5-2024

Essays on weak identification

Dennis Guo Wei LIM
Singapore Management University, dennis.lim.2019@phdecons.smu.edu.sg

ESSAYS ON WEAK IDENTIFICATION

By
DENNIS LIM GUO WEI

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment
of the Requirements for the Degree of PhD in Economics
2024

Supervisor of Dissertation

PhD in Economics, Programme Director

ESSAYS ON WEAK IDENTIFICATION

by

DENNIS LIM GUO WEI

Submitted to School of Economics in partial fulfillment of
the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Zhang Yichong (Supervisor/Chair)
Associate Professor of Economics
Singapore Management University

Li Jia
Professor of Economics
Singapore Management University

Yu Jun
Professor of Economics
Singapore Management University

Wang Wenjie
Assistant Professor of Economics
Nanyang Technological University

Singapore Management University

Abstract

This dissertation presents a comprehensive examination of inference techniques for weak instrumental variable (IV) models, crucial in addressing endogeneity and bias in econometric analyses. Comprising two interconnected chapters, the research explores innovative methodologies to enhance the reliability and robustness of IV regression estimations. Chapter 1 is concerned with maximizing the power of tests in the many weak IVs setting. This is done by introducing a novel approach that considers a linear combination of jackknife Anderson-Rubin (AR), jackknife Lagrangian multiplier (LM), and orthogonalized jackknife LM tests for inference in IV regressions with many weak instruments and heteroskedasticity. Following I. Andrews (2016), weights are adaptively chosen in a linear fashion based on a decision-theoretic rule, ensuring control of asymptotic size under weak and strong identifications. The proposed test exhibits optimal power against local alternatives, confirmed by simulations and empirical applications to Angrist and Krueger's (1991) dataset. Chapter 2 deals with inference under both fixed and diverging weak IVs simultaneously. In particular, conventional and jackknife Anderson-Rubin (AR) Tests are developed separately to conduct weak-identification-robust inference when the number of 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, two analytical and two bootstrap-based weak-identification-robust AR tests are introduced, all of which control asymptotic size whether the number of IVs is fixed or diverging - in particular, the number of instruments is allowed but not required to be greater than the sample size. Power properties of these uniformly valid AR tests under both fixed and diverging number of IVs are analysed.

Acknowledgements

Firstly, I want to thank my supervisor Yichong Zhang for his immaculate support. Everything I know about econometrics and measure theory, I learnt from him. He is one of the most brilliant, caring, and patient person I have been fortunate enough to encounter in this lifetime. I am proud that he is my advisor.

Secondly, I extend my gratitude to my co-author, Wenjie Wang, and my committee members, Jia Li and Jun Yu. They generously devoted their time to providing guidance throughout my research journey and offered invaluable support during my job search.

Thirdly, I extend my gratitude to my classmate Qiyuan Li. The doctoral journey is demanding, and there were moments when I contemplated quitting. Qiyuan has been a constant source of support, offering assistance whenever I needed it. His encouragement has been invaluable over the past five years.

Fourthly, I am grateful to my family who have given me unwavering support. Their love and encouragement have been a constant source of motivation and strength, enabling me to pursue my academic goals with confidence.

Finally, and most importantly, I want to thank my wife Si Jie, who have stood by me throughout this journey. She has had to take care of our newborn girl Ruby mostly without my help because of my commitment to the Ph.D. program, and has not once, in the slightest manner, expressed dissatisfaction. Her steadfast love has made my life more meaningful, and my journey more bearable. I would not have completed my doctorate without her.

Chapter Acknowledgements

The two chapters in my Thesis are joint work with Yichong Zhang and Wenjie Wang. I am grateful for their guidance and encouragement, without which this work would not have been possible. I am also grateful to Jia Li, Jun Yu, Liyu Dou, and all the participants at the SMU econometrics workshop for their valuable comments.

Contents

Chapter 1

A Conditional Linear Combination Test with Many Weak Instruments

1.1 Introduction

Various recent surveys in leading economics journals suggest that weak instruments remain important concerns for empirical practice. For instance, I.Andrews, Stock, and Sun (2019) survey 230 instrumental variable (IV) regressions from 17 papers published in the *American Economic Review* (AER). They find that many of the first-stage F-statistics (and non-homoskedastic generalizations) are in a range that raises such concerns, and virtually all of these papers report at least one first-stage F with a value smaller than 10. Similarly, in Lee, McCrary, Moreira, and Porter's (2022) survey of 123 AER articles involving IV regressions, 105 out of 847 specifications have first-stage Fs smaller than 10. Moreover, many IV applications involve a large number of instruments. For example, in their seminal paper, Angrist and Krueger (1991) study the effect of schooling on wages by interacting three base instruments (dummies for the quarter of birth) with state and year of birth, resulting in 180 instruments. Hansen, Hausman, and Newey (2008) show that using the 180 instruments gives tighter confidence intervals than using the base instruments even after adjusting for the effect of many instruments. In addition, as pointed out by Mikusheva and Sun (2022), in empirical papers that employ the "judge design" (e.g., see Maestas, Mullen, and Strand (2013), Sampat and Williams (2019), and Dobbie, Goldin, and Yang (2018)), the number of instruments (the number of judges) is typically proportional to the sample size, and the famous Fama-MacBeth two-pass regression in empirical asset pricing (e.g., see Fama and MacBeth (1973), Shanken (1992), and Anatolyev and Mikusheva (2022)) is equivalent to IV estimation with the number of instruments proportional to the number of assets. Similarly, Belloni, Chen, Chernozhukov, and Hansen (2012) consider an IV application involving more than one hundred instruments for the study of the effect of judicial eminent domain decisions on economic outcomes. Carrasco and Tchuente (2015) used many instruments in the estimation of the elasticity of intertemporal substitution in consumption. Furthermore, as pointed out by Goldsmith-Pinkham, Sorkin, and Swift (2020), the shift-share or Bartik instrument (e.g., see Bartik (1991) and Blanchard, Katz, Hall, and Eichengreen (1992)), which has been widely applied in many fields such as labor, public, development, macroeconomics, international trade, and finance, can be considered as a particular way of combining many instruments. For example, in the canonical setting of estimating the labor supply elasticity, the corresponding number of instruments is equal to the number of industries, which is also typically proportional to the sample size.

In this paper, following the seminal study by I.Andrews (2016), we propose a jackknife conditional linear combination (CLC) test that is robust to weak identification, many instruments, and heteroskedasticity. The proposed test also achieves efficiency under strong identification against local alternatives. The starting point of our analysis is the observation that, under strong identification, an orthogonalized jackknife Lagrangian multiplier (LM) test is the uniformly most powerful (UMP) test against local alternatives among the class of tests that are constructed based on jackknife LM and Anderson-Rubin (AR) tests and are either unbiased or invariant to sign changes. However, the orthogonalized LM test may not have good power under weak identification or against certain fixed alternatives. Therefore, we consider a linear combination of jackknife AR, jackknife LM, and orthogonalized LM tests. Specifically, we follow I.Andrews (2016) and determine the linear combination weights by minimizing the maximum power loss, which can be viewed as a maximum regret and is further calibrated based on the limit experiment of interest and a sufficient statistic for the identification strength under many instruments. Then, similar to I.Andrews (2016), we show such a jackknife CLC test is adaptive to the identification strength in the sense that (1) it achieves correct asymptotic size, (2) it is asymptotically and conditionally admissible under weak identification among certain class of tests, (3) it converges to the UMP test mentioned above under strong identification against local alternatives,[1] and (4) it has asymptotic power equal to 1 under strong identification against fixed alternatives. The properties of jackknife AR, jackknife LM, orthogonalized LM, and our CLC tests are summarized in Table 1.1. Simulations based on the limit experiment as well as calibrated data confirm the good power properties of our test. Then, we apply the new jackknife CLC test to Angrist and Krueger's (1991) dataset with the specifications of 180 and 1,530 instruments. We find that, in both specifications, our confidence intervals (CIs) are the shortest among those constructed by weak identification robust tests, namely, the jackknife AR, LM, and CLC tests, and the two-step procedure. Furthermore, our CIs are found to be even shorter than the non-robust Wald test CIs based on the jackknife IV estimator (JIVE) proposed by Angrist, Imbens, and Krueger (1999), which is in line with the theoretical result that the jackknife CLC test is adaptive to the identification strength and is efficient under strong identification.

	Weak ID, fixed alternative	Strong ID, local alternative	Strong ID, fixed alternative
Jackknife AR	Admissible	Not UMP	Power 1
Jackknife LM	Admissible	Not UMP	Power 1
Orthogonalized LM	Admissible	UMP	Non-monotonic power
CLC	Admissible	UMP	Power 1

Table 1.1: Power Comparison of the Tests

Relation to the literature. The contributions in the present paper relate to two strands of literature. First, it is related to the literature on many instruments; see, for example, Kunitomo (1980), Morimune (1983), Bekker (1994), Donald and Newey (2001), Chamberlain and Imbens (2004), Chao and Swanson (2005), Stock and Yogo (2005a), Han and Phillips (2006), D.Andrews and Stock (2007), Hansen et al. (2008), Newey and Windmeijer (2009), Anderson, Kunitomo, and Matsushita (2010), Kuersteiner and Okui (2010), Anatolyev and Gospodinov (2011), Belloni, Chernozhukov, and Hansen (2011), Okui (2011), Belloni et al. (2012), Carrasco (2012), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Hausman et al. (2012), Hansen and Kozbur (2014), Carrasco and Tchuente (2015), Wang and Kaffo (2016), Kolesár (2018), Matsushita and Otsu (2020), Sølvsten (2020), Crudu, Mellace, and

[1]We emphasize that the UMP property of our CLC test under strong identification holds within the class of sign-invariant or unbiased tests that are constructed based on jackknife AR and LM tests only. It may be possible to construct more efficient tests using test statistics besides the jackknife AR and LM. How to construct a globally optimal test under strong identification with many IVs and heteroskedastic errors is a topic that remains to be explored in future research.

Sándor (2021), and Mikusheva and Sun (2022), among others. In the context of many instruments and heteroskedasticity, Chao et al. (2012) and Hausman et al. (2012) provide standard errors for Wald-type inferences that are based on JIVE and jackknifed versions of the limited information maximum likelihood (LIML) and Fuller's (1977) estimators (HLIM and HFUL). These estimators are more robust to many instruments than the commonly used two-stage least squares (TSLS) estimator because they can correct the bias caused by the high dimension of IVs.[2] In simulations derived from the data in Angrist and Krueger (1991), which is representative of empirical labor studies with many instrument concerns, Angrist and Frandsen (2022, Section IV) show that such bias-corrected estimators outperform the TSLS that is based on the instruments selected by the least absolute shrinkage and selection operator (LASSO) introduced in Belloni et al. (2012) or the random forest-fitted first stage introduced in Athey, Tibshirani, and Wager (2019). Furthermore, under many weak moment asymptotics, Newey and Windmeijer (2009) provide new variance estimators for the jackknife GMM and the class of generalized empirical likelihood (GEL) estimators, which includes the continuous updating estimator (CUE) and EL estimator as special cases. In the linear heteroskedastic IV model, consistency and asymptotic normality of CUE require $m^2/n \to 0$ and $m^3/n \to 0$, respectively, where m and n denote the number of moment conditions and the sample size (e.g., see p.689 of Newey and Windmeijer (2009)). Such conditions are needed to simultaneously control the estimation error for all the elements of the heteroskedasticity consistent weighting matrix. Somewhat stronger rate conditions are required for other GEL estimators.

However, the Wald-type inference methods are invalid under weak identification, which occurs when the concentration parameter remains bounded as the sample size increases to infinity. In this case, all the estimators mentioned earlier become inconsistent, and there is no consistent test for the structural parameter of interest (see Section 3 of Mikusheva and Sun (2022)). For weak identification robust inference under many instruments, D.Andrews and Stock (2007) consider the AR test, the score test introduced in Kleibergen (2002),

[2]Specifically, the rate of growth of the concentration parameter, which measure the overall instrument strength, is denoted as μ_n^2. JIVE, HLIM, and HFUL remain consistent with heteroskedastic errors even when instrument weakness is such that μ_n^2 is slower than the number of instruments K, provided that $\mu_n^2/\sqrt{K} \to \infty$ as the number of observations $n \to \infty$ (Chao et al., 2012; Hausman et al., 2012). In contrast, TSLS is less robust to instrument weakness as it is shown to be consistent only under homoskedasticity if $\mu_n^2/K \to \infty$ (Chao and Swanson, 2005).

and the conditional likelihood ratio test introduced in Moreira (2003). Their IV model is homoskedastic and requires the number of instruments to diverge slower than the cube root of the sample size ($K^3/n \to 0$, where K denotes the number of instruments). Anatolyev and Gospodinov (2011) propose a modified AR test that allows for the number of instruments to be proportional to the sample size but still require homoskedastic errors. Recently, Crudu et al. (2021) and Mikusheva and Sun (2022) propose jackknifed versions of the AR test in a model with many instruments and heteroskedasticity. Both tests are robust to weak identification, but Mikusheva and Sun's (2022) jackknife AR test has better power properties due to the use of a cross-fit variance estimator. However, the jackknife AR tests may be inefficient under strong identification. To address this issue, Mikusheva and Sun (2022) also propose a new pre-test for weak identification under many instruments and apply it to form a two-stage testing procedure with a Wald test based on the JIVE introduced in Angrist et al. (1999). The JIVE-Wald test is more efficient than the jackknife AR under strong identification. Therefore, an empirical researcher can employ the jackknife AR if the pre-test suggests weak identification and the JIVE-Wald if the pre-test suggests strong identification. In addition to the jackknife AR, Matsushita and Otsu (2020) propose a jackknife LM test, which is also robust to weak identification, many instruments, and heteroskedastic errors. However, the jackknife CLC test introduced in our paper is more efficient than the jackknife AR, the jackknife LM, and the two-step test under strong identification and local alternatives, while still being robust to weak identification.

Second, our paper is related to the literature on weak identification under the framework of a fixed number of instruments or moment conditions, in which various robust inference methods are available for non-homoskedastic errors; see, for example, Stock and Wright (2000), Kleibergen (2005), D.Andrews and Cheng (2012), I.Andrews (2016), I.Andrews and Mikusheva (2016), I.Andrews (2018), Moreira and Moreira (2019), D.Andrews and Guggenberger (2019), and Lee et al. (2022). In particular, our jackknife CLC test extends the work of I.Andrews (2016) to the framework with many weak instruments. I.Andrews (2016) considers the convex combination between the generalized AR statistic (S statistic) introduced by Stock and Wright (2000) and the score statistic (K statistic) introduced by Kleibergen (2005). We find that under many weak instruments, the orthogonalized jackknife LM statistic plays a role similar to the K statistic. However, the trade-off between the jackknife AR and orthogonalized LM statistics turns out to be rather different from that between the S and K statistics. As pointed out by I.Andrews (2016), in the case with a fixed number of weak instruments (or moment conditions), the K statistic picks out a particular (random) direction corresponding to the span of a conditioning statistic that measures the identification strength and restricts attention to deviations from the null along this specific direction. In contrast to the K statistic, the S statistic treats all deviations from the null equally. Therefore, the trade-off between the K and S statistics is mainly from the difference in attention to deviation directions. We find that with many weak instruments, the jackknife AR and orthogonalized LM tests do not have such difference in deviation directions. Instead, their trade-off is mostly between local and non-local alternatives. Furthermore, although the standard LM test (without orthogonalization) is not weak identification robust under I.Andrews (2016)'s framework, the jackknife LM test is under many instruments. Therefore, we consider a linear combination of jackknife AR, jackknife LM, and orthogonalized jackknife LM tests and find that the resulting CLC test has good power properties in a variety of scenarios.

Notation: We denote $\mathcal{Z}(\mu)$ as the normal random variable with unit variance and expectation μ and $[n] = \{1, 2, \cdots, n\}$. We further simplify $\mathcal{Z}(0)$ as $\mathcal{Z}$, which is just a standard normal random variable. We denote z_α as the $(1-\alpha)$ quantile of a standard normal random variable and $\mathbb{C}_\alpha(a_1, a_2; \rho)$ as the $(1-\alpha)$ quantile of random variable $a_1\mathcal{Z}_1^2 + a_2(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2)^2 + (1-a_1-a_2)\mathcal{Z}_2^2$ where $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are two independent standard normal random variables, α is the significance level, ρ is a constant in $(-1, 1)$, and a_1 and a_2 are the weights of the first and second components in the random variable. We further simplify $\mathcal{C}_{0,0;\rho}$ as $\mathbb{C}_\alpha$, which is just the $1-\alpha$ quantile of $\mathcal{Z}^2$. We let $\mathbb{C}_{\alpha,\max}(\rho) = \sup_{(a_1,a_2)\in\mathbb{A}_0}\mathbb{C}_\alpha(a_1, a_2; \rho)$, where $\mathbb{A}_0 = \{(a_1, a_2) \in [0, 1] \times [0, 1], a_1 + a_2 \leq \overline{a}\}$ for some $\overline{a} < 1$. We suppress the dependence of $\mathbb{C}_{\alpha,\max}(\rho)$ on $\overline{a}$ for simplicity of notation. The operators $\mathbb{E}^*$ and $\mathbb{P}^*$ are expectation and probability taken conditionally on data, respectively. For example, $\mathbb{E}^*1\{\mathcal{Z}^2(\hat{\mu}) \geq \mathbb{C}_\alpha\}$, in which $\hat{\mu}$ is some estimator of the expectation μ based on data, means the expectation is taken over the normal random variable by treating $\hat{\mu}$ as deterministic. We use $\rightsquigarrow$ to denote convergence in distribution, $U \stackrel{d}{=} V$ to denote that U and V share the same distribution, and $\text{maxeig}(\mathcal{V})$ and $\text{mineig}(\mathcal{V})$ to denote maximum and minimum eigenvalues of a positive semidefinite matrix $\mathcal{V}$. For two sequences of random variables U_n and V_n, we write $U_n \stackrel{d}{=} V_n + o_P(1)$ if there exist $\tilde{U}_n \stackrel{d}{=} U_n$ and $\tilde{V}_n \stackrel{d}{=} V_n$ such that $\tilde{U}_n - \tilde{V}_n = o_P(1)$.

1.2 Setup and Limit Problems

We consider the linear IV regression with a scalar outcome Y_i, a scalar endogenous variable X_i, and a $K \times 1$ vector of instruments Z_i such that

$$Y_i = X_i\beta + e_i, \quad X_i = \Pi_i + V_i, \quad \forall i \in [n], \tag{1.2.1}$$

where $\Pi_i = \mathbb{E}X_i$ and $\{Z_i\}_{i\in[n]}$ is treated as fixed, following the many-instrument literature. We let K diverge with sample size n, allowing for the case that K is of the same order of magnitude as n. We further have $\mathbb{E}V_i = 0$ by construction, and $\mathbb{E}e_i = 0$ by IV exogeneity. We allow (e_i, V_i) to be heteroskedastic across i. Also, following the literature on many instruments (e.g., Mikusheva and Sun (2022)), we assume that there are no controls included in our model as they can be partialled out from (Y_i, X_i, Z_i). We provide more discussions about the effect of partialling out the covariates after Assumption 1 below.

We are interested in testing $\beta = \beta_0$. Let $e_i(\beta_0) = Y_i - X_i\beta_0 = e_i + X_i\Delta$, where $\Delta = \beta - \beta_0$. We collect the transpose of Z_i in each row of Z, an $n \times K$ matrix of instruments, and denote $P = Z(Z^\top Z)^{-1}Z^\top$. In addition, Let $Q_{a,b} = \frac{\sum_{i\in[n]}\sum_{j\neq i} a_i P_{ij} b_j}{\sqrt{K}}$ and $\mathcal{C} = Q_{\Pi,\Pi}$. Then, as pointed out by Mikusheva and Sun (2022), the rescaled $\mathcal{C}$ is the concentration parameter that measures the strength of identification in the heteroskedastic IV model with many instruments. Specifically, the parameter β is weakly identified if $\mathcal{C}$ is bounded and strongly identified if $|\mathcal{C}| \to \infty$. We consider drifting sequence asymptotics so that all quantities are implicitly indexed by the sample size n except specified otherwise. We omit such dependence for notation simplicity.

