_n\Pi'\Pi}{K} + C\Delta^4\frac{p_n\Pi'M\Pi}{K} = O\left(\Delta^2+\Delta^3+\Delta^4\right)
\end{aligned}$$

Next, to deal with T_3 we first show that

$$\mathbb{E}V_i^2(\Delta)\cdot(M_i'V(\Delta))^2 \leq C(1+\sum_{i\in[4]}\Delta^i) \tag{C.29}$$

Since $V(\Delta) = e + \Delta v$, it suffices to prove that

$$\mathbb{E}e_i^2(M_i'e)^2 = \mathbb{E}e_i^2((M_i^W)'\widetilde{e} - P_i'\widetilde{e})^2 \leq 2\mathbb{E}e_i^2((M_i^W)'\widetilde{e})^2 + 2\mathbb{E}e_i^2(P_i'\widetilde{e})^2 \leq C$$

as the other terms are shown in a similar manner. But this follows from

$$\begin{aligned}
&\mathbb{E}e_i^2((M_i^W)'\widetilde{e})^2 = \mathbb{E}\widetilde{e}_i^2((M_i^W)'\widetilde{e})^2 + \mathbb{E}((P_i^W)'\widetilde{e})^2((M_i^W)'\widetilde{e})^2 - 2\mathbb{E}\widetilde{e}_i(P_i^W)'\widetilde{e}((M_i^W)'\widetilde{e})^2 \\
&\leq C\left(\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + \sum_{\ell\in[n]}(P_{i\ell}^W)^2\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + (\sum_{\ell\in[n]}|P_{i\ell}^W M_{i\ell}^W|)^2 + CP_{ii}^W\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + M_{ii}^W\sum_{\ell\in[n]}|P_{i\ell}^W M_{i\ell}^W|\right) \\
&\leq C\left(M_{ii}^W + P_{ii}^W M_{ii}^W + (M_{ii}^W)^2 P_{ii}^W\right) \leq C.
\end{aligned}$$

Hence (C.29) is shown. Then

$$\begin{aligned}
\mathbb{E}|T_3| &\leq \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(V_i^2(\Delta)\cdot(M_i'V(\Delta))^2 + V_j^2(\Delta)\cdot(M_j'\Pi)^2) \\
&\overset{(C.27),(C.29)}{\leq} \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(M_j'\Pi)^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i)\frac{p_n\Pi'M\Pi}{K} = O\left(\sum_{i\in[5]}(1+\frac{p_n\Pi'M\Pi}{K})\Delta^i\right) = O\left(\sum_{i\in[5]}\Delta^i\right)
\end{aligned}$$

Next,

$$\begin{aligned}
&\mathbb{E}|T_4| \leq \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}\left(V_i^2(\Delta)(M_i'V(\Delta))^2 + \Pi_j^2(M_j'e(\beta_0))^2\right) \\
&\overset{(C.29)}{\leq} \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(M_j'e(\beta_0))^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(M_j'V(\Delta) + \Delta M_j'\Pi)^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(M_j'V(\Delta))^2 + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(\Delta M_j'\Pi)^2 \\
&\overset{(C.28)}{\leq} C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(M_j'\Pi)^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i)\frac{p_n\Pi'M\Pi}{K} = O\left(\sum_{i\in[5]}\Delta^i\right)
\end{aligned}$$

Finally,

$$\mathbb{E}|T_5| \leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \mathbb{E}\left(V_i^2(\Delta)(M_i'\Pi)^2 + \Pi_j^2(M_j'e(\beta_0))^2\right)$$
$$\overset{(C.27)}{\leq} \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 + \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(M_j'e(\beta_0))^2$$
$$\overset{(i)}{\leq} C\Delta^2 + C\Delta^2\frac{p_n\Pi'M\Pi}{K} = O(\Delta^2)$$

where (i) follows in the same way as T_4 above. By Markov inequality, we have shown that $T_\ell = O_p(1)$ for $\ell \in \{1, ..., 5\}$. Therefore (C.14) is shown, and the proof is complete.