Throughout the paper, we consider three scenarios: (1) weak identification and fixed alternatives in which $\mathcal{C} \to \widetilde{\mathcal{C}}$ for some fixed constant $\widetilde{\mathcal{C}} \in \Re$ and Δ is fixed and bounded, (2) strong identification and local alternatives in which $\mathcal{C} = \widetilde{\mathcal{C}}/d_n$, $\Delta = \widetilde{\Delta} d_n$, $\widetilde{\mathcal{C}}$ and $\widetilde{\Delta}$ are bounded constants independent of n, and $d_n \to 0$ is a deterministic sequence, and (3) strong identification and fixed alternatives in which $\mathcal{C} = \widetilde{\mathcal{C}}/d_n$ for the same $\widetilde{\mathcal{C}}$ and d_n defined in case (2) and Δ is fixed and bounded.[3] Many weak identification robust tests proposed in the

[3]If we follow the setup in Chao et al. (2012) and Hausman et al. (2012) and assume $\Pi_i = \mu_n \pi_i/\sqrt{n}$ so that $\infty > C \geq \sum_{i\in[n]}\sum_{j\neq i} \pi_i P_{ij} \pi_j / n \geq c > 0$ for some constants c, C, then $\mathcal{C} = \frac{\mu_n^2}{\sqrt{K}} \frac{\sum_{i\in[n]}\sum_{j\neq i}\pi_i P_{ij}\pi_j}{n}$, implying that $d_n = \sqrt{K}/\mu_n^2$. Then, our definition of strong identification ($d_n \to 0$) is equivalent to that defined in Chao et al. (2012) and Hausman et al. (2012) ($\mu_n^2/\sqrt{K} \to$

literature (namely, the jackknife AR tests proposed by Crudu et al. (2021) and Mikusheva and Sun (2022) and the jackknife LM test proposed by Matsushita and Otsu (2020)) depend on a subset of the following three quantities: $(Q_{e(\beta_0),e(\beta_0)}, Q_{X,e(\beta_0)}, Q_{X,X})$. Throughout the paper, we maintain the following high-level assumption.

Assumption 1. *Under both weak and strong identification, the following weak convergence holds:*

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix}\right), \tag{1.2.2}$$

for some $(\Phi_1, \Phi_{12}, \Phi_{13}, \Psi, \tau, \Upsilon)$.

Although there are no controls in the model (1.2.1), we further verify Assumption 1 in Section A.1 of the Appendix for a proper linear IV regression that includes a fixed dimension of exogenous control variables, which are then partialled out from the original outcome variable, endogenous variable, and instruments.[4]

Assumption 1 implies that,[5] under both strong and weak identification,

$$\begin{pmatrix} Q_{e(\beta_0),e(\beta_0)} - \Delta^2\mathcal{C} \\ Q_{X,e(\beta_0)} - \Delta\mathcal{C} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) & \Phi_{13}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) & \tau(\beta_0) \\ \Phi_{13}(\beta_0) & \tau(\beta_0) & \Upsilon \end{pmatrix}\right) + o_p(1), \tag{1.2.3}$$

where

$$\Phi_1(\beta_0) = \Delta^4\Upsilon + 4\Delta^3\tau + \Delta^2(4\Psi + 2\Phi_{13}) + 4\Delta\Phi_{12} + \Phi_1,$$

∞).

[4]Here, we focus on the case where the number of exogenous control variables is treated as fixed. In the case where the dimension of the exogenous variables is also large and assumed to diverge to infinity with the sample size, Chao, Swanson, and Woutersen (2023a) propose new versions of various jackknife IV estimators and show they are consistent and asymptotically normal under strong identification. We conjecture that it is possible to replace our jackknife construct (i.e. $Q_{a,b}$) by the new version and consider weak identification robust tests and their linear combinations in the same manner as studied in this paper. This is left as a topic for future research.

[5]Note that $\begin{pmatrix} Q_{e(\beta_0),e(\beta_0)} \\ Q_{X,e(\beta_0)} \\ Q_{X,X} \end{pmatrix} = \begin{pmatrix} 1 & 2\Delta & \Delta^2 \\ 0 & 1 & \Delta \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} \end{pmatrix}$.

$$
\begin{aligned}
\Phi_{12}(\beta_0) &= \Delta^3\Upsilon + 3\Delta^2\tau + \Delta(2\Psi + \Phi_{13}) + \Phi_{12}, \\
\Phi_{13}(\beta_0) &= \Delta^2\Upsilon + 2\Delta\tau + \Phi_{13}, \\
\Psi(\beta_0) &= \Delta^2\Upsilon + 2\Delta\tau + \Psi, \\
\tau(\beta_0) &= \Delta\Upsilon + \tau. \qquad (1.2.4)
\end{aligned}
$$

In particular, under strong identification, we have $Q_{X,X}d_n \xrightarrow{p} \widetilde{\mathcal{C}}$, which has a degenerate distribution. Also, under local alternatives, we have $\Delta = o(1)$ so that

$$
(\Phi_1(\beta_0), \Phi_{12}(\beta_0), \Phi_{13}(\beta_0), \Psi(\beta_0), \tau(\beta_0)) \to (\Phi_1, \Phi_{12}, \Phi_{13}, \Psi, \tau).
$$

To describe a feasible version of the test, we assume we have consistent estimates for all the variance components.

Assumption 2. *Let* $\rho(\beta_0) = \frac{\Phi_{12}(\beta_0)}{\sqrt{\Phi_1(\beta_0)\Psi(\beta_0)}}$, $\widehat{\gamma}(\beta_0) = (\widehat{\Phi}_1(\beta_0), \widehat{\Phi}_{12}(\beta_0), \widehat{\Phi}_{13}(\beta_0), \widehat{\Psi}(\beta_0), \widehat{\tau}(\beta_0), \widehat{\Upsilon}, \widehat{\rho}(\beta_0))$ *be an estimator, and* $\mathcal{B} \in \Re$ *be a compact parameter space. Then, we have* $\inf_{\beta_0 \in \mathcal{B}} \Phi_1(\beta_0) > 0$, $\inf_{\beta_0 \in \mathcal{B}} \Psi(\beta_0) > 0$, $\Upsilon > 0$, *and for* $\beta_0 \in \mathcal{B}$,

$$
||\widehat{\gamma}(\beta_0) - \gamma(\beta_0)||_2 = o_p(1),
$$

where $\gamma(\beta_0) \equiv (\Phi_1(\beta_0), \Phi_{12}(\beta_0), \Phi_{13}(\beta_0), \Psi(\beta_0), \tau(\beta_0), \Upsilon, \rho(\beta_0))$.

Several remarks on Assumption 2 are in order. First, Chao et al. (2012) propose a consistent estimator for Ψ where there is strong identification and many instruments. It is possible to compute $\widehat{\gamma}(\beta_0)$ based on Chao et al.'s (2012) estimator with their JIVE-based residuals $\hat{e}_i$ from the structural equation replaced by $e_i(\beta_0)$. Under weak identification and $\beta_0 = \beta$, Crudu et al. (2021) and Matsushita and Otsu (2021) establish the consistency of such estimators for $\Phi_1(\beta_0)$ and $\Psi(\beta_0)$, respectively. Similar arguments can be used to show the consistency of the rest of the elements in $\widehat{\gamma}(\beta_0)$ under both weak and strong identification. In addition, the consistency can be established under both local and fixed alternatives. We provide more details in Section A.2.1 in the Appendix. Second, motivated by Kline, Saggio, and Sølvsten (2020), Mikusheva and Sun (2022) propose cross-fit estimators $\widehat{\Phi}_1(\beta_0)$ and $\widehat{\Upsilon}$, which are consistent under both weak and strong identification and lead to better power properties. Following their lead, one can write down the cross-fit estimators for the rest of the elements in $\gamma(\beta_0)$ and show they are consistent.[6] We provide more details in Section A.2.2 in the Appendix. Note that both Crudu et al.'s (2021) and Mikusheva and Sun's (2022) estimators are consistent under heteroskedasticity and allow for K to be of the same order of n. Third, the consistency of $\widehat{\gamma}(\beta_0)$ over the entire parameter space under both strong and weak identifications is more than necessary and maintained mainly for simplicity of presentation. In fact, for our jackknife CLC test proposed below to control size, it suffices that $\widehat{\gamma}(\beta_0)$ and $(\widehat{\Phi}_1(\beta_0), \widehat{\Phi}_{12}(\beta_0), \widehat{\Psi}(\beta_0))$ are consistent under the null for weak and strong identifications, respectively. Furthermore, the power analyses under strong identification in Lemma 1.2.1, and subsequently, Theorems 1.4.2 and 1.4.4, only require consistency of $(\widehat{\Phi}_1(\beta_0), \widehat{\Phi}_{12}(\beta_0), \widehat{\Psi}(\beta_0))$ under local alternatives and $\widehat{\gamma}(\beta_0) = O_P(1)$ under both local and fixed alternatives.

Under this framework, Crudu et al. (2021) and Mikusheva and Sun (2022) consider the jackknife AR test

$$1\{AR(\beta_0) \geq z_\alpha\}, \quad AR(\beta_0) = \frac{Q_{e(\beta_0),e(\beta_0)}}{\widehat{\Phi}_1^{1/2}(\beta_0)}, \tag{1.2.5}$$

and Matsushita and Otsu (2020) consider the jackknife LM test

$$1\{LM^2(\beta_0) \geq \mathbb{C}_\alpha\}, \quad LM(\beta_0) = \frac{Q_{X,e(\beta_0)}}{\widehat{\Psi}^{1/2}(\beta_0)}. \tag{1.2.6}$$

Both tests are robust to weak identification, many instruments, and heteroskedasticity. Lemma 1.2.1 below characterizes the joint limit distribution of $(AR(\beta_0), LM(\beta_0))^\top$ under strong identification and local alternatives.

Lemma 1.2.1. *Suppose Assumptions 1 and 2 hold and we are under strong identification with local alternatives, that is, there exists a deterministic sequence $d_n \to 0$ such that $\mathcal{C} =$*

[6] For example, Mikusheva and Sun (2022, p.22) establish the limit of their cross-fit estimator $\widehat{\Psi}$ under weak identification and many instruments when the residual $\hat{e}_i$ from the structural equation is computed based on the JIVE estimator. We can construct $\widehat{\Psi}(\beta_0)$ by replacing $\hat{e}_i$ by $e_i(\beta_0)$. Then, the argument, as theirs with $Q_{X,e}/Q_{X,X}$ replaced by Δ, establishes that $\widehat{\Psi}(\beta_0) \xrightarrow{p} \Psi(\beta_0)$.

$\widetilde{\mathcal{C}}/d_n$ *and* $\Delta = \widetilde{\Delta} d_n$*, where* $\widetilde{\mathcal{C}}$ *and* $\widetilde{\Delta}$ *are bounded constants independent of* n*. Then, we have*

$$\begin{pmatrix} AR(\beta_0) \\ LM(\beta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{\Psi^{1/2}} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

where $\rho = \Phi_{12}/\sqrt{\Phi_1 \Psi}$*.*

Two remarks are in order. First, under strong identification, we consider local alternatives so that $\beta - \beta_0 \to 0$. This is why we have $(\Psi(\beta_0), \Phi_1(\beta_0), \Phi_{12}(\beta_0))$ converge to $(\Psi, \Phi_1, \Phi_{12})$, which are just the counterparts of $(\Psi(\beta_0), \Phi_1(\beta_0), \Phi_{12}(\beta_0))$ when β_0 is replaced by β. Second, although $AR(\beta_0)$ has zero mean, and hence, no power in this case, it is correlated with $LM(\beta_0)$. It is therefore possible to use $AR(\beta_0)$ to reduce the variance of $LM(\beta_0)$ and obtain a test that is more powerful than the LM test.

Lemma 1.2.2. *Consider the limit experiment in which researchers observe* $(\mathcal{N}_1, \mathcal{N}_2)$ *with*

$$\begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

know the value of ρ *and that* $\mathbb{E}\mathcal{N}_1 = 0$*, and want to test for* $\theta = 0$ *versus the two-sided alternative. In this case,* $1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$ *is UMP among level-*α *tests that are either invariant to sign changes or unbiased, where*

$$\mathcal{N}_2^* = (1 - \rho^2)^{-1/2}(\mathcal{N}_2 - \rho \mathcal{N}_1)$$

is the normalized residual from the projection of $\mathcal{N}_2$ *on* $\mathcal{N}_1$*.*

Let the orthogonalized jackknife LM statistic be $LM^*(\beta_0) = (1 - \widehat{\rho}(\beta_0)^2)^{-1/2}(LM(\beta_0) - \widehat{\rho}(\beta_0) AR(\beta_0))$. Then, Lemma 1.2.1 implies, under strong identification and local alternatives,

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{[(1-\rho^2)\Psi]^{1/2}} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right). \tag{1.2.7}$$

Lemma 1.2.2 with $\theta = \widetilde{\Delta}\widetilde{\mathcal{C}}\Psi^{-1/2}$ implies, in this case, that the test $1\{LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha\}$ is asymptotically strictly more powerful than the jackknife AR and LM tests based on $AR(\beta_0)$

and $LM(\beta_0)$ against local alternatives as long as $\rho \neq 0$. In addition, under strong identification and local alternatives, Mikusheva and Sun's (2022) two-step test statistic is asymptotically equivalent to $LM(\beta_0)$, and thus, is less powerful than $LM^*(\beta_0)$ too.

Next, we compare the behaviors of $AR(\beta_0)$, $LM(\beta_0)$, and $LM^*(\beta_0)$ under strong identification and fixed alternatives.

Lemma 1.2.3. *Suppose Assumption 2 holds,* $(Q_{e(\beta_0),e(\beta_0)}-\Delta^2\mathcal{C}, Q_{X,e(\beta_0)}-\Delta\mathcal{C}, Q_{X,X}-\mathcal{C})^\top = O_p(1)$*, and we are under strong identification so that* $d_n\mathcal{C} \to \widetilde{\mathcal{C}}$ *for some* $d_n \to 0$*. Then, we have, for any fixed* $\Delta \neq 0$*,*

$$d_n^2 \begin{pmatrix} AR^2(\beta_0) \\ LM^2(\beta_0) \\ LM^{*2}(\beta_0) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \Phi_1^{-1}(\beta_0)\Delta^4\widetilde{\mathcal{C}}^2 \\ \Psi^{-1}(\beta_0)\Delta^2\widetilde{\mathcal{C}}^2 \\ (1-\rho^2(\beta_0))^{-1}(\Psi^{-1/2}(\beta_0)-\rho(\beta_0)\Phi_1^{-1/2}(\beta_0)\Delta)^2\Delta^2\widetilde{\mathcal{C}}^2 \end{pmatrix}.$$

Given $d_n \to 0$ and both $\Phi_1^{-1}(\beta_0)\Delta^4\widetilde{\mathcal{C}}^2 > 0$ and $\Phi_1^{-1}(\beta_0)\Delta^2\widetilde{\mathcal{C}}^2 > 0$, $AR^2(\beta_0)$ and $LM^2(\beta_0)$ have power 1 against fixed alternatives asymptotically. By contrast, $LM^{*2}(\beta_0)$ may not have power if $\Delta = \Delta_*(\beta_0) \equiv \Phi_1^{1/2}(\beta_0)\Psi^{-1/2}(\beta_0)\rho^{-1}(\beta_0)$.

Next, we compare the performance of $AR(\beta_0)$ and $LM^*(\beta_0)$ under weak identification and fixed alternatives.

Lemma 1.2.4. *Suppose Assumptions 1 and 2 hold and we are under weak identification so that* $\mathcal{C} \to \widetilde{\mathcal{C}} \in \Re$*. Then, we have, for any fixed* $\Delta \neq 0$*,*

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \tag{1.2.8}$$

where $\rho(\beta_0) = \frac{\Phi_{12}(\beta_0)}{\sqrt{\Psi(\beta_0)\Phi_1(\beta_0)}}$ *and*

$$\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \end{pmatrix} = \begin{pmatrix} \Phi_1^{-1/2}(\beta_0)\Delta^2\widetilde{\mathcal{C}} \\ (1-\rho^2(\beta_0))^{-1/2}\Psi^{-1/2}(\beta_0)\Delta\widetilde{\mathcal{C}} - \rho(\beta_0)(1-\rho^2(\beta_0))^{-1/2}\Phi_1^{-1/2}(\beta_0)\Delta^2\widetilde{\mathcal{C}} \end{pmatrix}.$$

In particular, as $\Delta \to \infty$, *we have*

$$m_1(\Delta) \to \frac{\widetilde{\mathcal{C}}}{\Upsilon^{1/2}} \quad \textit{and} \quad m_2(\Delta) \to \frac{\widetilde{\mathcal{C}}}{\Upsilon^{1/2}} \frac{\rho_{23}}{(1-\rho_{23}^2)^{1/2}},$$

where $\rho_{23} = \frac{\tau}{(\Psi\Upsilon)^{1/2}}$ *is the correlation between* $Q_{X,e}$ *and* $Q_{X,X}$.[7]

By comparing the means of the normal limit distribution in (1.2.8), we notice that under weak identification and fixed alternatives, neither $LM^*(\beta_0)$ dominates $AR(\beta_0)$ or vice versa. We also notice from Lemma 1.2.4 that for testing distant alternatives, the power of $LM^*(\beta_0)$ is different from $AR(\beta_0)$ by a factor of $\rho_{23}/\sqrt{1-\rho_{23}^2}$, so that it will be lower when $|\rho_{23}| \leq 1/\sqrt{2}$. Under weak identification and homoskedasticity,[8] we have $\rho_{23} = \rho = \Phi_{12}/\sqrt{\Psi\Phi_1}$. Therefore, although the test $1\{LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha\}$ has a power advantage under strong identification against local alternatives, it may lack power under weak identification against distant alternatives if the degree of endogeneity is low. Furthermore, $LM^*(\beta_0)$ may not have power if $\Delta = \Delta_*(\beta_0)$.

In the current setting with many instruments, $AR(\beta_0)$ and $LM^*(\beta_0)$ play roles similar to that of Stock and Wright's (2000) S and Kleibergen's (2005) K statistics in I.Andrews's (2016) setting, respectively. In the fixed number of IVs case, the power trade-off between S and K statistics is based on the direction of deviations from the null. However, as shown in Lemma 1.2.4 (the case with weak identification and fixed alternatives), the deviations of $AR(\beta_0)$ and $LM^*(\beta_0)$ from the null do not have such a difference in direction under the many-instrument setting because $\widetilde{\mathcal{C}}$ is just a scalar. Instead, their power trade-off is between local and non-local alternatives. This is in stark contrast to the setting in I.Andrews (2016).

To achieve the advantages of $AR(\beta_0)$, $LM(\beta_0)$, and $LM^*(\beta_0)$ in all three scenarios above, we need to combine them in a way that is adaptive to the identification strength. Following I.Andrews (2016), we consider the linear combination of $AR^2(\beta_0)$, $LM^2(\beta_0)$, and $LM^{*2}(\beta_0)$. Recall that $(\mathcal{N}_1, \mathcal{N}_2^*)$ are the limits of $(AR(\beta_0), LM^*(\beta_0))$ in either strong or weak identification. See (1.2.7) and (1.2.8) for their expressions in these two cases. Then, in the limit

[7] We suppress the dependence of $m_1(\Delta)$ and $m_2(\Delta)$ on $\gamma(\beta_0)$ and $\widetilde{\mathcal{C}}$ for notation simplicity.

[8] Specifically, we say the data are homoskedastic if the covariance matrices of (e_i, V_i) are constant across i.

experiment, the linear combination test can be written as

$$\phi_{a_1,a_2,\infty} = 1\{a_1\mathcal{N}_1^2 + a_2(\tilde{\rho}\mathcal{N}_1 + (1-\tilde{\rho}^2)^{1/2}\mathcal{N}_2^*)^2 + (1-a_1-a_2)\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha(a_1,a_2;\tilde{\rho})\}, \tag{1.2.9}$$

where $(a_1, a_2) \in \mathbb{A}_0$ are the combination weights, $\mathcal{N}_1 \sim \mathcal{Z}(\theta_1)$, and $\mathcal{N}_2^* \sim \mathcal{Z}(\theta_2)$; the mean parameters θ_1 and θ_2 are defined in Lemmas 1.2.1 and 1.2.4 for strong and weak identification, respectively; and $\tilde{\rho}$ is the limit of $\widehat{\rho}(\beta_0)$.[9] Let the eigenvalue decomposition of the matrix $\begin{pmatrix} a_1 + a_2\tilde{\rho}^2 & a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} \\ a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} & 1-a_1-a_2\tilde{\rho}^2 \end{pmatrix}$ be

$$\begin{pmatrix} a_1 + a_2\tilde{\rho}^2 & a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} \\ a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} & 1-a_1-a_2\tilde{\rho}^2 \end{pmatrix} = \mathcal{U}\begin{pmatrix} \nu_1(a_1,a_2) & 0 \\ 0 & \nu_2(a_1,a_2) \end{pmatrix}\mathcal{U}^\top \tag{1.2.10}$$

where, by construction, $\nu_1(a_1, a_2) \geq \nu_2(a_1, a_2) \geq 0$ and $\mathcal{U}$ is a 2×2 unitary matrix. We highlight the dependence of eigenvalues (ν_1, ν_2) on the weights (a_1, a_2). The dependence of $\mathcal{U}$ on (a_1, a_2) is suppressed for notation simplicity. Then, we have

$$a_1\mathcal{N}_1^2 + a_2(\tilde{\rho}\mathcal{N}_1 + (1-\tilde{\rho}^2)^{1/2}\mathcal{N}_2^*)^2 + (1-a_1-a_2)\mathcal{N}_2^{*2} = \nu_1(a_1,a_2)\widetilde{\mathcal{N}}_1^2 + \nu_2(a_1,a_2)\widetilde{\mathcal{N}}_2^2$$

and $\phi_{a_1,a_2,\infty} = 1\{\nu_1(a_1,a_2)\widetilde{\mathcal{N}}_1^2 + \nu_2(a_1,a_2)\widetilde{\mathcal{N}}_2^2 \geq \mathbb{C}_\alpha(a_1,a_2;\tilde{\rho})\}$, where

$$\begin{pmatrix} \widetilde{\mathcal{N}}_1 \\ \widetilde{\mathcal{N}}_2 \end{pmatrix} = \mathcal{U}^\top \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \tag{1.2.11}$$

and $\widetilde{\mathcal{N}}_1$ and $\widetilde{\mathcal{N}}_2$ are independent normal random variables with unit variance. This implies that $\phi_{a_1,a_2,\infty}$ can be viewed as a linear combination test of two independent chi-squared random variables with one degree of freedom, and those two chi-squared random variables are obtained by properly rotating $\mathcal{N}_1$ and $\mathcal{N}_2^*$ (i.e., the limits of $AR(\beta_0)$ and $LM^*(\beta_0)$).

Theorem 1.2.1 states the key properties of $\phi_{a_1,a_2,\infty}$ under the limit experiment.

Theorem 1.2.1. *(i) Suppose we are under weak identification and fixed alternatives and let $\mathcal{N}_1 \sim \mathcal{Z}(\theta_1)$, $\mathcal{N}_2^* \sim \mathcal{Z}(\theta_2)$, and they are independent, where $\theta_1 = m_1(\Delta)$ and*

[9] Under fixed alternatives, $\tilde{\rho} = \rho(\beta_0)$; under local alternatives, $\tilde{\rho} = \rho$.

$\theta_2 = m_2(\Delta)$ as in (1.2.8). We consider the test of $H_0 : \theta_1 = \theta_2 = 0$ against $H_1 : \theta_1 \neq 0$ or $\theta_2 \neq 0$. Let Φ_α denote the class of size-α tests for $H_0 : \theta_1 = \theta_2 = 0$ constructed based on $(\widetilde{\mathcal{N}}_1^2, \widetilde{\mathcal{N}}_2^2)$ defined in (1.2.11). Then, for any $(a_1, a_2) \in \mathbb{A}_0$, $\phi_{a_1,a_2,\infty}$ defined in (1.2.9) is an admissible test within Φ_α. In addition, let $(\widetilde{\theta}_1, \widetilde{\theta}_2) = (\theta_1, \theta_2)\mathcal{U}$. If $(\widetilde{\theta}_1^2, \widetilde{\theta}_2^2) = b \cdot (\nu_1(a_1, a_2), \nu_2(a_1, a_2))$ for some positive constant b, then for any test $\phi \in \Phi_\alpha$, there exists some $\bar{b} > 0$ such that for any $0 < b < \bar{b}$, we have $\mathbb{E}\phi \leq \mathbb{E}\phi_{a_1,a_2,\infty}$.

(ii) Suppose we are under strong identification and local alternatives and

$$\begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where $\theta = \frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{\Psi^{1/2}}$. We consider the test of $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. Then, $\phi_{a_1,a_2,\infty}$ defined in (1.2.9) is UMP among the class of level-α tests that are constructed based on $(\mathcal{N}_1, \mathcal{N}_2)$ and invariant to the sign change if and only if $a_1 = 0$ and $a_2\rho = 0$. In this case, this test is also UMP among the class of unbiased level-α tests that are constructed based on $(\mathcal{N}_1, \mathcal{N}_2)$.

(iii) Suppose Assumption 2 holds, $(Q_{e(\beta_0),e(\beta_0)} - \Delta^2\mathcal{C}, Q_{X,e(\beta_0)} - \Delta\mathcal{C}, Q_{X,X} - \mathcal{C})^\top = O_p(1)$, and we are under strong identification with fixed alternatives. If $1 \geq a_{1,n} \geq \frac{\tilde{q}\Phi_1(\beta_0)}{\mathcal{C}^2\Delta_*^4(\beta_0)}$ for some constant $\tilde{q} > \mathbb{C}_{\alpha,max}(\rho(\beta_0))$ and $(a_{1,n}, a_{2,n}) \in \mathbb{A}_0$, where $\Delta_*(\beta_0) = \Phi_1^{1/2}(\beta_0)\Psi^{-1/2}(\beta_0)\rho^{-1}(\beta_0)$, then

$$1\{a_{1,n}AR^2(\beta_0) + a_{2,n}LM^2(\beta_0) + (1 - a_{1,n} - a_{2,n})LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha(a_{1,n}, a_{2,n}; \widehat{\rho}(\beta_0))\} \xrightarrow{p} 1.$$

Several remarks are in order. First, unlike the one-sided jackknife AR test proposed by Mikusheva and Sun (2022), we construct the jackknife CLC test based on $AR^2(\beta_0)$ for several reasons. First, under weak identification, when the concentration parameter $\mathcal{C}$, and thus, $m_1(\Delta)$ defined in Lemma 1.2.4 is nonnegative, the one-sided test has good power. However, even in this case, the power curves simulation in Section 1.5.1 shows that our jackknife CLC test is more powerful than the one-sided AR test in most scenarios. Second, our jackknife CLC test will have good power even when $\mathcal{C}$ is negative.[10] Third, we show below that under

[10]We note that $\mathcal{C} = \frac{\sum_{i\in[n]}\sum_{j\neq i}\Pi_i P_{ij}\Pi_j}{\sqrt{K}} = \frac{\sum_{i\in[n]}(1-P_{ii})\Pi_i^2 - \Pi^\top M\Pi}{\sqrt{K}}$, where $M = I - P$. If $\Pi^\top M\Pi$ and

strong identification and local alternatives, our jackknife CLC test converges to the UMP test $1\{\mathcal{N}_2^{*2} > \mathbb{C}_\alpha\}$ whereas both the one- and two-sided tests based on $AR(\beta_0)$ have no power, as shown in Lemma 1.2.1. Fourth, under strong identification and fixed alternatives, our jackknife CLC test has asymptotic power equal to 1, as shown in Lemma 1.2.3 and Theorem 1.4.4 below. In this case, using the one-sided jackknife AR test cannot further improve the power. Fifth, combining $LM^{*2}(\beta_0)$ with $AR^2(\beta_0)$ (and $LM^2(\beta_0)$), rather than $AR(\beta_0)$, can substantially mitigate the impact of power loss of $LM^*(\beta_0)$ at $\Delta_*(\beta_0)$, as shown in the numerical investigation in Section 1.5.

Second, Theorem 1.2.1(i) implies that $\phi_{a_1,a_2,\infty}$ is admissible among tests that are also quadratic functions of $\mathcal{N}_1$ and $\mathcal{N}_2^*$ with the same rotation $\mathcal{U}$ but different eigenvalues $(\tilde{\nu}_1, \tilde{\nu}_2)$; that is,

$$(\mathcal{N}_1, \mathcal{N}_2^*)\mathcal{U}\begin{pmatrix}\tilde{\nu}_1 & 0 \\ 0 & \tilde{\nu}_2\end{pmatrix}\mathcal{U}^\top\begin{pmatrix}\mathcal{N}_1 \\ \mathcal{N}_2^*\end{pmatrix}.$$

Specifically, in the special case with $a_2 = 0$ (i.e., we put zero weight on $LM^2(\beta_0)$), the rotation matrix $\mathcal{U} = I_2$ and $\phi_{a_1,0,\infty}$ is admissible among level-α tests based on the test statistics of the form $a_1\mathcal{N}_1^2 + (1-a_1)\mathcal{N}_2^{*2}$ for $a_1 \in [0,1]$, which is similar to the result for the linear combination of S and K statistics in I.Andrews (2016).

Third, similar to I.Andrews (2016, Theorem 2.1), Theorem 1.2.1(i) also shows that our linear combination test is optimal against certain alternatives under weak identification. Additionally, in the case with $a_2 = 0$, the power optimality result in 1.2.1(i) also carries over to $\phi_{a_1,0,\infty}$ among level-α tests of the form $a_1\mathcal{N}_1^2 + (1-a_1)\mathcal{N}_2^{*2}$ for $a_1 \in [0,1]$.

Fourth, when $a_1 = 0$ and $a_2\rho = 0$ and under strong identification and local alternatives, we have $\phi_{a_1,a_2,\infty} = 1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$, which is both the UMP invariant and unbiased test. When $\rho = 0$ and under local alternatives, $a_2\mathcal{N}_2^{*2}$ in the second and third terms of $\phi_{a_1,a_2,\infty}$ cancels out, implying that $\phi_{a_1,a_2,\infty} = 1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$ as long as $a_1 = 0$.

Fifth, we note that both the rotation matrix $\mathcal{U}$ and the eigenvalues ν_1 and ν_2 in (1.2.10) are functions of (a_1, a_2). We choose this specific parametrization so that $\phi_{a_1,a_2,\infty}$ can be written as a linear combination of $AR^2(\beta_0)$, $LM^2(\beta_0)$, and $LM^{*2}(\beta_0)$. It is possible to use

$\sum_{i\in[n]} P_{ii}\Pi_i^2$ are sufficiently large, $\mathcal{C}$ can be negative. Mikusheva and Sun (2022) further assume that $\Pi^\top M\Pi \leq \frac{C\Pi^\top\Pi}{K}$ for some constant $C > 0$, which implies that $\mathcal{C} > 0$.

alternative parametrizations to combine $AR(\beta_0)$ and $LM^*(\beta_0)$. For example, let

$$\mathcal{O}(\zeta) = \begin{pmatrix} \cos(\zeta) & -\sin(\zeta) \\ \sin(\zeta) & \cos(\zeta) \end{pmatrix}$$

be a rotation matrix with angle ζ and $\begin{pmatrix} AR^\dagger(\beta_0,\zeta) \\ LM^\dagger(\beta_0,\zeta) \end{pmatrix} = \mathcal{O}(\zeta) \begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix}$. Then, in the limit experiment, the linear combination test statistic can be written as

$$a\mathcal{N}_1^{\dagger 2} + (1-a)\mathcal{N}_2^{\dagger 2}, \tag{1.2.12}$$

where $(\mathcal{N}_1^\dagger, \mathcal{N}_2^\dagger)$ are the limits of $(AR^\dagger(\beta_0,\zeta), LM^\dagger(\beta_0,\zeta))$ under either weak or strong identification. In the following, we will use a minimax procedure to determine the optimal weights (a_1, a_2) for our jackknife CLC test $\phi_{a_1,a_2,\infty}$. Similarly, we can use this procedure to select the value of a and ζ for the new parametrization in (1.2.12). Under strong identification and local alternatives, Lemma 1.2.2 shows that the test $1\{LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha\}$ is the most powerful test against local alternatives. This is achieved by our jackknife CLC test $\phi_{a_1,a_2,\infty}$ with $a_1 = 0$ and $a_2\rho = 0$. In this case, the alternative parametrization does not bring any additional power.

1.3 A Conditional Linear Combination Test

In this section, we determine the weights (a_1, a_2) in the jackknife CLC test via a minimax procedure. Under weak identification, the limit test statistic of the jackknife CLC test with weights (a_1, a_2) is

$$\phi_{a_1,a_2,\infty} = 1\left\{ \begin{array}{c} a_1\mathcal{Z}_1^2(m_1(\Delta)) + a_2(\rho(\beta_0)\mathcal{Z}_1(m_1(\Delta)) + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2(m_2(\Delta)))^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(m_2(\Delta)) \geq \mathbb{C}_\alpha(a_1,a_2;\rho(\beta_0)) \end{array} \right\}, \tag{1.3.1}$$

where $m_1(\Delta)$ and $m_2(\Delta)$ are defined in Lemma 1.2.4, and $\mathcal{Z}_1(\cdot)$ and $\mathcal{Z}_2(\cdot)$ are independent. In this case, we can be explicit and write $\phi_{a_1,a_2,\infty} = \phi_{a_1,a_2,\infty}(\Delta)$. However, the limit power of the jackknife CLC test will typically remain unknown as the true parameter β (and hence

Δ) is unknown. To overcome this issue, we follow I.Andrews (2016) and calibrate the power, i.e, $\mathbb{E}\phi_{a_1,a_2,\infty}(\delta)$, where δ ranges over all possible values that Δ can potentially take; we define $\phi_{a_1,a_2,\infty}(\delta)$ as well as the range of potential values of Δ below.

Let $\widehat{D} = Q_{X,X} - (Q_{e(\beta_0),e(\beta_0)}, Q_{X,e(\beta_0)}) \begin{pmatrix} \widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0) \end{pmatrix}$ be the residual from the projection of $Q_{X,X}$ on $(Q_{e(\beta_0),e(\beta_0)}, Q_{X,e(\beta_0)})$. By (1.2.3), under weak identification,

$$\widehat{D} = D + o_p(1), \quad D \stackrel{d}{=} \mathcal{N}(\mu_D, \sigma_D^2),$$

where

$$\mu_D = \widetilde{\mathcal{C}} \left[1 - (\Delta^2, \Delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right] \quad \text{and}$$

$$\sigma_D^2 = \Upsilon - \left((\Phi_{13}(\beta_0), \tau(\beta_0)) \begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right).$$

We note that $\widehat{D}$ is a sufficient statistic for μ_D, which contains information about the concentration parameter $\mathcal{C}$ and is asymptotically independent of $AR(\beta_0)$, $LM(\beta_0)$, and hence $LM^*(\beta_0)$.

Under weak identification, we observe that $m_1(\Delta)$ and $m_2(\Delta)$ in Lemma 1.2.4 can be written as

$$\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \end{pmatrix} = \begin{pmatrix} C_1(\Delta) \\ C_2(\Delta) \end{pmatrix} \mu_D, \tag{1.3.2}$$

where

$$\begin{pmatrix} C_1(\Delta) \\ C_2(\Delta) \end{pmatrix} \equiv \begin{pmatrix} \Phi_1^{-1/2}(\beta_0)\Delta^2 \\ (1-\rho^2(\beta_0))^{-1/2}(\Psi^{-1/2}(\beta_0)\Delta - \rho(\beta_0)\Phi_1^{-1/2}(\beta_0)\Delta^2) \end{pmatrix}$$
$$\times \left[1 - (\Delta^2, \Delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right]^{-1}. \tag{1.3.3}$$

By (1.3.2), we see that $\phi_{a_1,a_2,\infty} = \phi_{a_1,a_2,\infty}(\Delta)$ defined in (1.2.9) can be written as

$$1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2(C_1(\Delta)\mu_D) + a_2(\rho(\beta_0)\mathcal{Z}_1(C_1(\Delta)\mu_D) + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2(C_2(\Delta)\mu_D))^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(C_2(\Delta)\mu_D) \geq \mathbb{C}_\alpha(a_1,a_2;\rho(\beta_0)) \end{array}\right\}.$$

This motivates the definition that

$$\phi_{a_1,a_2,\infty}(\delta) = 1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2(C_1(\delta)\mu_D) + a_2(\rho(\beta_0)\mathcal{Z}_1(C_1(\delta)\mu_D) + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2(C_2(\delta)\mu_D))^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(C_2(\delta)\mu_D) \geq \mathbb{C}_\alpha(a_1,a_2;\rho(\beta_0)) \end{array}\right\}. \tag{1.3.4}$$

To emphasize the dependence of $\phi_{a_1,a_2,\infty}(\delta)$ on μ_D and $\gamma(\beta_0)$, we further write $\phi_{a_1,a_2,\infty}(\delta)$ as $\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))$.

The range of values that Δ can take is defined as $\mathcal{D}(\beta_0) = \{\delta : \delta + \beta_0 \in \mathcal{B}\}$, where $\mathcal{B}$ is the parameter space. For instance, in their empirical application of returns to education, Mikusheva and Sun (2022) assume that β (i.e., the return to education) ranges from -0.5 to 0.5, with $\mathcal{B} = [-0.5, 0.5]$. We adopt the same practice in our simulations based on calibrated data in Section 1.5.2 and empirical application in Section 1.6. Specifying the parameter space is almost inevitable for any weak-identification-robust inference method, but additional simulation results in Section A.21 of the Appendix show that our method is insensitive to the choice of parameter space.

Following the lead of I.Andrews (2016), we define the highest attainable power for each $\delta \in \mathcal{D}(\beta_0)$ as $\mathcal{P}_{\delta,\mu_D} = \sup_{(a_1,a_2)\in\mathbb{A}(\mu_D,\gamma(\beta_0))} \mathbb{E}\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))$, which means that

$$\mathcal{P}_{\delta,\mu_D} - \mathbb{E}\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))$$

is the power loss when the weights are set as (a_1, a_2). Here we denote the domain of (a_1, a_2) as $\mathbb{A}(\mu_D, \gamma(\beta_0))$ and define it as $\mathbb{A}(\mu_D, \gamma(\beta_0)) = \{(a_1, a_2) \in \mathbb{A}_0, a_1 \in [\underline{a}(\mu_D, \gamma(\beta_0)), 1]\}$ where $\mathbb{A}_0 = \{(a_1, a_2) \in [0, 1] \times [0, 1], a_1 + a_2 \leq \overline{a}\}$ for some $\overline{a} < 1$,

$$\underline{a}(\mu_D, \gamma(\beta_0)) = \min\left(p_1, \frac{p_2\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}\right), \tag{1.3.5}$$

the two tuning parameters $(p_1, p_2) = (0.01, 1.1)$, $\Delta_*(\beta_0) = \Phi_1^{1/2}(\beta_0)\Psi^{-1/2}(\beta_0)\rho^{-1}(\beta_0)$ as defined after Lemma 1.2.3, and

$$c_{\mathcal{B}}(\beta_0) = \sup_{\delta \in \mathcal{D}(\beta_0)} \left[1 - (\delta^2, \delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right]^2 .$$

The maximum power loss over $\delta \in \mathcal{D}(\beta_0)$ can be viewed as a maximum regret. Then, we choose (a_1, a_2) that minimizes the maximum regret; that is,

$$(a_1(\mu_D, \gamma(\beta_0)), a_2(\mu_D, \gamma(\beta_0))) \in \operatorname*{arg\,min}_{(a_1,a_2) \in \mathbb{A}(\mu_D, \gamma(\beta_0))} \sup_{\delta \in \mathcal{D}(\beta_0)} (\mathcal{P}_{\delta,\mu_D} - \mathbb{E}\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))). \tag{1.3.6}$$

Four remarks on the domain of (a_1, a_2) (i.e., $\mathbb{A}(\mu_D, \gamma(\beta_0))$) are in order. First, the lower bound $\underline{a}(\mu_D, \gamma(\beta_0))$ is motivated by Theorem 1.2.1(iii). Specifically, we require $p_1 \in (0, 1)$ and close to 0 and $p_2 > 1$. In the Appendix, we provide a detailed report on the finite sample performance of our CLC test for both simulation designs analyzed in Section 1.5 and the empirical application in Section 1.6, where we consider different values of p_1 and p_2. The results indicate that our test's finite sample performance is not affected by the specific values chosen for (p_1, p_2), as all the results are very close to those reported in the main paper. Second, under weak identification, μ_D is bounded, and $\frac{p_2 \mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}$ may be larger than p_1. In this case, we have $\mathbb{A}(\mu_D, \gamma(\beta_0)) = \{(a_1, a_2) \in \mathbb{A}_0, a_1 \in [p_1, 1]\}$. Third, under strong identification and local alternatives, $\frac{p_2 \mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}$ will converge to zero so that

$$\mathbb{A}(\mu_D, \gamma(\beta_0)) = \left\{ (a_1, a_2) \in \mathbb{A}_0, a_1 \in \left[\frac{p_2 \mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}, 1 \right] \right\}.$$

We show in Theorem 1.4.2 below that in this case, the minimax jackknife CLC test converges to $1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$ defined in Lemma 1.2.2, which is the UMP invariant and unbiased test. Furthermore, the minimax a_1 satisfies the requirement in Theorem 1.2.1(iii) with $\tilde{q} = 1.1\mathbb{C}_{\alpha,\max}(\rho(\beta_0))$ so that under strong identification, our CLC test has asymptotic power 1 against fixed alternatives, as shown in Theorem 1.4.4. Fourth, we require $\overline{a} < 1$ for some technical reason. In our simulations, we have not observed the minimax $a_1 + a_2$ reaching the upper bound. Therefore, setting the upper bound to $\overline{a}$ or 1 does not have any numerical impact.