D Limit problem for fixed and diverging instruments

D.1 Limit Problem For Diverging Instruments

Define $Q_{a,b} := \frac{1}{\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}a_ib_j$. In the context of diverging K, we say that we have strong identification whenever $\overline{\mathcal{C}} := Q_{\Pi,\Pi} \to \infty$ and weak identification otherwise. Under the arguments of Chao et al. (2012) and Mikusheva and Sun (2022), one can obtain the following asymptotics for diverging K: Under both Weak and Strong Identification, for $K \to \infty$,

$$\begin{pmatrix} Q_{\widetilde{e},\widetilde{e}} \\ Q_{\widetilde{X},\widetilde{e}} \\ Q_{\widetilde{X},\widetilde{X}} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix}\right) \tag{D.1}$$

for $\mathcal{C} := Q_{\widetilde{\Pi},\widetilde{\Pi}}$, for some $(\Phi_1, \Phi_{12}, \Phi_{13}, \Psi, \tau, \Upsilon)$. We can take (D.1) as given for simplicity. Under the alternative we have the following asymptotic for our $\widehat{Q}(\beta_0)$-statistic in the case of diverging K

Theorem D.1.1 (Theorem A.1. of Lim, Wang, and Zhang (2023))**.** *Suppose Assumptions 1, 2 and (D.1) holds. Then for $K \to \infty$,*

$$Q_{e(\beta_0),e(\beta_0)} \rightsquigarrow \mathcal{N}(\Delta^2\overline{\mathcal{C}}, \Phi_1(\beta_0))$$

where $\overline{\mathcal{C}} := Q_{\Pi,\Pi}$, $\Phi_1(\beta_0) = \Delta^4\overline{\Upsilon} + 4\Delta^3\overline{\tau} + \Delta^2(4\overline{\Psi} + 2\overline{\Phi}_{13}) + 4\Delta\Phi_{12} + \Phi_1$ and

$$\overline{\Phi}_{13} = \frac{2}{K}\sum_{i=1}^n\sum_{j\neq i} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\gamma}_i + \frac{2}{K}\sum_{i=1}^n \widetilde{\gamma}_i\left(\sum_{j\neq i} P_{ij}\Pi_j\right)^2,$$

$$\overline{\Psi} \equiv \frac{1}{K}\sum_{i\neq j} P_{ij}^2\widetilde{\gamma}_i\widetilde{\gamma}_j + \frac{1}{K}\sum_{i\neq j} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\varsigma}_j^2 + \frac{1}{K}\sum_{i=1}^n(\sum_{j\neq i} P_{ij}\Pi_j)^2\widetilde{\sigma}_i^2,$$

$$\overline{\Upsilon} \equiv \frac{2}{K}\sum_{i=1}\sum_{j\neq i} P_{ij}^2\widetilde{\varsigma}_i^2\widetilde{\varsigma}_j^2 + \frac{4}{K}\sum_{i=1}\widetilde{\varsigma}_i^2\left(\sum_{j\neq i} P_{ij}\Pi_j\right)^2,$$

$$\overline{\tau} \equiv \frac{2}{K}\sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^2 \widetilde{\varsigma}_i^2 \widetilde{\gamma}_j + \frac{2}{K}\sum_{i=1}^{n}\widetilde{\gamma}_i\left(\sum_{j\neq i} P_{ij}\Pi_j\right)^2$$

Theorem D.1.2 (Diverging K asymptotics)**.** *Suppose Assumption 1, 2 and (D.1) holds. Then for $K \to \infty$,*

$$\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) \rightsquigarrow \mathcal{N}(\Delta^2\overline{\mathcal{C}}, \Phi_1(\beta_0))$$

D.2 Limit Problem For Fixed Instruments

Consider now the case of fixed K. Recall that $U := Z(Z'Z)^{-1/2} \in \mathbb{R}^{n\times K}$ so that $U'U = I_K$ and $UU' = P$. To deal with the convergence of $\widehat{Q}(\beta_0)$, we can assume that $(\widetilde{e}, \widetilde{v})$ are jointly normal by the strong approximation. Therefore we can assume

$$\begin{pmatrix} U'e \\ U'X \end{pmatrix} = \begin{pmatrix} U'\widetilde{e} \\ U'\widetilde{X} \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ U'\Pi \end{pmatrix}, \begin{pmatrix} U'\Lambda_{\widetilde{\sigma}}U & U'\Lambda_{\widetilde{\gamma}}U \\ U'\Lambda_{\widetilde{\gamma}}U & U'\Lambda_{\widetilde{v}}U \end{pmatrix}\right)$$

implying that

$$U'e(\beta_0) = U'e + \Delta U'X \stackrel{d}{=} \mathcal{N}\left(\Delta U'\Pi, U'\Lambda U\right)$$