Since we cannot observe the values of μ_D and $\gamma(\beta_0)$ in practice, we adopt the plug-in method described in Section 6 of I.Andrews (2016). Specifically, we replace $\gamma(\beta_0)$ with its consistent estimator $\widehat{\gamma}(\beta_0)$ as specified in Assumption 2. To obtain a proxy of μ_D,[11] we define

$$\widehat{\sigma}_D = \left(\widehat{\Upsilon} - (\widehat{\Phi}_{13}(\beta_0), \widehat{\tau}(\beta_0)) \begin{pmatrix} \widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0) \end{pmatrix}\right)^{1/2},$$

which is a function of $\widehat{\gamma}(\beta_0)$ and a consistent estimator of σ_D by Assumption 2. Then, under weak identification, we have $\widehat{D}^2/\widehat{\sigma}_D^2 = D^2/\sigma_D^2 + o_p(1) \overset{d}{=} \mathcal{Z}^2(\mu_D/\sigma_D) + o_p(1)$ and D^2/σ_D^2 is a sufficient statistic for μ_D^2. Let $\widehat{r} = \widehat{D}^2/\widehat{\sigma}_D^2$. We consider two estimators for μ_D as functions of $\widehat{D}$ and $\widehat{\sigma}_D$, namely, $f_{pp}(\widehat{D}, \widehat{\gamma}(\beta_0)) = \widehat{\sigma}_D\sqrt{\widehat{r}_{pp}}$ and $f_{krs}(\widehat{D}, \widehat{\gamma}(\beta_0)) = \widehat{\sigma}_D\sqrt{\widehat{r}_{krs}}$, where $\widehat{r}_{pp} = \max(\widehat{r} - 1, 0)$ and

$$\widehat{r}_{krs} = \widehat{r} - 1 + \exp\left(-\frac{\widehat{r}}{2}\right)\left(\sum_{j=0}^{\infty}\left(-\frac{\widehat{r}}{2}\right)^j \frac{1}{j!(1+2j)}\right)^{-1}.$$

Specifically, Kubokawa, Robert, and Saleh (1993) show that $\widehat{r}_{krs}$ is positive as long as $\widehat{r} > 0$ and $\widehat{r} \geq \widehat{r}_{krs} \geq \widehat{r} - 1$. It is also possible to consider the MLE based on a single observation $\widehat{D}^2/\widehat{\sigma}_D^2$. However, such an estimator is harder to use because it does not have a closed-form expression.

In practice, we estimate $\mathbb{E}\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))$ by $\mathbb{E}^*\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$ for $s \in \{pp, krs\}$, where

$$\begin{aligned} &\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0)) \\ &= 1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2(\widehat{C}_1(\delta) f_s(\widehat{D}, \widehat{\gamma}(\beta_0))) \\ +a_2\left[\widehat{\rho}(\beta_0)\mathcal{Z}_1(\widehat{C}_1(\delta) f_s(\widehat{D}, \widehat{\gamma}(\beta_0))) + (1-\widehat{\rho}^2(\beta_0))^{1/2}\mathcal{Z}_2(\widehat{C}_2(\delta) f_s(\widehat{D}, \widehat{\gamma}(\beta_0)))\right]^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(\widehat{C}_2(\delta) f_s(\widehat{D}, \widehat{\gamma}(\beta_0)) \geq \mathbb{C}_\alpha(a_1, a_2; \widehat{\rho}(\beta_0)) \end{array}\right\}, \end{aligned} \tag{1.3.7}$$

[11] In fact, as $\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))$ only depends on μ_D^2, we aim to find a good estimator for μ_D^2.

and $(\widehat{C}_1(\delta), \widehat{C}_2(\delta))$ are similarly defined as $(C_1(\delta), C_2(\delta))$ in (1.3.3) with $\gamma(\beta_0)$ replaced by $\widehat{\gamma}(\beta_0)$; that is,

$$\begin{pmatrix} \widehat{C}_1(\delta) \\ \widehat{C}_2(\delta) \end{pmatrix} \equiv \begin{pmatrix} \widehat{\Phi}_1^{-1/2}(\beta_0)\delta^2 \\ (1-\widehat{\rho}^2(\beta_0))^{-1/2}(\widehat{\Psi}^{-1/2}(\beta_0)\delta - \widehat{\rho}(\beta_0)\widehat{\Phi}_1^{-1/2}(\beta_0)\delta^2) \end{pmatrix}$$
$$\times \left[1 - (\delta^2, \delta) \left(\begin{pmatrix} \widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0) \end{pmatrix} \right) \right]^{-1}.$$

Let $\mathcal{P}_{\delta,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) = \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))} \mathbb{E}^*\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$. Then, for $s \in \{pp, krs\}$, we can estimate $a(\mu_D, \gamma(\beta_0))$ in (1.3.6) by $\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0)) = (\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)), \mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0)))$ defined as

$$\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0)) \in \underset{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))}{\arg\min} \sup_{\delta\in\mathcal{D}(\beta_0)} (\mathcal{P}_{\delta,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))), \tag{1.3.8}$$

where $\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$ is defined in (1.3.7),

$$\mathbb{A}(f_s(\widehat{D}, \widehat{\gamma}(\beta_0)), \widehat{\gamma}(\beta_0)) = \{(a_1, a_2) \in \mathbb{A}_0, a_1 \in [\underline{a}(f_s(\widehat{D}, \widehat{\gamma}(\beta_0)), \widehat{\gamma}(\beta_0)), \overline{a}]\},$$

$$\underline{a}(f_s(\widehat{D}, \widehat{\gamma}(\beta_0)), \widehat{\gamma}(\beta_0)) = \min\left(0.01, \frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) f_s^2(\widehat{D}, \widehat{\gamma}(\beta_0))}\right),$$

$$\widehat{c}_{\mathcal{B}}(\beta_0) = \sup_{\delta\in\mathcal{D}(\beta_0)} \left[1 - (\delta^2, \delta) \left(\begin{pmatrix} \widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0) \end{pmatrix} \right) \right]^2,$$

and $\widehat{\Delta}_*(\beta_0) = \widehat{\Phi}_1^{1/2}(\beta_0)\widehat{\Psi}^{-1/2}(\beta_0)\widehat{\rho}^{-1}(\beta_0)$. Then, the feasible jackknife CLC test is, for

$s \in \{pp, krs\}$,

$$\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} = 1 \left\{ \begin{array}{c} \mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0))AR^2(\beta_0) + \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))LM^2(\beta_0) \\ +(1 - \mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) - \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0)))LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha(\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)); \widehat{\rho}(\beta_0)) \end{array} \right\}. \tag{1.3.9}$$

1.4 Asymptotic Properties

We first consider the asymptotic properties of the jackknife CLC test under weak identification and fixed alternatives, in which $\mathcal{C} \to \widetilde{\mathcal{C}}$ and Δ is treated as fixed so that we have

$$\widehat{D} \rightsquigarrow D \stackrel{d}{=} \mathcal{N}(\mu_D, \sigma_D^2).$$

We see from (1.3.6) and (1.3.8) that $\mathcal{A}_s(d,r) = (a_1(f_s(d,r),r), a_2(f_s(d,r),r))$ is a function of $(d,r) \in \Re \times \Gamma$, where Γ is the parameter space for $\gamma(\beta_0)$ and $s \in \{pp, krs\}$. We make the following assumption on $\mathcal{A}_s(\cdot)$.

Assumption 3. *Let $\mathcal{S}_s$ be the set of discontinuities of $\mathcal{A}_s(\cdot,\gamma(\beta_0)) : \Re \mapsto [0,1] \times [0,1]$. Then, we assume $\mathcal{A}_s(d,r)$ is continuous in r for any $d \in \Re/\mathcal{S}_s$, and the Lebesgue measure of $\mathcal{S}_s$ is zero for $s \in \{pp, krs\}$.*

Assumption 3 is a technical condition that allows us to apply the continuous mapping theorem. It is mild because $\mathcal{A}_s(\cdot)$ is allowed to be discontinuous in its first argument. In practice, we can approximate $\mathcal{A}_s(\cdot)$ by a step function defined over a grid of d so that there is a finite number of discontinuities. The continuity of $\mathcal{A}_s(\cdot)$ in its second argument is due to the smoothness of the bivariate normal PDF with respect to the covariance matrix. Therefore, in this case, Assumption 3 holds automatically.

Theorem 1.4.1. *Suppose we are under weak identification and fixed alternatives and that Assumptions 1–3 hold. Then, for $s \in \{pp, krs\}$,*

$$\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)) \rightsquigarrow \mathcal{A}_s(D,\gamma(\beta_0)) = (a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)), a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)))$$

and[12]

$$\mathbb{E}\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} \to \mathbb{E}\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(\Delta,\mu_D,\gamma(\beta_0)),$$

where $\phi_{a_1,a_2,\infty}(\delta)$ *is defined in* (1.3.4) *and* $a_l(f_s(D,\gamma(\beta_0)),\gamma(\beta_0))$ *is interpreted as* $a_l(\mu_D,\gamma(\beta_0))$ *defined in* (1.3.6) *with* μ_D *replaced by* $f_s(D,\gamma(\beta_0))$ *for* $l = 1,2$ *defined in Section 1.3.*

In addition, let BL_1 *be the class of functions* $h(\cdot)$ *of* D *that is bounded and Lipschitz with Lipschitz constant 1. Then, if the null hypothesis holds such that* $\Delta = 0$*, we have*

$$\mathbb{E}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} - \alpha)h(\widehat{D}) \to 0, \quad \forall h \in BL_1.$$

Several remarks on Theorem 1.4.1 are in order. First, we see that the power of our jackknife CLC test is $\mathbb{E}\phi_{\mathcal{A}_s(D,\gamma(\beta_0)),\infty}(\Delta,\mu_D,\gamma(\beta_0))$, which does not exactly match the minimax power

$$\mathbb{E}\phi_{a_1(\mu_D,\gamma(\beta_0)),a_2(\mu_D,\gamma(\beta_0)),\infty}(\Delta,\mu_D,\gamma(\beta_0))$$

in the limit problem. This is because under weak identification, it is impossible to consistently estimate μ_D, or equivalently, the concentration parameter. A similar result holds under weak identification with a fixed number of moment conditions in I.Andrews (2016). The best we can do is to approximate μ_D by reasonable estimators based on D such as $f_{pp}(D,\gamma(\beta_0))$ and $f_{krs}(D,\gamma(\beta_0))$, which are random even asymptotically. Second, Theorem 1.4.1 implies that our jackknife CLC test controls size asymptotically conditionally on $\widehat{D}$, and thus, unconditionally. Last, according to Theorem 1.4.1, the CLC test's asymptotic power, with weights (a_1,a_2) chosen through the minimax procedure, is equivalent to the limit experiment's asymptotic power when the weights are $\mathcal{A}_s(D,\gamma(\beta_0))$, which is a function of D. As D is independent of the normal random variables in $\phi_{a_1,a_2,\infty}(\delta)$ in (1.3.4), the two optimality results stated in Theorem 1.2.1(i) also hold asymptotically, conditional on $\widehat{D}$. To make this statement precise, we define the eigenvalue decomposition

$$\begin{pmatrix} \mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) + \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}^2(\beta_0) & \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}(\beta_0)(1-\hat{\rho}^2(\beta_0))^{1/2} \\ \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}(\beta_0)(1-\hat{\rho}^2(\beta_0))^{1/2} & 1 - \mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) - \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}^2(\beta_0) \end{pmatrix}$$

[12] We assume that $\frac{C}{0} = +\infty$ if $C > 0$ and $\min(C, +\infty) = C$.

$$= \mathcal{U}_s(\widehat{D}, \widehat{\gamma}(\beta_0)) \begin{pmatrix} \nu_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) & 0 \\ 0 & \nu_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) \end{pmatrix} \mathcal{U}_s(\widehat{D}, \widehat{\gamma}(\beta_0))^\top. \quad (1.4.1)$$

Define a class of tests

$$\Phi_\alpha = \left\{ \begin{array}{c} \tilde{\phi}(\mathcal{Z}_1^2, \mathcal{Z}_2^2, d, r) : \mathbb{E}\tilde{\phi}(\mathcal{Z}_1^2, \mathcal{Z}_2^2, d, r) \leq \alpha, \text{ for any } (d, r) \in \Re \times \Gamma, \\ \tilde{\phi}(\mathcal{Z}_1^2, \mathcal{Z}_2^2, d, r) \text{ is continuous in } r, \\ \text{the discontinuities of } \tilde{\phi}(\mathcal{Z}_1^2, \mathcal{Z}_2^2, d, r) \text{ w.r.t.} \\ \text{the first three arguments have zero Lebesgue measure} \end{array} \right\},$$

where $(\mathcal{Z}_1, \mathcal{Z}_2)$ are two independent standard normal random variables. Further define, for $s \in \{pp, krs\}$,

$$\begin{pmatrix} \widetilde{AR}_s(\beta_0) \\ \widetilde{LM}^*_s(\beta_0) \end{pmatrix} = \mathcal{U}_s(\widehat{D}, \widehat{\gamma}(\beta_0))^\top \begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix}.$$

Assumption 4. *Suppose $\mathcal{U}_s(d, r)$ is continuous in r and the set of discontinuities of $\mathcal{U}_s(\cdot)$ w.r.t. its first argument has zero Lebesgue measure.*

Corollary 1.4.1. *Suppose we are under weak identification and fixed alternatives and that Assumptions 1–4 hold. Let $\tilde{\phi}(\cdot) \in \Phi_\alpha$ and for any $d \in \Re$, denote $(\theta_1, \theta_2) = (m_1(\Delta), m_2(\Delta))\mathcal{U}_s(d, \gamma(\beta_0))$. Then, the following two optimality results hold.*

(i) If for some $d \in \Re$ and $s \in \{pp, krs\}$, we have

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\tilde{\phi}(\widetilde{AR}_s^2(\beta_0), \widetilde{LM}_s^{*2}(\beta_0), \widehat{D}, \widehat{\gamma}(\beta_0)) 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}} \geq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\hat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}},$$

for all $(\theta_1, \theta_2) \in \Re^2$, then

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\tilde{\phi}(\widetilde{AR}_s^2(\beta_0), \widetilde{LM}_s^{*2}(\beta_0), \widehat{D}, \widehat{\gamma}(\beta_0)) 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}} = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\hat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}},$$

for all $(\theta_1, \theta_2) \in \Re^2$.

(ii) If $(\theta_1^2, \theta_2^2) = b \cdot (\nu_{1,s}(d, \gamma(\beta_0)), \nu_{2,s}(d, \gamma(\beta_0)))$ *for some positive constant* b*, then there exists* $\overline{b} > 0$ *such that if* $0 < b < \overline{b}$*, we have*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\tilde{\phi}(\widetilde{AR}_s^2(\beta_0), \widetilde{LM}_s^{*2}(\beta_0), \widehat{D}, \widehat{\gamma}(\beta_0)) 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}}$$
$$\leq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathbb{E}\hat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} 1\{|\widehat{D} - d| \leq \varepsilon\}}{\mathbb{E}1\{|\widehat{D} - d| \leq \varepsilon\}},$$

Corollary 1.4.1 shows that under weak identification and fixed alternatives, our jackknife CLC test is asymptotically admissible and optimal against certain alternatives conditional on $\widehat{D}$.

Next, we consider the performance of $\widehat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))}$ defined in (1.3.9) under strong identification and local alternatives. To precisely state the optimality result, we further consider the class of level-α tests against $\theta = 0$ v.s. the two-sided alternative that are constructed based on one observation of $(\mathcal{N}_1, \mathcal{N}_2)$, where $\theta = \widetilde{\Delta}\widetilde{\mathcal{C}}\Psi^{-1/2}$ and

$$\begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

Specifically, denote

$$\Phi_\alpha^I = \left\{ \begin{array}{c} \phi(\cdot) : \mathbb{E}\phi(\mathcal{N}_1, \mathcal{N}_2) \leq \alpha \quad \text{under the null}, \\ \phi(\mathcal{N}_1, \mathcal{N}_2) = \phi(\mathcal{N}_1, -\mathcal{N}_2), \\ \text{the discontinuities of } \phi(\cdot) \text{ has zero Lebesgue measure} \end{array} \right\}$$

and

$$\Phi_\alpha^U = \left\{ \begin{array}{c} \phi(\cdot) : \mathbb{E}\phi(\mathcal{N}_1, \mathcal{N}_2) \leq \alpha \quad \text{under the null}, \\ \mathbb{E}\phi(\mathcal{N}_1, \mathcal{N}_2) \geq \alpha \quad \text{under the alternative}, \\ \text{the discontinuities of } \phi(\cdot) \text{ has zero Lebesgue measure} \end{array} \right\}$$

as the classes of sign-invariant and unbiased tests, respectively.

Theorem 1.4.2. *Suppose that Assumptions 1 and 2 hold. Further suppose that we are under strong identification and local alternatives as described in Lemma 1.2.1. Then, for* $s \in \{pp, krs\}$*, we have*

$$\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) \xrightarrow{p} 0, \quad \mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0))\rho \xrightarrow{p} 0, \quad \text{and} \quad \widehat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} \rightsquigarrow 1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\},$$

where $\mathcal{N}_2^* \overset{d}{=} \mathcal{N}\left(\frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{[(1-\rho^2)\Psi]^{1/2}}, 1\right)$. *In addition, suppose* $\breve{\phi}_n$ *is a generic test such that* $\breve{\phi}_n = \phi(AR(\beta_0), LM(\beta_0)) + o_P(1)$ *for some* $\phi \in \Phi_\alpha^I \cup \Phi_\alpha^U$ *and the sequence* $\{\breve{\phi}_n\}_{n\geq 1}$ *is uniformly integrable. Then, we have*

$$\lim_{n\to\infty} \mathbb{E}\widehat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} = \sup_{\phi \in \Phi_\alpha^I \cup \Phi_\alpha^U} \lim_{n\to\infty} \mathbb{E}\phi(AR(\beta_0), LM(\beta_0)) \geq \lim_{n\to\infty} \mathbb{E}\breve{\phi}_n.$$

Five remarks are in order. First, under strong identification, μ_D, and thus, D approaches infinity, and so does our estimator $\widehat{D}$. This is how our estimator $\widehat{D}$ can detect the identification strength. In addition, we show in the proof of Theorem 1.4.2 that under strong identification, the calibrated power gap $\mathcal{P}_{\delta,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$ is maximized when δ is in the region of local alternatives. However, in this region, as shown by Lemma 1.2.2, the maximum power gap can achieve zero if all the weights are put on $LM^*(\beta_0)$, which leads to the first result in Theorem 1.4.2. Second, our jackknife CLC test is adaptive to identification strength. In practice, econometricians do not know whether the true value β is close to the null β_0. Therefore, our jackknife CLC test calibrates power across all possible values of δ (i.e., $\delta \in \mathcal{D}(\beta_0)$), which include both local and fixed alternatives. Yet, Theorem 1.4.2 shows that the minimax procedure can produce the most powerful test as if it is known that β belongs to the region of local alternatives. Third, Theorem 1.4.2 shows that under strong identification and local alternatives, our jackknife CLC test converges to the UMP level-α test that is either invariant to the sign change or unbiased and constructed based on $AR(\beta_0)$ and $LM(\beta_0)$. Therefore, it is more powerful than both the jackknife AR and LM tests. Fourth, under strong identification and local alternatives, the JIVE-based Wald test proposed by Chao et al. (2012) is asymptotically equivalent to the jackknife LM test, which implies that the jackknife AR and JIVE-Wald-based two-step test in Mikusheva and Sun (2022) is also dominated by the jackknife CLC test. Fifth, consider the HLIM based Wald test statistic proposed by Hausman et al. (2012), which is denoted as $W_h(\beta_0)$. In Section A.20 in the Appendix, we show that, under local alternative and strong identification,

$$W_h(\beta_0) = \frac{\Psi^{1/2}}{\Psi_h^{1/2}} LM(\beta_0) - \frac{\tilde{\rho}\Phi_1^{1/2}}{\Psi_h^{1/2}} AR(\beta_0) + o_P(1),$$

where $\tilde{\rho} = plim_{n\to\infty} X^\top e(\beta_0)/(e(\beta_0)^\top e(\beta_0))$ and $\Psi_h = \Psi - 2\tilde{\rho}\Phi_{12} + \tilde{\rho}^2\Phi_1$ is the corresponding asymptotic variance. Then, by letting $\breve{\phi}_n = 1\{W_h^2(\beta_0) \geq \mathbb{C}_\alpha\}$ and

$$\phi(AR(\beta_0), LM(\beta_0)) = 1\left\{\left[\frac{\Psi^{1/2}}{\Psi_h^{1/2}} LM(\beta_0) - \frac{\tilde{\rho}\Phi_1^{1/2}}{\Psi_h^{1/2}} AR(\beta_0)\right]^2 \geq \mathbb{C}_\alpha\right\},$$

Theorem 1.4.2 implies our jackknife CLC test is more powerful than the HLIM based Wald test under strong identification against local alternatives. In fact, by direct calculation, we can see that, for $\theta = \widetilde{\Delta}\widetilde{\mathcal{C}}\Psi^{-1/2}$,

$$\frac{\Psi^{1/2}}{\Psi_h^{1/2}} LM(\beta_0) - \frac{\tilde{\rho}\Phi_1^{1/2}}{\Psi_h^{1/2}} AR(\beta_0) \rightsquigarrow \mathcal{Z}(\tilde{\theta}), \quad \text{where} \quad \tilde{\theta}^2 = \frac{\theta^2}{1-\rho^2 + \left(\tilde{\rho}\Phi_1^{1/2}\Psi^{-1/2} - \rho\right)^2} \leq \frac{\theta^2}{(1-\rho^2)}.$$

The noncentrality parameter for the HLIM based Wald test is weakly smaller than that of the CLC test, which explains the power comparison. The equality holds if $\tilde{\rho}\Phi_1^{1/2}\Psi^{-1/2} = \rho$, which further holds in the special case of many weak IVs and homoskedasticity in the sense that $\Pi^\top\Pi/K = o(1)$ and $\mathbb{E}(V_i, e_i)^\top(V_i, e_i)$ does not vary across i.

Combining Theorems 1.4.1 and 1.4.2, we can show the uniform size control of our jackknife CLC test no matter the identification is strong or weak. Let $\lambda_n \in \Lambda_n$ be the data generating process of n observations of (e, V, Z). Under λ_n, the covariance matrix of $(Q_{e,e}, Q_{X,e}, Q_{X,X})$ is denoted as $\mathbb{V}_n$. We impose the following restriction on the sequence of classes of DGPs

$(\{\Lambda_n\}_{n\geq 1})$:[13]

$$\left(\begin{array}{c} \{V_i, e_i\}_{i\in[n]} \text{ are independent, } \mathbb{E}e_i = \mathbb{E}V_i = 0, \\ \max_i \mathbb{E}e_i^4 + \max_i \mathbb{E}V_i^4 \leq C_1 < \infty, \\ \mathcal{C}_n = \frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} \Pi_i P_{ij} \Pi_j \in \Re, \\ P_{ii} \leq C_2 < 1, \\ 0 < \kappa_1 \leq \text{mineig}(\mathbb{V}_n) \leq \text{maxeig}(\mathbb{V}_n) \leq \kappa_2 < \infty, \\ \text{where } C_1,\ C_2,\ \kappa_1, \text{ and } \kappa_2 \text{ are some fixed constants,} \\ \text{and Assumption 2 holds for } \beta_0 = \beta. \end{array}\right) \tag{1.4.2}$$

In Sections A.2.1 and A.2.2 of the Appendix, we further verify that Assumption 2 holds, respectively, for the standard variance estimators, which follow the construction in Crudu et al. (2021), and the cross-fit variance estimators, which follow Mikusheva and Sun (2022). Theorem 1.4.3 shows that our jackknife CLC test has correct asymptotic size, under similar arguments as those in Andrews, Cheng, and Guggenberger (2020a) and I.Andrews (2016).

Theorem 1.4.3. *Suppose Assumption 3 holds, $\{\Lambda_n\}_{n\geq 1}$ satisfies (1.4.2), and we are under the null hypothesis that $\beta_0 = \beta$. Then, we have*

$$\liminf_{n\to\infty} \inf_{\lambda_n\in\Lambda_n} \mathbb{E}_{\lambda_n}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}) = \limsup_{n\to\infty} \sup_{\lambda_n\in\Lambda_n} \mathbb{E}_{\lambda_n}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}) = \alpha.$$

Last, we show that, under strong identification, the jackknife CLC test $\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}$ defined in (1.3.9) has asymptotic power 1 against fixed alternatives.

Theorem 1.4.4. *Suppose Assumption 2 holds, and $(Q_{e(\beta_0),e(\beta_0)} - \Delta^2\mathcal{C}, Q_{X,e(\beta_0)} - \Delta\mathcal{C}, Q_{X,X} - \mathcal{C})^\top = O_p(1)$. Further suppose that we are under strong identification with fixed alternatives so that $\Delta = \beta - \beta_0$ is nonzero and fixed. Then, we have*

$$\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} \xrightarrow{p} 1.$$

[13] In (1.4.2), we focus on the model without exogenous control variables. The independence and moment conditions for (e_i, V_i) are sufficient for Assumption 1. We further verify in Section A.1 of the Appendix that the joint asymptotic normality (Assumption 1) holds in the case with exogenous controls.

1.5 Simulation

1.5.1 Power Curve Simulation for the Limit Problem

In this section, we present simulation results to compare the power performance of various tests under the limit problem described in Section 1.2. We consider the following tests with a nominal rate of 5%: (i) our jackknife CLC test, where μ_D is estimated using either pp or krs method, (ii) the one-sided jackknife AR test defined in (1.2.5), (iii) the jackknife LM test defined in (1.2.6), and (iv) the test that is based on the orthogonalized jackknife LM statistic $LM^{*2}(\beta_0)$ defined in this paper. We conduct 5,000 simulation replications to obtain stable simulation results.

We set the parameter space for β as $\mathcal{B} = [-6/\mathcal{C}, 6/\mathcal{C}]$, where $\mathcal{C} = 3$ and 6. The choice of parameter space follows that in I.Andrews (2016, Section 7.2). We set $\beta_0 = 0$, and the values of the covariance matrix in (1.2.2) are set as follows: $\Phi_1 = \Psi = \Upsilon = 1$, and $\Phi_{12} = \Phi_{13} = \tau = \rho$, where $\rho \in \{0.2, 0.4, 0.7, 0.9\}$. We then compute $\gamma(\beta_0)$ based on (1.2.4) as β ranges over $\mathcal{B}$ and generate $AR(\beta_0)$ and $LM(\beta_0)$ based on (1.2.3). Last, we implement our CLC test purely based on $AR(\beta_0)$, $LM(\beta_0)$, $\gamma(\beta_0)$, and $\mathcal{B}$ without assuming the knowledge of $(\mathcal{C}, \beta)$. We have tried to simulate under alternative settings of the covariance matrix, and the obtained patterns of the power behavior are very similar.

Figures 1.1–1.4 plot the power curves for $\rho = 0.2, 0.4, 0.7$, and 0.9. In each figure, we report the results under both $\mathcal{C} = 3$ and 6. We observe that overall, the two jackknife CLC tests have the best power properties in terms of minimizing the maximum regret. Especially when the identification is relatively strong ($\mathcal{C} = 6$) and/or the degree of endogeneity is not very low ($\rho = 0.4, 0.7$, or 0.9), the jackknife CLC tests outperform their AR and LM counterparts by a large margin. In addition, we notice that when $\mathcal{C} = 3$, for some parameter values $LM^*(\beta_0)$ can suffer from substantial declines in power relative to the other tests, which is in line with our theoretical predictions. By contrast, our jackknife CLC tests are able to guard against such substantial power loss because of the adaptive nature of their minimax procedure. In Section A.21.1 of the Appendix, we further report power curves for alternative values of the tuning parameters (p_1, p_2) in (1.3.5) and of $\mathcal{C}$, and find that the overall patterns remain very similar.

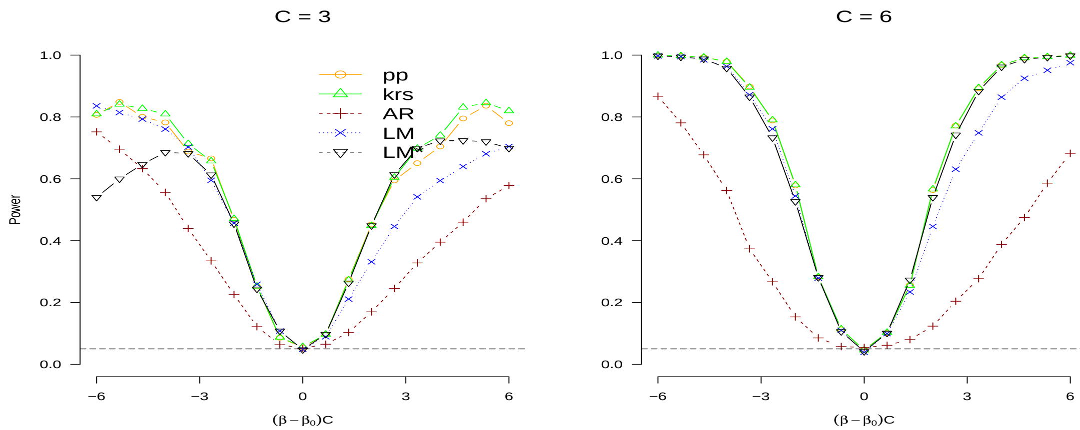

Figure 1.1: Power Curve for $\rho = 0.2$ with nominal size represented by the horizontal dotted line.
Note: The orange line with circle represents *pp*, which is the probability of rejection by using the test $\phi_{a_1,a_2,pp}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$; the green line with upward-pointing triangle represents *krs*, which is the probability of rejection by using the test $\phi_{a_1,a_2,krs}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$; the brown dash line with additive sign represents AR test given in (1.2.5); the blue dotted line with cross represents LM test given in (1.2.6); the dark dash line with downward-pointing triangle represents LM^* test defined just above (1.2.7).

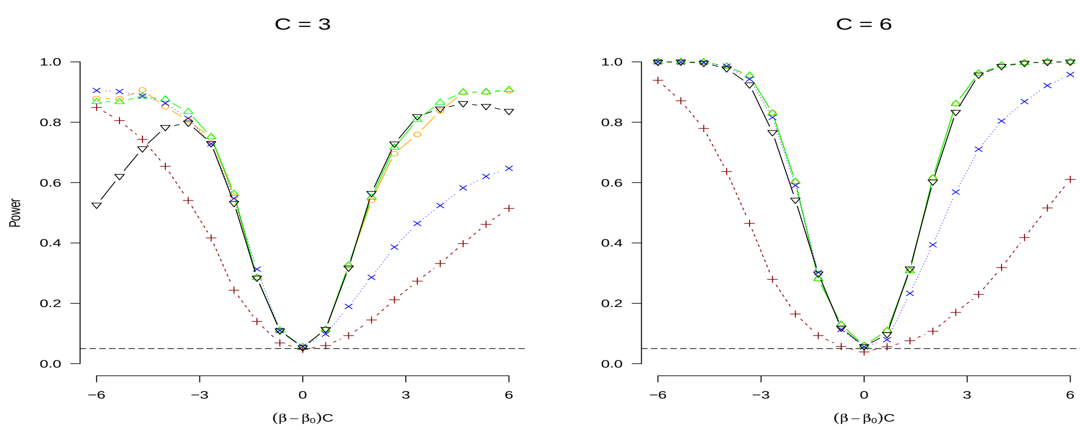

Figure 1.2: Power Curve for $\rho = 0.4$ with nominal size represented by the horizontal dotted line.
Note: The lines are explained under Figure 1.1.

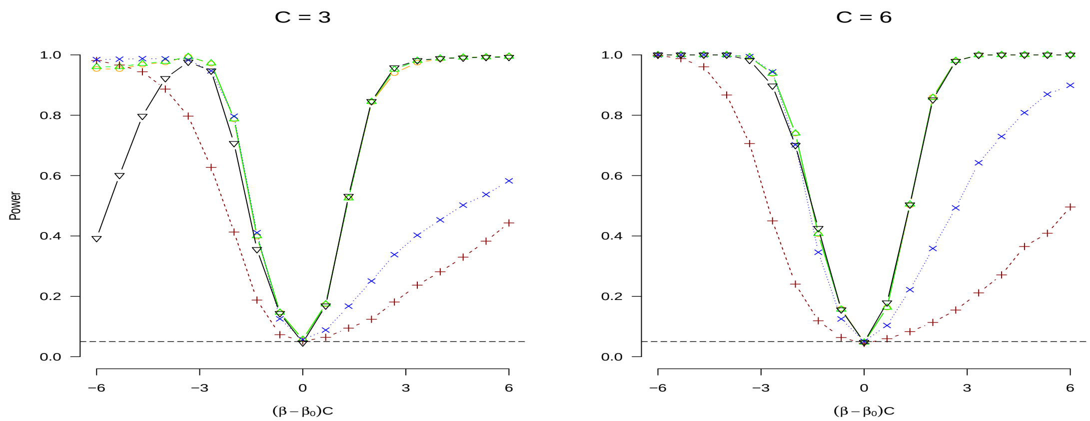

Figure 1.3: Power Curve for $\rho = 0.7$ with nominal size represented by the horizontal dotted line.

Note: The lines are explained under Figure 1.1.

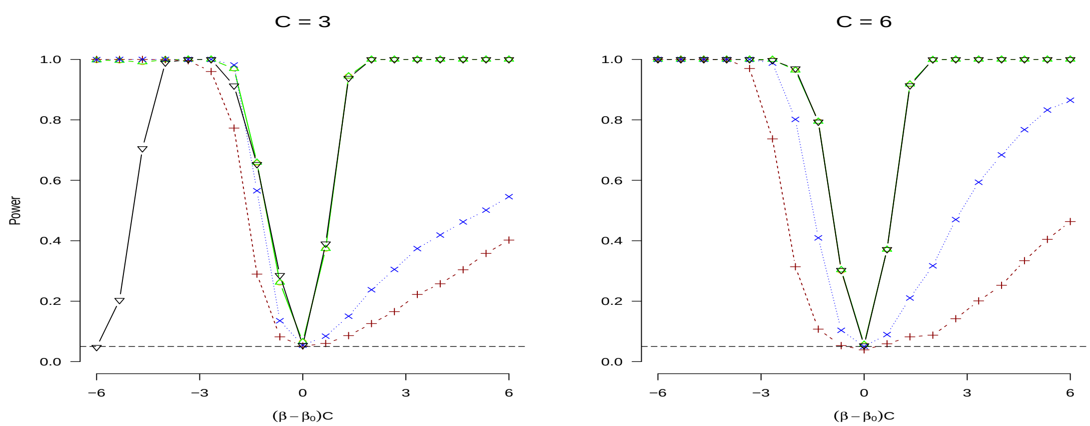

Figure 1.4: Power Curve for $\rho = 0.9$ with nominal size represented by the horizontal dotted line.

Note: The lines are explained under Figure 1.1.

1.5.2 Simulation Based on Calibrated Data

We follow the approach of Angrist and Frandsen (2022) and Mikusheva and Sun (2022) and use a data generating process (DGP) calibrated based on the 1980 census dataset from Angrist and Krueger (1991). We define the instruments as

$$\tilde{Z}_i = \left((1\{Q_i = q, C_i = c\})_{q \in \{2,3,4\}, c \in \{31, \cdots, 39\}}, (1\{Q_i = q, P_i = p\})_{q \in \{2,3,4\}, p \in \{51 \text{ states}\}} \right),$$

where Q_i, C_i, P_i are individual i's quarter of birth (QOB), year of birth (YOB) and place of birth (POB), respectively, so that there are 180 instruments. Note that the dummy with $q = 1$ and $c = 30$ is omitted in $\tilde{Z}_i$. We denote $\tilde{Y}_i$ as income, $\tilde{X}_i$ as the highest grade completed, and $\tilde{W}_i$ as the full set of YOB-POB interactions; that is,

$$\tilde{W}_i = \left(1\{C_i = c, P_i = p\}_{c\in\{30,\ldots,39\},p\in\{51 \text{ states}\}}\right),$$

which is a 510×1 matrix.

As in Angrist and Frandsen (2022), using the full 1980 sample (consisting of 329,509 individuals), we first obtain the average $\tilde{X}_i$ for each QOB-YOB-POB cell; we call this $\bar{s}(q, c, p)$. Next we use LIML to estimate the structural parameters in the following linear IV regression:

$$\tilde{Y}_i = \tilde{X}_i\beta_X + \tilde{W}_i^\top\beta_W + e_i,$$

$$\tilde{X}_i = \tilde{Z}_i^\top\Gamma_Z + \tilde{W}_i^\top\Gamma_W + V_i,$$

where $\tilde{X}$ is endogenous and instrumented by $\tilde{Z}_i$ and $\tilde{W}_i$ is the exogenous control variable. Denote the LIML estimate for $\beta_{X,W} \equiv (\beta_X^\top, \beta_W^\top)^\top$ as $\widehat{\beta}_{LIML}^\top = (\widehat{\beta}_{LIML,X}^\top, \widehat{\beta}_{LIML,W}^\top)$. We let $\widehat{y}(C_i, P_i) = \tilde{W}_i^\top\widehat{\beta}_{LIML,W}$ and

$$\omega(Q_i, C_i, P_i) = \tilde{Y}_i - \tilde{X}_i\widehat{\beta}_{LIML,X} - \tilde{W}_i^\top\widehat{\beta}_{LIML,W}.$$

Based on the LIML estimate and the calibrated $\omega(Q_i, C_i, P_i)$, we simulate the following two DGPs:

1. DGP 1:

$$\widetilde{y}_i = \bar{y} + \beta\widetilde{s}_i + \omega(Q_i, C_i, P_i)(\nu_i + \kappa_2\xi_i) \tag{1.5.1}$$

$$\widetilde{s}_i \sim Poisson(\mu_i),$$

   where β is the parameter of interest, ν_i and ξ_i are independent standard normal, $\bar{y} = \frac{1}{n}\sum_{i=1}^n \widehat{y}(C_i, P_i)$, $\mu_i \equiv max\{1, \gamma_0 + \gamma_Z^\top\tilde{Z}_i + \kappa_1\nu_i\}$, and $\gamma_0 + \gamma_Z^\top\tilde{Z}_i$ is the projection of $\bar{s}_i(q, c, p)$ onto a constant and $\tilde{Z}_i$. We set $\kappa_1 = 1.7$ and $\kappa_2 = 0.1$ as in Mikusheva and Sun (2022).

2. DGP 2: Same as DGP 1 except that $\kappa_1 = 2.7$ and

$$\widetilde{s}_i \sim \lfloor Poisson(2\mu_i)/2 \rfloor$$

We consider sample sizes of 0.5%, 1%, and 1.5% of the full sample size. Upon obtaining n observations, we exclude instruments with $\sum_{i=1}^{n} \tilde{Z}_{ij} < 5$. This results in three different sample sizes: small, medium, and large, with 1,648, 3,296, and 4,943 observations, respectively. The number of instruments also varies across sample sizes, with 119, 142, and 150 instruments for small, medium, and large samples, respectively. Our DGP 1 is exactly the same as that in Mikusheva and Sun (2022), with the correlation parameter of $\rho = 0.41$. DGP 2 has a higher correlation parameter of $\rho = 0.7$. The identification strength increases with the sample size. For DGP 1, the concentration parameters $\mathcal{C}/\Upsilon^{1/2}$ for small, medium, and large samples are 2.15, 3.62, and 4.85, respectively. For DGP 2, they are 2.38, 3.97, 5.28, respectively.

We emphasize that following Angrist and Frandsen (2022) and Mikusheva and Sun (2022), we only use $\tilde{W}_i$ to compute the LIML estimator and calibrate $\omega(Q_i, C_i, P_i)$, but do not use it to generate new data. Therefore, for the simulated data, the outcome variable is $\tilde{y}_i$, the endogenous variable is $\tilde{s}_i$, the IV $\tilde{Z}_i$ is viewed to be fixed, and the exogenous control variable is just an intercept. We then denote the demeaned versions of $\tilde{y}_i$, $\tilde{s}_i$, and $\tilde{Z}_i$ as Y_i, X_i, and Z_i, respectively, in (1.2.1) and implement various inference methods described below. Following Mikusheva and Sun (2022), we test the null hypothesis that $\beta = \beta_0$ for $\beta_0 = 0.1$ while varying the true value $\beta \in \mathcal{B}$. The parameter space is set as $\mathcal{B} = [-0.5, 0.5]$, which is consistent with the choice of parameter space for the empirical application below. The results below are based on 1,000 simulation repetitions. We provide more details about the implementation in Section A.3 in the Appendix. We set $(p_1, p_2) = (0.01, 1.1)$ in (1.3.5). Additional simulation results using other choices of (p_1, p_2) and $\mathcal{B}$ are reported in Section A.21.2 in the Appendix. All of them are very close to what we report here.

We compare the following tests with a nominal rate of 5%:

1. pp: our jackknife CLC test when μ_D is estimated by the method pp.

2. krs: our jackknife CLC test when μ_D is estimated by the method krs.

3. AR: the one-sided jackknife AR test with the cross-fit variance estimator proposed by

Mikusheva and Sun (2022).

4. LM_CF: Matsushita and Otsu's (2021) jackknife LM test, but with a cross-fit variance estimator (details are given in Section A.2.2 in the Appendix).

5. 2-step: Mikusheva and Sun's (2022) two-step estimator in which the overall size is set at 5%.

6. LM*: LM* test defined in this paper.

7. LM_MO: Matsushita and Otsu's (2021) original jackknife LM test.

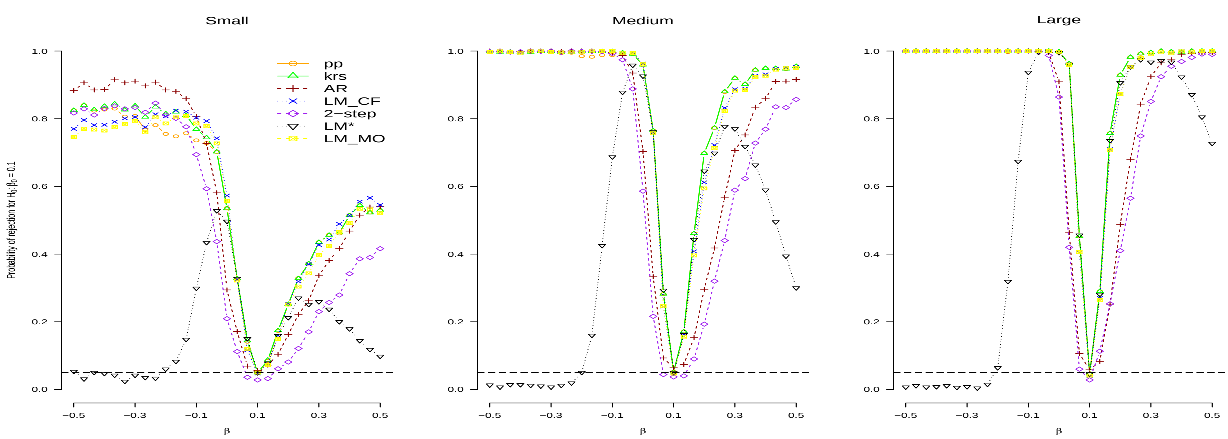

Figure 1.5: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 1.1)$ and nominal size of 5% represented by the horizontal dotted line

Note: The orange line with circle represents *pp* test; the green line with upward-pointing triangle represents *krs* test; the brown dash line with additive sign represents AR test given in (1.2.5); the blue dotted line with cross represents LM test with cross-fit variance; the purple dash line with diamond represents the 2-step test proposed by Mikusheva and Sun (2022) with overall 5% significance level; dark line with downward-pointing triangle represents LM^*; the yellow dash line with rectangle represents the LM test proposed by Matsushita and Otsu (2021).