where $\Lambda(\beta_0) = \Lambda_{\widetilde{\sigma}} + 2\Delta\Lambda_{\widetilde{\gamma}} + \Delta^2\Lambda_{\widetilde{\varsigma}}$, $\Lambda_{\widetilde{\sigma}} := diag(\widetilde{\sigma}_1^2, ..., \widetilde{\sigma}_n^2), \Lambda_{\widetilde{\gamma}} := diag(\widetilde{\gamma}_1, ..., \widetilde{\gamma}_n), \Lambda_{\widetilde{\varsigma}} := diag(\widetilde{\varsigma}_1^2, ..., \widetilde{\varsigma}_n^2)$. We use the variance estimator $e_i^2(\beta_0) := (Y_i - X_i\beta_0)^2$ to estimate $\sigma_i^2(\beta_0) \equiv \widetilde{\sigma}_i^2 + 2\Delta\widetilde{\gamma}_i + \Delta^2\widetilde{\varsigma}_i^2$.

Theorem D.2.1 (Fixed K asymptotics)**.** *Suppose Assumption 1 and 2 holds. Then for fixed K, under the null*

$$\widehat{Q}(\beta_0) \stackrel{d}{=} \sum_{i\in[K]} w_{i,n}\chi_{1,i}^2 + o_p(1)$$

where the $\chi_{1,i}^2$ are independent chi-squares with one degree-of-freedom and $D_n := diag(w_{1,n}, ..., w_{K,n})$ are the eigenvalues of $\frac{(Z'\Lambda Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}$

D.3 Proofs for Section D

D.3.1 Proof of Theorem D.1.1

By Lim et al. (2023)[Theorem A.1.], we have $Q_{e,e} = Q_{\widetilde{e},\widetilde{e}} + o_p(1)$, $Q_{X,e} = Q_{\overline{X},e} + o_p(1)$ and $Q_{X,X} = Q_{\overline{X},\overline{X}} + o_p(1)$, where $\overline{X} := \Pi + \widetilde{v}$. An application of (D.1) yields

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \overline{\mathcal{C}} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix}\right)$$

Since $Q_{e(\beta_0),e(\beta_0)} = Q_{e+\Delta X,e+\Delta X} = Q_{e,e} + \Delta^2 Q_{X,X} + 2\Delta Q_{X,e}$, then

$$Q_{e(\beta_0),e(\beta_0)} - \Delta^2\overline{\mathcal{C}} = \begin{pmatrix} 1 & 2\Delta & \Delta^2 \end{pmatrix} \begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \overline{\mathcal{C}} \end{pmatrix} \rightsquigarrow \mathcal{N}(0, \Phi_1(\beta_0))$$

D.3.2 Proof of Theorem D.1.2

We can express

$$\left(\widehat{Q}(\beta_0) - 1\right) = \frac{\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij} e_i(\beta_0) e_j(\beta_0)}{\frac{1}{K}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)} = \frac{\frac{1}{\sqrt{K}} Q_{e(\beta_0),e(\beta_0)}}{\frac{1}{K}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}.$$

By Theorem D.1.1,

$$\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) = Q_{e(\beta_0),e(\beta_0)} \rightsquigarrow \mathcal{N}(\Delta^2\overline{\mathcal{C}}, \Phi_1(\beta_0))$$

D.3.3 Proof of Theorem D.2.1

By Lemma B.1 and Theorem 1, we can obtain

$$\begin{aligned}
\widehat{Q}(\beta_0) &= \frac{e'UU'e}{\sum_{i\in[n]} P_{ii} e_i^2} = \frac{e'UU'e}{\sum_{i\in[n]} P_{ii}\sigma_i^2} \frac{\sum_{i\in[n]} P_{ii}\sigma_i^2}{\sum_{i\in[n]} P_{ii} e_i^2} \overset{d}{=} \left(\frac{\mathcal{E}'UU'\mathcal{E}}{\sum_{i\in[n]} P_{ii}\sigma_i^2} + o_p(1)\right)(1 + o_p(1)) \\
&= \mathcal{E}'Z(Z'\Lambda Z)^{-1/2}\frac{(Z'\Lambda Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\sigma_i^2}(Z'\Lambda Z)^{-1/2}Z'\mathcal{E}) + o_p(1) \\
&= \mathcal{Z}'D_n\mathcal{Z} + o_p(1)
\end{aligned}$$

where $\mathcal{Z} \sim \mathcal{N}(0, I_K)$.