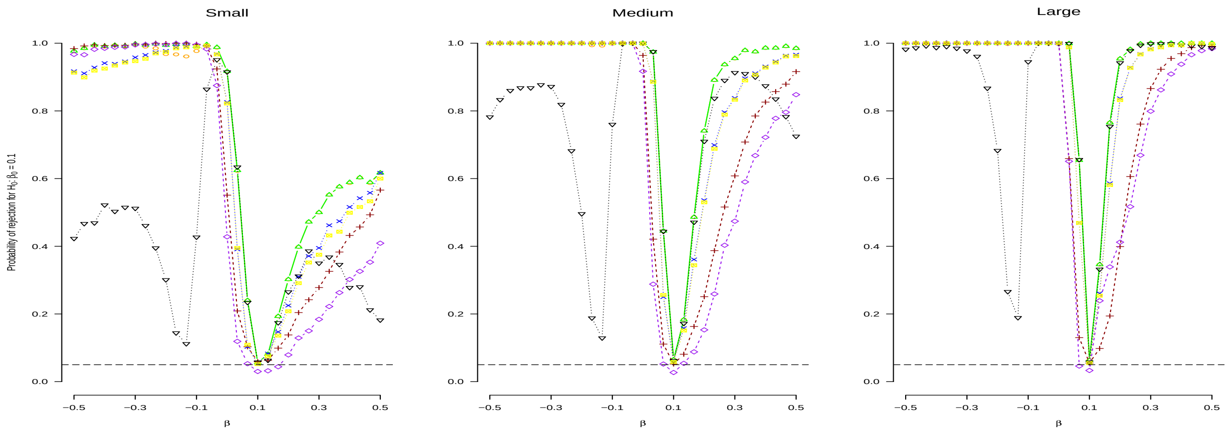

Figure 1.6: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 1.1)$ and nominal size of 5% represented by the horizontal dotted line
Note: The lines are explained in Figure 1.5.

Figures 1.5 and 1.6 plot the power curves of the aforementioned tests. We can make four observations. First, all methods control size well because they are all weak identification robust. Second, the performance of the jackknife CLC test with krs is slightly better than that with pp, which is consistent with the power curve simulation in Section 1.5.1. Third, in DGP 1 with a small sample size, the power of the jackknife AR test is at most about 9.2% higher than that of the krs test when β is around -0.3. However, for alternatives close to the null (e.g., when β is around 0), the power of the krs test is 24% higher, which implies that the power of the krs test is still better than that for the jackknife AR test in the minimax sense. The power of the jackknife LM tests is similar to that of the krs test in DGP 1 with a small sample size. Fourth, for the rest of the scenarios, the power of the krs test is the highest in most regions of the parameter space. The power of the jackknife AR and LM is at most 0.7% higher than that of the krs test at some point. For DGP 1 with medium and large sample sizes, the maximum power gaps between our krs test and the jackknife LM are about 8.6% and 5.6%, and about 43.2% and 50% compared with the jackknife AR. Furthermore, they are 23.3%, 19.5%, and 18.5% compared with the jackknife LM for DGP 2 with small, medium, and large sample sizes, respectively, and about 41.5%, 55.3%, and 55.85% compared with the jackknife AR.

Figures 1.7 and 1.8 show the average values of (a_1, a_2), which represents the weights assigned to $AR(\beta_0)$ and $LM(\beta_0)$ in our CLC tests, under DGPs 1 and 2, respectively. The weight assigned to $LM^*(\beta_0)$ is simply $1 - a_1 - a_2$. As shown in Table 1.1, under weak identification and fixed alternatives, there is no clear winner among $AR(\beta_0)$, $LM(\beta_0)$, and $LM^*(\beta_0)$, and thus, our CLC test assigns weights to all the three tests. However, under strong identification and local alternative, $LM^*(\beta_0)$ is the UMP test and should carry all the weights, which means $a_1 + a_2$ should be minimum. On the other hand, under strong identification and for some fixed alternatives, $LM^*(\beta_0)$ may lack power while both $AR(\beta_0)$ and $LM(\beta_0)$ have power 1. In this case, as long as we do not assign all weights on $LM^*(\beta_0)$, our CLC test should also have power 1. We observe that our simulation results are consistent with these theoretical predictions. First, when β_0 is close to the null 0.1, both a_1 and a_2 are small, indicating that most of the weights are put on $LM^*(\beta_0)$. Second, we observe from Figures 1.5 and 1.6 that the power of $LM^*(\beta_0)$ drops rapidly when β is smaller than around zero. Therefore, our CLC test assigns more weights on $AR(\beta_0)$ and $LM(\beta_0)$. Third, for distant alternatives, significant weights are assigned to $AR(\beta_0)$ and $LM(\beta_0)$, which ensures the good power of our CLC test. Additionally, we note that the weights assigned to $AR(\beta_0)$ (a_1) are higher on the left side of the parameter space relative to the right, since $AR(\beta_0)$ is more powerful on the left.

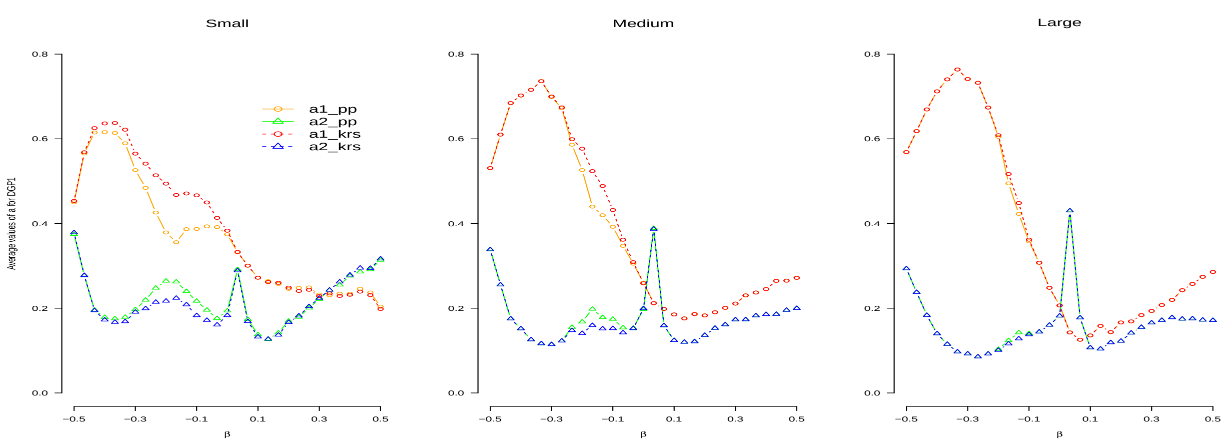

Figure 1.7: Average Values of a for DGP 1.
Note: The orange line with circle represents the average value of a_1 in the pp test; the green line with upward-pointing triangle represents the average value of a_2 in the pp test; the red dotted line with circle represents the average value of a_1 in the krs test; the blue dotted line with upward-pointing triangle represents the average value of a_2 in the krs test.

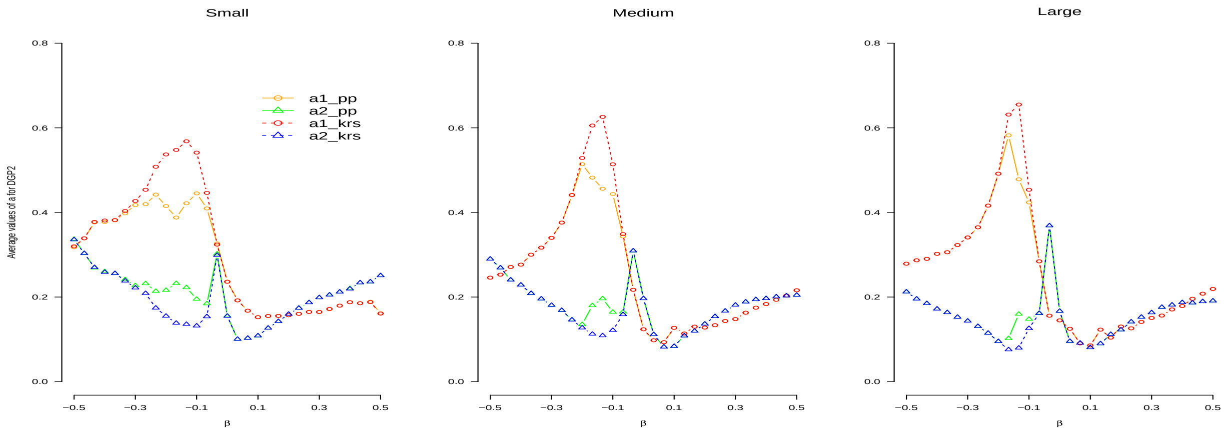

Figure 1.8: Average Values of a for DGP 2

Note: The lines are explained in Figure 1.7.

1.6 Empirical Application

In this section, we consider the linear IV regressions with the specification underlying Angrist and Krueger (1991, Table VII, column (6)), using the full original dataset.[14] The outcome variable Y and endogenous variable X are log weekly wages and schooling, respectively. We follow Angrist and Krueger (1991) and focus on two specifications with 180 and 1,530 instruments. The 180 instruments consist of 30 quarter and year of birth interactions (QOB-YOB) and 150 quarter and place of birth interactions (QOB-POB). The second specification includes full interactions among QOB-YOB-POB, resulting in 1,530 instruments. The exogenous control variables have been partialled out from the outcome, endogenous variables, and IVs. Further details on the empirical application can be found in Section A.4 in the Appendix. The considered tests are similar to those in the previous section. The jackknife AR test is defined in (1.2.5) with $\widehat{\Phi}_1$ being the cross-fit estimator in Mikusheva and Sun (2022). The jackknife LM test is defined in (1.2.6) with the cross-fit estimator for $\Psi(\beta_0)$. The pp and krs tests are our jackknife CLC tests. The two-step procedure is given by Mikusheva and Sun (2022, Section 5). Specifically, the researcher accepts the null if $\widetilde{F} > 9.98$

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

and $Wald(\beta_0) < \mathbb{C}_{0.02}$[15] or if $\widetilde{F} \leq 9.98$ and $AR(\beta_0) < z_{0.02}$. In the case of 180 instruments, because $\widetilde{F} = 13.42 > 9.98$, the lower and upper bounds of the 95% confidence interval (CI) for the two-step procedure correspond respectively to the minimum and maximum of the set $\{\beta_0 \in \Re : Wald(\beta_0) < \mathbb{C}_{0.02}\}$; similarly, for the 1,530 instruments, as $\widetilde{F} = 6.32 \leq 9.98$, the lower and upper bounds of the CI for the two-step procedure correspond respectively to the minimum and maximum of the set $\{\beta_0 \in \Re : AR(\beta_0) < z_{0.02}\}$. We also report the 95% Wald test CI based on the JIVE estimator, denoted as JIVE-t. Table 1.2 reports the 95% CIs by inverting the corresponding 5% tests mentioned above for the parameter space $\mathcal{B} = [-0.5, 0.5]$. Note all CIs except JIVE-t are robust to weak identification. As $\widetilde{F}$'s are higher than 4.14 in both cases, the JIVE-t (5%) has the Stock and Yogo (2005b)-type guarantee with at most a 5% size distortion (i.e., the overall size is less than 10%). We set (p_1, p_2) in (1.3.5) as $(0.01, 1.1)$. The empirical results with other choices of (p_1, p_2) and $\mathcal{B}$ are reported in Section A.22 of the Appendix. All of them are very close to what we report here.

	jackknife AR (5%)	jackknife LM (5%)	JIVE-t (5%)	Two-step (5%)	pp (5%)	krs (5%)
180 IVs	[0.008,0.201]	[0.067,0.135]	[0.066,0.132]	[0.059,0.139]	[0.067,0.128]	[0.067,0.128]
1530 IVs	[-0.035,0.22]	[0.036,0.138]	[0.035,0.133]	[-0.051,0.242]	[0.037,0.133]	[0.037,0.133]

Table 1.2: **Confidence Intervals**

Notes: The $\widetilde{F}$'s for 180 and 1,530 instruments are 13.42 and 6.32, respectively. The grid-search used for our confidence interval was over 10,000 equidistant grid-points for $\beta_0 \in [-0.5, 0.5]$. Our jackknife AR confidence interval for 1530 instruments differs from that in Mikusheva and Sun (2022) because they used year-of-birth 1930-1938 dummies for the QOB-YOB-POB interactions, whereas we used 1930-1939 dummies. More details are provided in Section A.4 in the Appendix.

Table 1.2 highlights that the CIs generated by our jackknife CLC tests are the shortest among all the weak identification robust CIs (i.e., pp, krs, jackknife AR, jackknife LM, and two-step). Furthermore, the jackknife CLC CIs are 7.6% and 2.0% shorter than the non-robust JIVE-t CIs with 180 and 1,530 instruments, respectively, which is in line with our

[15] $\widetilde{F} = Q_{X,X}/\widehat{\Upsilon}$, where $\widehat{\Upsilon}$ is the cross-fit estimator. $Wald(\beta_0)$ is defined as $\left(\frac{\hat{\beta}-\beta_0}{\hat{V}}\right)^2$, where $\hat{\beta}$ is the JIVE estimator and $\hat{V}$ is a cross-fit estimator of the asymptotic variance of $\hat{\beta}$. We refer interested readers to Mikusheva and Sun (2022, Section 5) for more details.

theoretical result that the CLC tests are adaptive to the identification strength and efficient under strong identification.

Chapter 2

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

2.1 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 et al. (1999), Chao and Swanson (2005), Andrews and Stock (2007), Chao et al. (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 (2.2.1)). 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

[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]

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 instrumental variable (JIVE) estimator 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 for weak-identification-robust inference. On one hand the researcher could use the classical AR test for a fixed number of instruments (defined as $AR_{classical}$ in section 2.6.1), which has size control under a fixed number of instruments but has power deficit when the number of instruments is large (See Lemma B.2.5). On the other hand, the researcher could instead use the jackknifed AR tests proposed by Crudu et al. (2021) and Mikusheva and Sun (2022) (defined as $AR_{standard}$ and AR_{cf}, respectively, in section 2.6.1), which provides good size control whenever the number of instruments is large, but in general has size distortion when the number of instruments is small (e.g., see the discussions in Section 2.2.2). Since the two types of AR statistics are important components of many other weak-identification-robust test statistics proposed in the literature, we expect a similar non-uniformity issue for these statistics as well.

A simple simulation illustrates this issue.[3] Figure 2.1 demonstrates the case of a moderate number of instruments, with the number of instruments K equal to 15 and the sample size n equal to 200. In this paper, we propose four new tests that are robust to both weak identification and the number of instruments, two of which are denoted as $Q_{standard}$ and

[3]The tests in Figure 2.1 are simulated based on the design of section 2.6.2, except we have reduced the sample size from 400 to 200. The concentration parameter $\mathcal{G} \approx 70$. Note that using a different (higher or lower) concentration parameter does not change the size, shape, power-ranking, and percentage difference in power among the tests. In fact, $\mathcal{G} \approx 70$ was a result of $\pi_K \approx 0.25$, which is very small in practice.

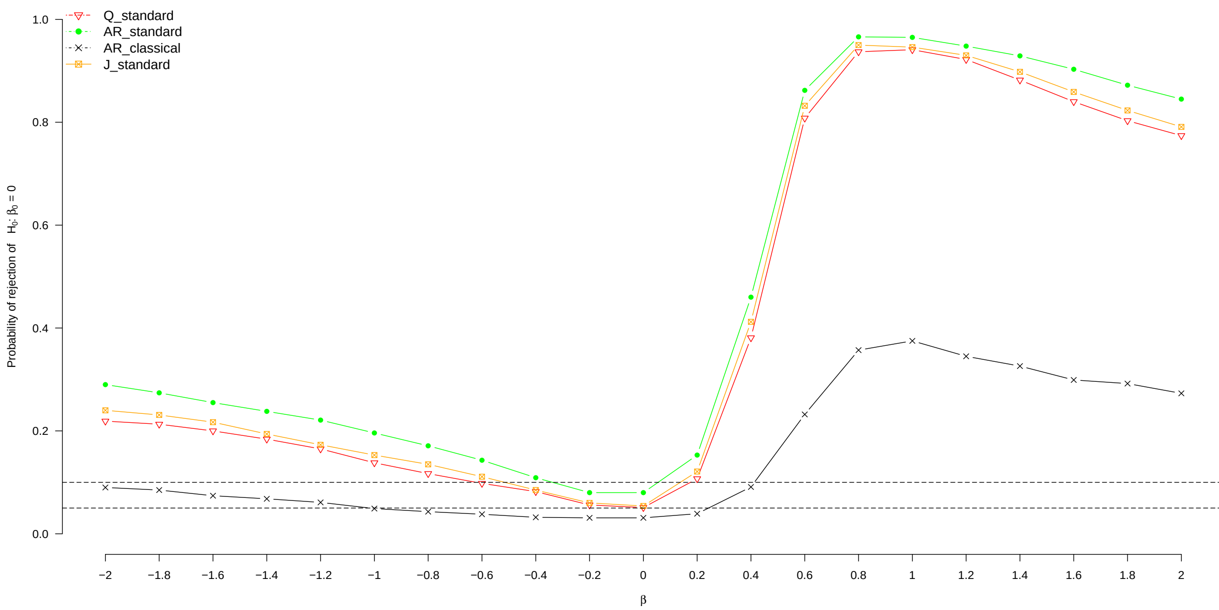

Figure 2.1: Power curve for $K = 15$ and $n = 200$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the green line with a colored-circle represents $AR_{standard}$; the black dotted line with 'x' represents $AR_{classical}$; the orange-line with colored-square represents $J_{standard}$. The first horizontal dotted black line represents 5%, while the second represents 10%.

$J_{standard}$ in Figure 2.1 (see section 2.6.1 for the detailed descriptions of these tests). At the true parameter $\beta = 0$, the null rejection probabilities of $AR_{classical}$, $AR_{standard}$, $Q_{standard}$, and $J_{standard}$ are 3.1%, 8%, 5.3%, and 5.4%, respectively. In addition, we observe that the power of $AR_{classical}$ is low throughout, while our tests $Q_{standard}$ and $J_{standard}$ have 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 controls grow slower than the fourth root of the number of instruments[4]. The analytical

[4] Chao, Swanson, and Woutersen (2023b) 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 6)

critical value defined in (2.2.8) is related to Anatolyev and Solvsten (2023),[5] and we extend their result to the problem of weak instruments.

Relation to the literature: Tests that allow for both fixed and diverging instruments dates back to Anatolyev and Gospodinov (2011). They proposed an estimator that is robust to the number of instruments, but requires errors to be homoskedastic. To improve finite sample performance Kaffo and Wang (2017) proposes bootstrapping as an alternative, although it relies on homoskedastic errors once again. Maurice J. G. Bun and Poldermans (2020) relaxes the assumption of homoskedastic errors but requires $Z_i e_i$ to be identically and independently distributed (i.i.d.), where Z_i is the instrument and e_i is the second-stage error. Relaxing the i.i.d. assumption, Boot and Ligtenberg (2023) proposed an estimator based on a continuous updating objective function (see their Corollary 2), but their approach relies on an invariance assumption on the second stage error term. Belloni et al. (2012) relaxes the i.i.d. and invariance assumption, but require the first-stage IV moment to be sparse. However, Kolesar, Muller, and Roelsgaard (2023) advised against making sparsity assumption whenever the number of instruments is less than the sample size. In contrast to the aforementioned approaches, our test procedure allow for heteroskedastic error but does not rely on invariance or sparsity assumption.

Structure of the paper: Section 2.2 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 2.3 provides a new strong approximation result for any 'AR-type' tests. Section 2.4 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. Note that the number of instruments is assumed to be

[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.

less than the sample size in sections 2.2–2.4 in order to simplify our discussion. Section 2.5 relaxes this and allow the number of instruments to be possibly larger than the sample-size. In particular, this section discusses the case of instruments being rank-deficient, and includes high-dimensional instruments as a special case. Section 2.6.2 provides simulation results for our power-curve based on calibrated data, which lends itself to our theory. Section 2.6.3 provides an application of our theory to empirical data. Proofs of Theorems, Lemmas, and Corollaries stated in the main text are given in Appendix B.1, while Auxiliary Lemmas are provided in Appendix B.2. In Appendix B.3 we provide details on the two estimators satisfying (2.2.12). In Appendix B.4 we discuss general limit problems under fixed and diverging instruments. Appendix B.5 provides more detail on the rank-deficiency procedure of Section 2.5.

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, and write WPA1 to mean 'with probability approaching one'. We define ι_i to be a vector of zeros, with value 1 only on the ith element. For any set S, we write S^c to mean the complement of the set, and use the symbol '$\otimes$' to denote Kronecker product. We write $\mathcal{Z}_K(J)$ to represent a standard Gaussian plus a constant $J \in \mathbb{R}^K$, i.e. $\mathcal{Z}_K(J) := \mathcal{N}(J, I_K)$. For any statistic T, denote $q_{1-\alpha}(T)$ to be the $(1-\alpha)$-quantile of the law of T.

2.2 Setup and Testing Procedure

2.2.1 Setup

Consider the model

$$\begin{aligned}\widetilde{Y} &= \widetilde{X}\beta + W\Gamma + \widetilde{e} \\ \widetilde{X} &= \widetilde{\Pi} + \widetilde{v}\end{aligned} \tag{2.2.1}$$

where $\widetilde{X} \in \mathbb{R}^{n\times d_X}$, $W \in \mathbb{R}^{n\times d_W}$, $\widetilde{Y}, \widetilde{e} \in \mathbb{R}^{n\times 1}$, and $\widetilde{Z} \in \mathbb{R}^{n\times K}$ is a full-rank matrix of instruments. $\widetilde{\Pi}_i \equiv \mathbb{E}(\widetilde{X}_i|\widetilde{Z}_i, W_i) \in \mathbb{R}^{1\times d_X}$ [6] Also, $\beta \in \mathbb{R}^{d_X}$ and $\Gamma \in \mathbb{R}^{d_W\times 1}$. We observe $(\widetilde{Y}, \widetilde{X}, W, \widetilde{Z})$, and assume that W is a full-ranked matrix of **exogenous** control variables 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.[7]

We are interested in testing

$$H_0 : \beta = \beta_0 \quad versus \quad H_1 : \beta \neq \beta_0, \tag{2.2.2}$$

where d_X, the dimension of β, is fixed. We aim to obtain a test that guarantees a correct size control irrespective of identification strength and asymptotic frameworks with regard to K and d_W. Specifically, our test remains valid no matter the instruments are strong or weak, and it remains valid no matter the dimensions of the instruments and control variables, K and d_W, are fixed or diverge to infinity as $n \to \infty$.[8] Whenever they do diverge, we allow K to grow at the same rate as the sample size n, while d_W must grow at a slower rate than n. For now we assume that $K < n$, but we will relax this assumption in Section 2.5.

[6] Note that assuming $\widetilde{Z}$ is of full rank implies that the number of instruments must be less than the sample-size

[7] See Remark 1

[8] The number of instruments K should be better written as K_n to reflect its dependence on sample size n, but we drop this notational dependence and simply write K whenever it does not cause confusion.

To proceed, we first partial out the exogenous control variables W (we give appropriate regularity conditions for d_W below) and rewrite the model as

$$\begin{aligned} Y &= X\beta + e \\ X &= \Pi + v \end{aligned} \tag{2.2.3}$$

where $Y = M_W \widetilde{Y}$, $X = M_W \widetilde{X}$, $\Pi = M_W \widetilde{\Pi}$, $e = M_W \widetilde{e}$, $v = M_W \widetilde{v}$, $Z = M_W \widetilde{Z}$, $M_W = I_n - P^W$, and $P^W := W(W'W)^{-1}W'$. 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 $P := Z(Z'Z)^{-1}Z'$.[9] We define $e_i(\beta_0) := Y - X\beta_0 = e + \Delta X$, where $\Delta := \beta - \beta_0$. Similarly, 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 of $e(\beta_0)$ 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 throughout the rest of the paper.

Assumption 5. *Suppose that the errors $(\widetilde{e}_i, \widetilde{v}_i)$ are mean zero and independent across i.*

Assumption 6 (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 that $\underline{C} \leq \lambda_{\min}(W'W/n) \leq \lambda_{\max}(W'W/n) \leq \overline{C}$ and Z has a full rank.*

We note that for a balanced-instrument design without control variables, $p_n = \frac{K}{n}$. Hence, for both fixed and diverging K, $\frac{p_n}{K} = \frac{1}{n} = o(1)$. Furthermore, $p_n \leq 1$ since each element on the diagonal of a projection matrix is always bounded by one. As mentioned above, we allow the number of controls d_W to diverge to infinity. However, in order for p_n^W to shrink to zero in Assumption 6, d_W must grow at a slower rate than n, i.e. $d_W = o(n)$, since $p_n^W \geq \frac{d_W}{n}$ by definition. In particular, we require that $d_W = O(K^{(1-\eta)/4})$ for any $\eta > 0$. Such an assumption ensures that we can strongly approximate our test statistics (see Theorem 1 and

[9] This implies that the partialled-out instrument matrix Z is full-ranked. In section 2.5 we discuss what to do in the event Z is not full-ranked.

the discussions after it). In the case of fixed K,

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

while in the case of diverging K,

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

2.2.2 Some Background and Motivations

In this section, we briefly discuss the general difficulties of constructing a weak-identification-robust test that achieves a simultaneous size control under both fixed and diverging number of instruments with heteroskedastic errors. First, let us consider the classical case of a fixed number of instruments and homoskedastic errors. For simplicity, we assume for the moment that control variables are not present in the model of (2.2.1). Under the null, a consistent estimator of the error variance σ^2 can be given by $\widehat{\sigma}^2 := \frac{1}{n}\sum_{i\in[n]} e_i^2$. Then, under standard regularity conditions, for the classical AR test statistic, we have

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

Now, consider the case of a diverging number of 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, $\frac{\sum_{i\in[n]} P_{ii}e_i^2}{K\widehat{\sigma}^2} = \frac{\sum_{i\in[n]} P_{ii}\sigma^2}{K\sigma^2} = \frac{\sum_{i\in[n]} P_{ii}}{K} = 1$ (See Lemma B.2.1). Therefore, we have

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

We observe that there are two distinct limiting distributions for the same (classical) statistic under two different scenarios for K. Indeed, for the case with diverging K, $e'Pe$ itself would diverge to infinity, so that the denominator K acts as a form of normalization. Such 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 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 K, a natural idea would be to focus on the jackknifed statistic, where the diagonals are removed, i.e.,

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

which converges weakly to a $\frac{\chi^2_K - K}{\sqrt{2K}}$-distribution under fixed K. On the other hand, 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 H_0 whenever

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

which has correct asymptotic size control no matter K is fixed or diverging, under homoskedasticity. However, under general heteroskedasticity, the problem becomes more complicated. To see why, suppose we have certain consistent variance estimator $\widehat{\Phi}_1(\beta_0)$ in the case with heteroskedastic errors so that under the null,[10]

$$\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_i e_j}{\sqrt{2K\widehat{\Phi}_1(\beta_0)}} \rightsquigarrow \mathcal{N}(0,1),$$

when $K \to \infty$. However, when K is fixed, the asymptotic distribution of this statistic is no longer $(\chi^2_K - K)/\sqrt{2K}$, resulting in size distortions in this case (this is also confirmed by our simulations in Section 2.6.2). Nevertheless, as we will explain in the next section, even under heteroskedastic errors, our proposed tests are able to provide a correct asymptotic size control simultaneously for both fixed and diverging numbers of instruments (and control variables).

[10]See section 2.2.5 for more details on this estimator

2.2.3 Analytical Tests

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

$$\widehat{Q}(\beta_0) := \frac{e(\beta_0)'Pe(\beta_0)}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} \tag{2.2.4}$$

Our analytical test compares the test statistic $\widehat{Q}(\beta_0)$ with a robust critical value $C_{\alpha,df}(\widehat{\Phi}_1(\beta_0))$, where $\alpha \in (0,1)$ is the significance level and under the null, $\widehat{\Phi}_1(\beta_0)$ is a consistent estimator of $\Phi_1(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$, with more details provided in section 2.2.5. We will reject $H_0 : \beta = \beta_0$ at the α significance level if

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

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 $\beta = \beta_0$, in which case $e_i(\beta_0)$ has mean zero and variance equal to $\sigma_i^2(\beta_0)$. When K is fixed, under regularity conditions, we can show that

$$\widehat{Q}(\beta_0) \rightsquigarrow \mathcal{Z}'D_n\mathcal{Z} = \sum_{k\in[K]} w_{n,i}\chi^2_{1,k}, \tag{2.2.5}$$

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

$$\Omega(\beta_0) := \frac{(Z'\Lambda(\beta_0)Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda(\beta_0)Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}, \tag{2.2.6}$$

where $\Lambda(\beta_0) = \text{diag}(\sigma_1^2(\beta_0), \cdots, \sigma_n^2(\beta_0))$, and $\{\chi^2_{1,k}\}_{k\in[K]}$ are K independent chi-squared random variables with 1 degree of freedom. $\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)$, the denominator of $\Omega(\beta_0)$, 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.

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

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

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-squared random variable with weights $\widetilde{w}_n$ (i.e., $F_{\widetilde{w}_n} = \sum_{i\in[K]} \widetilde{w}_{i,n}\chi^2_{1,i}$), 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

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

where $\Phi_1(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0)$ and df is our degree-of-freedom adjustment. In particular, df is some deterministic sequence such that[11]

$$df^{-1} = o(K^{-1/2}). \tag{2.2.7}$$

In fact, we allow df to take the value of ∞ so that $1/df$ can be taken to be zero. For generality we simply assume df satisfies (2.2.7). This degree-of-freedom correction is asymptotically negligible, but is included for better finite-sample performance.

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

$$C_{\alpha,df}(\widehat{\Phi}_1(\beta_0)) := 1 + \frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}\left(\frac{q_{1-\alpha}(F_{\widetilde{w}_n}) - 1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2 + 1/df}}\right). \tag{2.2.8}$$

[11] In our simulation (section 2.6.2), we let $df = (n-K)/2$. To see why this holds, note that by assumption 6, $\max_i P_{ii} \leq \delta < 1$, so that $\frac{K}{n} = \frac{\sum_{i\in[n]P_{ii}}}{n} \leq \delta < 1$. Therefore $K^{1/2}df^{-1} = 2\sqrt{\frac{1}{n/K-1}}\sqrt{\frac{1}{n-K}} \leq 2\sqrt{\frac{1}{1/\delta-1}}\sqrt{\frac{1}{n-K}} = O(1)\sqrt{\frac{1}{n-K}} = o(1)$, where the last equality follows from $n-K\to\infty$ since $\frac{K}{n}\leq\delta<1$.

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

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

$$q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2 + 1/df}} - 1\right). \tag{2.2.9}$$

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

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

implying that the second term in (2.2.9) is asymptotically negligible. This guarantees that our analytical test achieves the correct asymptotic size under fixed K as well.

2.2.4 Bootstrap Tests

The test statistic for our bootstrap tests is defined as

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij} e_i(\beta_0) e_j(\beta_0)}{\sqrt{K\widehat{\Phi}_1(\beta_0)}}, \tag{2.2.10}$$

with $\widehat{\Phi}_1(\beta_0)$ satisfying (2.2.12) and having the additional requirement that it can be constructed from using only $e(\beta_0)$ and P. The two variance estimators $\widehat{\Phi}_1(\beta_0)^{standard}$ and $\widehat{\Phi}_1(\beta_0)^{cf}$ discussed in section 2.2.5 satisfy this requirement. We reject $H_0 : \beta = \beta_0$ at the α significance level if

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L}),$$

where $C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L})$ is a bootstrap-based critical value that depends on (1) some large positive integer B, (2) the significance level α, (3) i.i.d. random variables $\{\kappa_i\}_{i\in[n]}$ following the probability law $\mathcal{L}$ with the property that its mean is zero, variance is one, fourth moment is bounded, and (4) the structure of the variance estimator $\widehat{\Phi}_1(\beta_0)$.

Specifically, the bootstrap 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 i.i.d. 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 as $\eta_{i,\ell} := \kappa_{i,\ell}e_i(\beta_0)$. 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, we can construct the bootstrap 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 $\ell \in [B]$, we obtain a collection of statistics $\{\widehat{J}^{BS,\ell}\}_{\ell\in[B]}$. Then

$$C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L}) := inf\left\{z \in \mathbb{R} : 1-\alpha \leq \frac{\sum_{\ell\in[B]} 1\left\{\widehat{J}^{BS,\ell} \leq z\right\}}{B}\right\} + 1/df_{BS} \quad (2.2.11)$$

where $df_{BS}^{-1} = o(1)$ is a deterministic sequence that is asymptotically negligible, but is included for better finite-sample performance.[12]

2.2.5 Estimators for $\Phi_1(\beta_0)$

In this section, 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

$$\widehat{\Phi}_1(\beta_0) = \Phi_1(\beta_0) + \mathcal{D}(\Delta) + o_p(1 + \sum_{i\in[4]} \Delta^i) \quad (2.2.12)$$

[12]In section 2.6.1 we take $df_{BS}^{-1} = (3\ log(n-K))/(n-K)$. To see that this is an $o(1)$ term, simply note that $n-K \to \infty$ by assumption 6, and apply L'Hopital rule. Furthermore, note that $\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0))$ has the same form as the jackkinfe AR statistics in Crudu et al. (2021) and Mikusheva and Sun (2022), which are asymptotically valid with standard normal critical values under diverging K. In this paper, we propose bootstrap tests for $\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0))$ and show the bootstrap validity under both diverging and fixed K.

where

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

and

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

We introduce two estimators that satisfy (2.2.12) under both fixed and diverging K (and d_W) – this is shown in Appendix B.3. The first variance estimator is due to Crudu et al. (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[13]

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

In order to reduce the bias of the variance estimator under the alternative, we further consider the cross-fit variance estimator due to Mikusheva and Sun (2022), which is defined as

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

where $M := I_n - Z(Z'Z)^{-1}Z'$ and $\widetilde{P}_{ij}^2 := \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$, and we show it also satisfies (2.2.12) under both fixed and diverging K (and d_W) in Appendix B.3. In particular, its corresponding asymptotic property as well as the expression of $\mathcal{D}^{cf}(\Delta)$ is provided in Theorem B.3.0.2.[14] 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

[13] This is shown in Theorem B.3.0.1

[14] 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.

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 (2.2.1), 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 control variables 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 (2.2.3). We show that the cross-fit variance estimator remains consistent under the null with the assumption that $p_n^W = \max_i P_{ii}^W = o(1)$.

2.3 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 (2.2.4) 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 5 holds and* $\sup_{i\in\mathbb{N}} \mathbb{E}(\widetilde{e}_i)^4 < \infty$*. Then we have*

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

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

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

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 6. 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.2.3, diverging K case).

2.4 Asymptotic Properties

2.4.1 Asymptotic Size

We discuss the size properties of our test in this section. We begin by making the following assumption, which ensures that we have uniform size-control.

Assumption 7. *Suppose* $p_n \leq \overline{C} \frac{K}{n}$ *for some* $\overline{C} < \infty$

Intuitively, Assumption 7 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 6, 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 ($\{\Lambda_n\}_{n\geq 1}$):

$$\left(\begin{array}{c} \{\widetilde{e}_i, \widetilde{v}_i\}_{i\in[n]} \text{ are independent, } \mathbb{E}\widetilde{e}_i = \mathbb{E}\widetilde{v}_i = 0, \\ \frac{p_n}{K} = o(1), p_n^W = o(1), d_W = O(K^{(1-\eta)/4}) \text{ for any } \eta > 0, \\ \max_i \Pi_i^2 + \max_i \mathbb{E}\widetilde{e}_i^8 + \max_i \mathbb{E}\widetilde{v}_i^8 \leq \overline{C} < \infty, \\ \Pi'\Pi, \sigma_i^2(\beta_0), \zeta_i^2(\beta_0) \geq \underline{C} \text{ under the null,} \\ \underline{C} \leq \lambda_{min}(\frac{W'W}{n}) \leq \lambda_{max}(\frac{W'W}{n}) \leq \overline{C}, \\ 0 \leq P_{ii} \leq \delta < 1, \\ \widehat{\Phi}_1(\beta_0) \text{ satisfies (2.2.12) under the null,} \\ \text{where } 0 < \underline{C}, \overline{C}, \delta < \infty \text{ are some fixed constants} \end{array}\right) \tag{2.4.1}$$

Then our test has size-control uniformly over the set of DGPs that satisfy (2.4.1). We formalize the statement as follows:

Theorem 2. *Suppose $\{\Lambda_n\}_{n\geq 1}$ satisfies (2.4.1), (2.2.7), and assumption 7 holds. Then under the null, for both fixed and diverging instruments, we have exact size-control for the proposed tests, i.e.*

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

and

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

Remark 1. *Note that Theorem 2 still holds when β is multivariate (instead of a scalar in (2.2.1)). 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 uniform size-control for any $\beta \in \mathbb{R}^{d_X}$ with fixed $d_X \geq 1$.*

2.4.2 Asymptotic Power

In this section we show that under strong identification, for both fixed and diverging K, our tests consistently differentiate the null from the alternative, where strong identification means $\mathcal{C} := Q_{\Pi,\Pi} \to \infty$. The concentration parameter $\mathcal{C}$ was introduced by Mikusheva and Sun (2022).[15] 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 $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$.[16] 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 $\mathcal{C} \to \infty$, which explains why $\mathcal{C}$ should be the right measure of identification strength. [17]

The Case with Diverging K

We want to evaluate the power of our test $\widehat{Q}(\beta_0)$ and $\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0))$ under different scenarios. In particular, we consider three cases for some sequence $d_n \to 0$: (1) Strong identification and local alternative, where $d_n \mathcal{C} = \widetilde{\mathcal{C}}$ 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 $d_n \mathcal{C} = \widetilde{\mathcal{C}}$ and $\Delta = \widetilde{\Delta}$; (3) Weak identification and fixed alternative, where $\mathcal{C} = \widetilde{\mathcal{C}}$ and $\Delta = \widetilde{\Delta}$.

Theorem 3. *Suppose Assumption 5, 6, 7, (2.2.7) and $\frac{\Pi'\Pi}{K} = O(1)$ holds. Then for any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (2.2.12), we have under strong identification and fixed alternative*

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

[15] Section B.4 provides more detail regarding the concentration parameter $\mathcal{C}$

[16] See Assumption 2 of their paper

[17] To see this, note that we can express the concentration parameter as $\mathcal{C} = \frac{\pi' Z' Z \pi}{\sqrt{K}} - \frac{\sum_{i\in[n]} P_{ii}(\pi' Z_i)^2}{\sqrt{K}}$, so that by assumption 6, $(1-\delta)\frac{\pi' Z' Z \pi}{\sqrt{K}} \le \mathcal{C} \le \frac{\pi' Z' Z \pi}{\sqrt{K}}$. We can then see that the order between $\frac{\pi' Z' Z \pi}{\sqrt{K}}$ and $\mathcal{C}$ are the same.

and

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}\left(\widehat{J}(\beta_0,\widehat{\Phi}_1(\beta_0)) > C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L})\right) = 1$$

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 necessarily zero, for diverging K we have that[18]

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

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

Theorem 4. *Suppose Assumption 5, 6, 7, (2.2.7) and* $\frac{\Pi'\Pi}{K} = O(1)$ *holds. Then 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*

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

and

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

where $F(\cdot)$ *denotes the cumulative distribution function (CDF) of a standard normal distribution. In particular, if we further assume* $\Pi' M\Pi \le \frac{\Pi'\Pi}{K} \to 0$*, then* $\widehat{\Phi}_1(\beta_0)$ *can be taken as* $\widehat{\Phi}^{\ell}_1(\beta_0)$ *for* $\ell = \{standard, cf\}$ *given in section 2.2.5.*

The assumption of $\frac{\Pi'\Pi}{K} \to 0$ automatically ensures that $\widehat{\Phi}^{standard}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$, while the additional requirement of $\Pi' M\Pi \le \frac{\Pi'\Pi}{K}$ is made to ensure that $\widehat{\Phi}^{cf}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$ 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.

[18]See the proof of Theorem 3

Theorem 5. *Suppose Assumption 5, 6, 7,* (2.2.7) *and* $\frac{\Pi'\Pi}{K} = O(1)$ *holds. Then for* $K \to \infty$ *and any estimator* $\widehat{\Phi}_1(\beta_0)$ *that satisfies (2.2.12), under strong identification and local alternative we have*

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

and

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

The Case with Fixed K

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 $\mathcal{C}$ defined for the diverging K case discussed earlier: $\widetilde{\mu}_n^2$ and $\mathcal{C}$ have the same order. To see this, note that under the 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 $\frac{\Pi'P\Pi}{\sqrt{K}} = \mathcal{C} + \frac{\sum_{i\in[n]}P_{ii}\Pi_i^2}{\sqrt{K}}$ has the same order as $\mathcal{C}$. Combining these facts yield the result that $\widetilde{\mu}_n^2$ has the same order as $\mathcal{C}$.

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 8. *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* $\Sigma(\Delta^\dagger)$ *is positive-semi-definite and* Q_{ZZ} *is positive-definite matrices. Furthermore, assume that* $\sup_i ||Z_i||_F < \infty$.

Under the assumption that the errors in the DGP of (2.2.1) 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 8 can be removed.

Recall from (2.2.9) that the power of our proposed test involves the critical value that is itself random. This randomness comes from the limit of the eigenvalues from $D_{\widetilde{w}_n} := diag(\widetilde{w}_{1,n}, ..., \widetilde{w}_{K,n})$. Since this is generally unknown, in order to show that our proposed tests consistently differentiates the null from the alternative whenever we have strong identification (under fixed instruments), under minimal assumptions, we begin by showing some intermediate asymptotic properties pertaining to the critical value (2.2.8).

Lemma 2.4.1. *Suppose Assumption 5, 6, 8 holds and we are under fixed* K. *Assume* (2.2.7) *holds and consider any estimator* $\widehat{\Phi}_1(\beta_0)$ *satisfying (2.2.12). Then for fixed* Δ *we have*

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

Under the alternative, for fixed K, the limiting distribution of the critical value $C_{\alpha,df}(\widehat{\Phi}_1(\beta_0))$ (see (2.2.8) for its expression) becomes that of a weighted chi-square $F_{w^{limit}}$-distribution.

Given that the limit w^{limit} 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 $w^{limit} = (w_1^{limit}, ..., w_K^{limit})$ such that $||w^{limit}||_F = 1$, $w_i^{limit} \geq 0$ and $q_{1-\alpha}(F_{w^{limit}})$ is the largest it can be. We have the following result due to Fleiss (1971):

Lemma 2.4.2. *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 (B.1.20), Lemma 2.4.1 and 2.4.2, 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 2.4.1 and 2.4.2, 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 5, 6, 8,* (2.2.7) *holds and we are under fixed K. For any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (2.2.12), our test consistently differentiates the null from alternative, i.e.*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L})\right) = 1$$

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, $\mu_K \equiv \widetilde{\mu}$,[19] 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 (2.2.6). We have the following result:

Theorem 7. *Suppose Assumption 5, 6, 8, (2.2.7) 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 any estimator* $\widehat{\Phi}_1(\beta_0)$ *that satisfies (2.2.12),*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}(\beta_0,\widehat{\Phi}_1(\beta_0)) > C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L})\right) = \mathbb{P}\left(\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)'\Omega^*(\beta_0)\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*}$$

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

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

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 proposed tests (i.e. analytical and bootstrap-based tests) have certain desirable properties; in particular, our tests are admissible within

[19]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.

some class of tests. Consider the test

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

Then we have the following result due to Marden (1982):

Corollary 2.4.1. *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 2.4.1 relates back to Theorem 7 in the sense that our proposed tests are admissible over the class of tests that are based on χ^2_1 or some combination of independent chi-squares (not necessarily a linear combination), under strong identification and local alternative. Finally, we can express the asymptotic power of our tests under weak identification and fixed alternative as follows:

Theorem 8. *Suppose Assumption 5, 6, 8, (2.2.7) holds and we are under fixed K. Assume $\Omega^*(\beta_0)$ is well defined and consider any estimator $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$. Then under weak identification and fixed alternative, if we further assume that $\Pi'\Pi = O(1)$, we have*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L})\right) = \mathbb{P}\left(\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right)'\Omega^*(\beta_0)\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*})\right)$$

where w^* are the eigenvalues of $\Omega^*(\beta_0)$. 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 2.2.5.

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

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

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 tests have power strictly greater than α. The asymptotic rejection criteria for both our tests can be written as

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

Analogous to Theorem 7, we have the result that under weak-identification and fixed-alternative, our tests are admissible within some class of tests. This follows from the following corollary.

Corollary 2.4.2. *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 $\overline{\Phi}_\alpha$.*

2.5 Rank-Deficiency and High-Dimensional Instruments

In this section we explore the problem of rank-deficiency in instruments (i.e. Z is not full-ranked). Under such rank-deficiency, the projection matrix $P := Z(Z'Z)^{-1}Z'$ is not well-defined. To overcome this, we consider the ridged-projection-matrix defined as

$$P_{\gamma_n} := Z(Z'Z + \gamma_n I_K)^{-1}Z$$

for some (sequence of) $\gamma_n \geq 0$. Following Dovi, Kock, and Mavroeidis (2023), we set the parameter γ_n to equal

$$\gamma_n^* := \max \arg\max_{\gamma_n \in \Gamma_n} \sum_{i\in[n]} \sum_{j\neq i} P^2_{ij,\gamma_n}$$

where $\Gamma_n := \{\gamma_n \in \mathbb{R} : \gamma_n \geq 0 \text{ if } r = K \text{ and } \gamma_n \geq \gamma_- > 0 \text{ if } r < K\}$ and $r := Rank(Z)$. We make the additional assumption to ensure that γ_n^* exists. In fact, whenever assumption 6 holds, assumption 9 will automatically hold,[20] so that assumption 9 is seen as a "generalized" version of the balanced-design assumption (i.e. $p_n \leq \delta < 1$).

Assumption 9 (Assumption 3 of Dovi et al. (2023))**.** *There exists constants $c, \gamma_- > 0$ not depending on n, some $h \geq 1$ and some sequence $\gamma_n \in [\overline{\gamma}, \infty)$ such that*

$$\sum_{i \in [n]} \sum_{j \neq i} P_{ij,\gamma_n}^2 \geq cr^h$$

where $\overline{\gamma} = 0$ if $r = K$ and $\overline{\gamma} = \gamma_-$ if $r < K$

Recall from sections 2.2.3–2.2.5 that the estimators involved depend on the number of instruments K. The reason is that we assumed the instruments have full rank (i.e. $r = K$). When instrument rank is deficient, we should focus instead on the rank of the instruments. In particular, we should replace P and K by P_{γ_n} and r respectively in the previous sections. Note that under these changes, our proposed analytical and bootstrap-based tests will once again control for size, even if the number of instruments exceed the sample-size. For clarity of exposition, we provide details of the testing procedure as well as its asymptotic properties in Appendix B.5

Remark 2. *Note that in section 2.2 we assumed that $\widetilde{Z}$ is of full-rank. This assumption implies that the number of instruments must be less than the sample size (i.e. $K < n$). Throughout the rest of section 2.5, however, we do not make this assumption. Instead, we focus on the rank-deficiency of partialled-out instrument Z. This allows for the number of instruments to be much larger than the sample size (i.e. $K >> n$), which includes the high-dimensional case.*

2.6 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

[20] In particular, we simply require $p_n \leq \delta < 1$ from assumption 6. See the proof of Proposition 1 in Dovi et al. (2023)

these tests and their corresponding critical-values.

2.6.1 Description of Tests

We consider the following tests, letting $df = (n-K)/2$, $df_{BS} = (n-K)/(3\, log(n-K))$, law $\mathcal{L}$ following a Rademacher distribution (i.e. equal probability of -1 and 1), and $\alpha = 0.05$ (i.e. 95% confidence level):

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

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

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

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

(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

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

(7) The bootstrapped-based test using $\widehat{\Phi}_1^{standard}(\beta_0)$ as variance estimator, which rejects whenever

$$\widehat{J}(\beta_0, \widehat{\Phi}_1^{standard}(\beta_0)) > C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L});$$

(8) The bootstrapped-based test using $\widehat{\Phi}_1^{cf}(\beta_0)$ as variance estimator, which rejects whenever

$$\widehat{J}(\beta_0, \widehat{\Phi}_1^{cf}(\beta_0)) > C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L}).$$

We denote the tests (1), (2), (3), (4), (5), (6), (7), (8) by $Q_{standard}$, Q_{cf}, AR_{cf}, $AR_{standard}$, $AR_{classical}$, JAR_{homo}, $J_{standard}$ and J_{cf} respectively.

2.6.2 Simulation Based on Hausman et al. (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

$$\begin{aligned}
&Y = \beta X + W\Gamma + D_{z_1} U_1 \\
&X = \pi_K z_1 + U_2 \\
&W = (1, ..., 1)' \in \mathbb{R}^n \\
&U_1 = \rho_1 U_2 + \sqrt{\frac{1-\rho_1^2}{\phi^2 + 0.86^4}}(\phi v_1 + 0.86 v_2), \\
&z_{i1} \sim \mathcal{N}(0.5, 1), \quad v_{1i} \sim z_{1i}(Beta(0.5, 0.5) - 0.5), \quad v_{2i} \sim \mathcal{N}(0, 0.86^2), \\
&D_{z_1} := diag(\sqrt{1 + z_{11}^2}, \sqrt{1 + z_{21}^2}, ..., \sqrt{1 + z_{n1}^2}) \\
&U_{2i} \sim exponential(0.2) - 5, \quad \phi = 0.3, \quad \rho_1 = 0.3
\end{aligned}$$

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 $\mathcal{C} \approx 70$.[21] 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 7 is not too large. We conduct $1,000$ simulation replications to obtain stable results and detail the probability of rejection under the null of $\beta = \beta_0$ in the following table.

Table 2.1 provides the probability of rejection under the null for different values of K;

[21] 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 $\mathcal{C} := \frac{(\pi z_1)' P_0 (\pi z_1)}{\sqrt{K}} = 70$ 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 $\mathcal{C}$ changes by a little, say ± 20.

Table 2.1: **Rejection Probability under Null**

	$AR_{standard}$ (5%)	$\mathbf{Q_{standard}}$ (5%)	AR_{cf} (5%)	$\mathbf{Q_{cf}}$ (5%)	$AR_{classical}$ (5%)	JAR_{homo} (5%)	$\mathbf{J_{standard}}$ (5%)	$\mathbf{J_{cf}}$ (5%)
$K = 1$	0.072	0.06	0.072	0.061	0.06	0.062	0.06	0.06
$K = 2$	0.079	0.054	0.08	0.055	0.046	0.054	0.048	0.049
$K = 3$	0.066	0.048	0.07	0.053	0.044	0.053	0.047	0.044
$K = 4$	0.08	0.058	0.086	0.065	0.052	0.068	0.052	0.053
$K = 5$	0.077	0.05	0.083	0.056	0.059	0.06	0.049	0.048
$K = 6$	0.08	0.061	0.128	0.099	0.053	0.098	0.059	0.061
$K = 8$	0.073	0.047	0.106	0.08	0.049	0.082	0.056	0.06
$K = 10$	0.073	0.05	0.098	0.082	0.047	0.081	0.051	0.055
$K = 15$	0.083	0.054	0.111	0.089	0.039	0.087	0.057	0.062
$K = 20$	0.07	0.048	0.10	0.069	0.04	0.079	0.051	0.052
$K = 40$	0.062	0.041	0.092	0.061	0.023	0.074	0.047	0.048
$K = 100$	0.048	0.035	0.075	0.058	0.001	0.068	0.046	0.045
$K = 200$	0.059	0.043	0.103	0.086	0	0.098	0.056	0.061
$K = 300$	0.066	0.065	0.134	0.131	0	0.125	0.056	0.067

Note: We reject at the 95% confidence-level, i.e. $\alpha = 0.05$

we make four observations. First, the $AR_{standard}$ suffers from size issues when the number of instruments is small-moderate. Our corresponding proposed tests $Q_{standard}$ and $J_{standard}$ resolves this. Second, severe size distortion also occurs for AR_{cf} under small-moderate amount of instruments;[22] our corresponding analytical test Q_{cf} tries to resolve this, albeit partially successful. However, notice that Q_{cf} reduces the size distortion by about 20%−30%. The bootstrap-based cross-fit test J_{cf} has more success in that size-distortion is mostly negligible, even when its counterpart AR_{cf} experiences severe size-distortion. Third, the

[22]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 B.3.0.2 that one of the key assumptions in assuring $\widehat{\Phi}_1^{cf}(\beta_0)$ satisfies (2.2.12) is that $p_n \leq \delta < 1$ for some $\delta > 0$. Note that even though this assumption was made in Theorem B.3.0.1, 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.

classical AR-test for fixed instruments $AR_{classical}$ generally does not suffer size-distortion for any number of instruments; however, we will see that it suffers from substantial power decline when the number of instruments is larger, say $K \geq 6$, as seen from Figure 2.4–2.8. Finally, JAR_{homo} suffers from size-distortion even for small instruments, say $K = 3$. This is to be expected since the critical value of JAR_{homo} is based on homoskedastic errors, while the errors of the DGP are heteroskedastic.

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. We then reject the tests whenever the test-statistic is greater than this null-computed quantile, i.e. we compute the size-corrected power.[23]

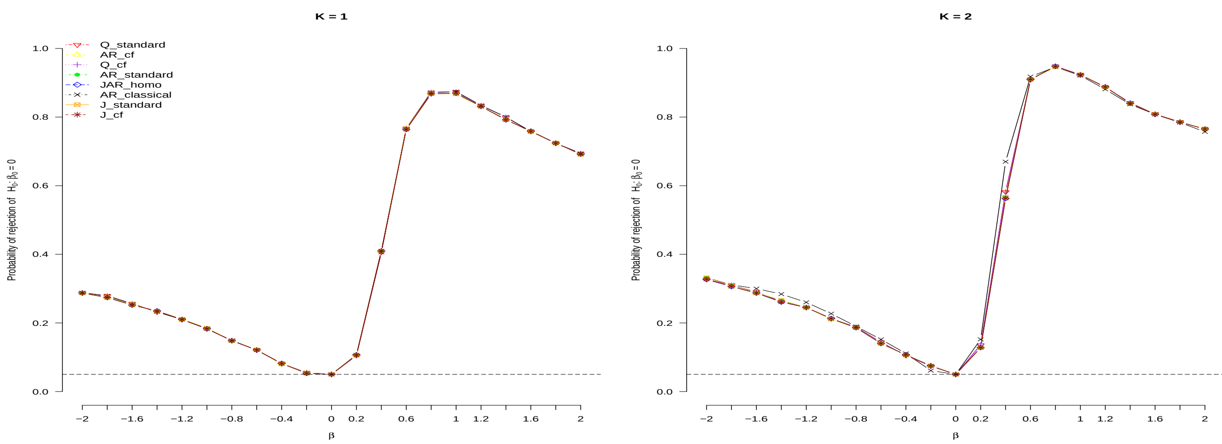

Figure 2.2: Power curve for $K = 1, 2$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

[23]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

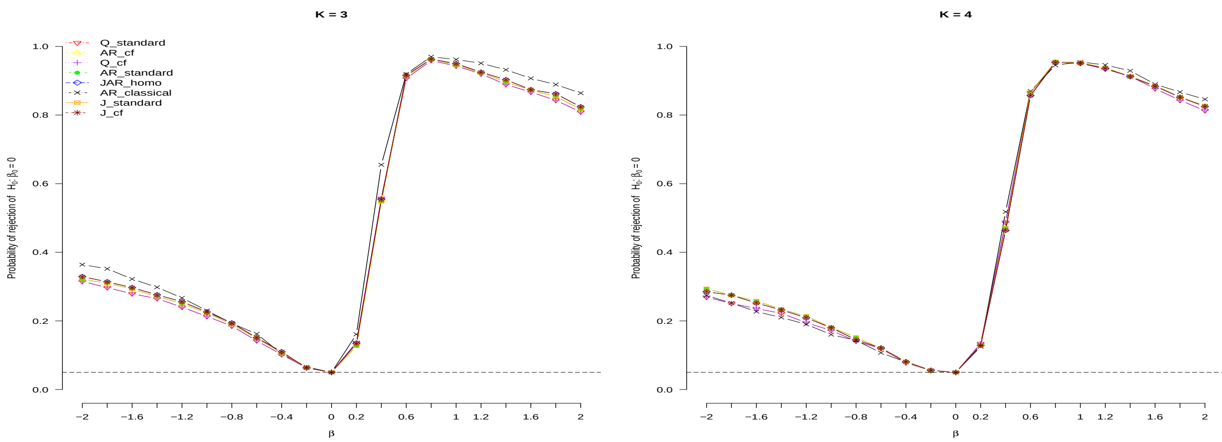

Figure 2.3: Power curve for $K = 3, 4$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

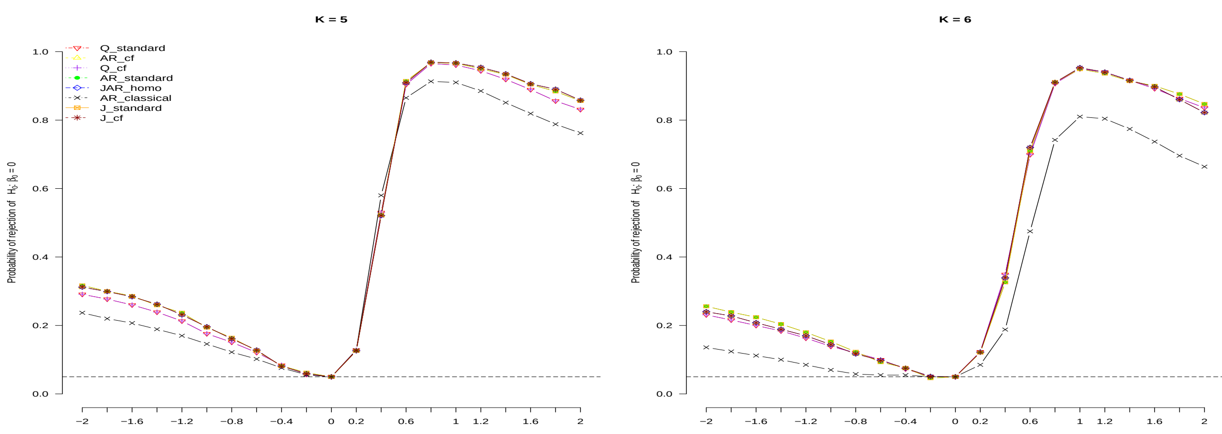

Figure 2.4: Power curve for $K = 5, 6$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

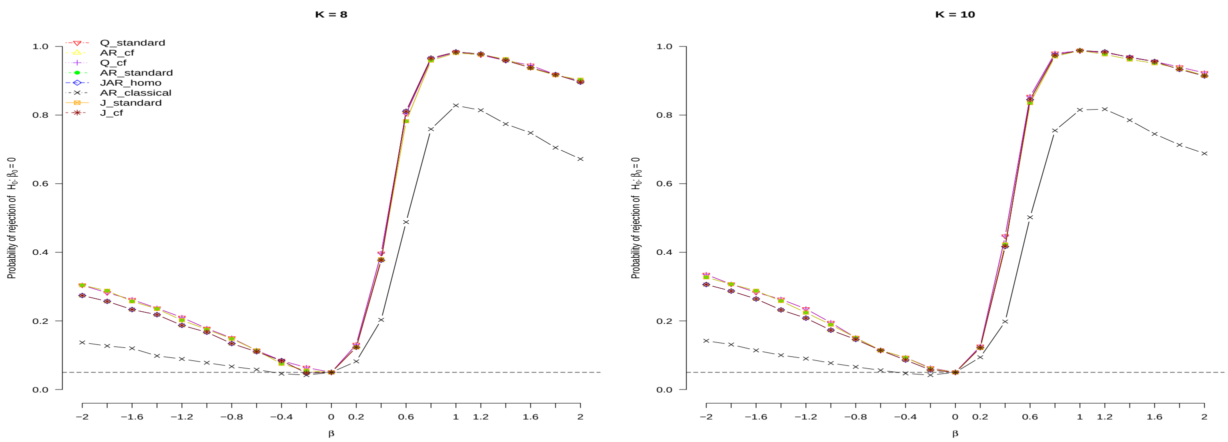

Figure 2.5: Power curve for $K = 8, 10$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

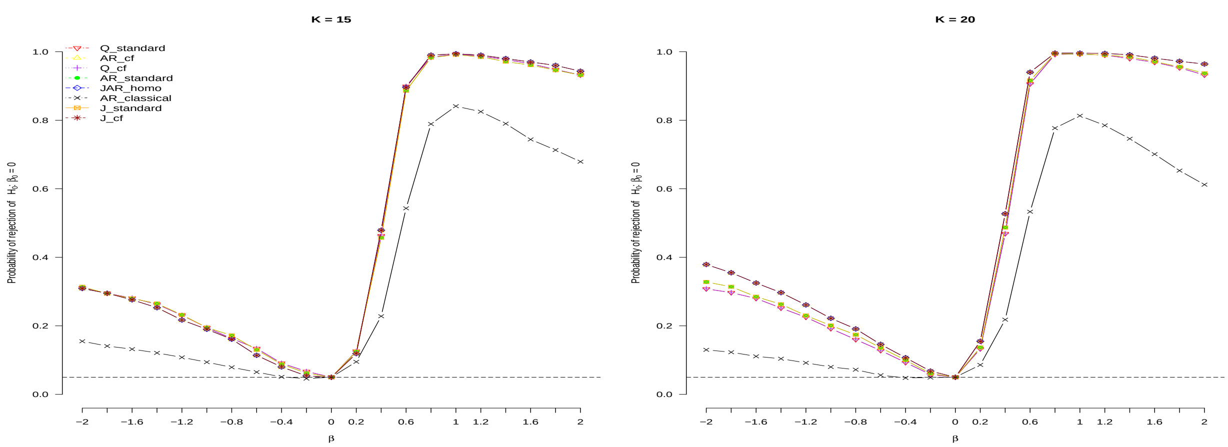

Figure 2.6: Power curve for $K = 15, 20$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

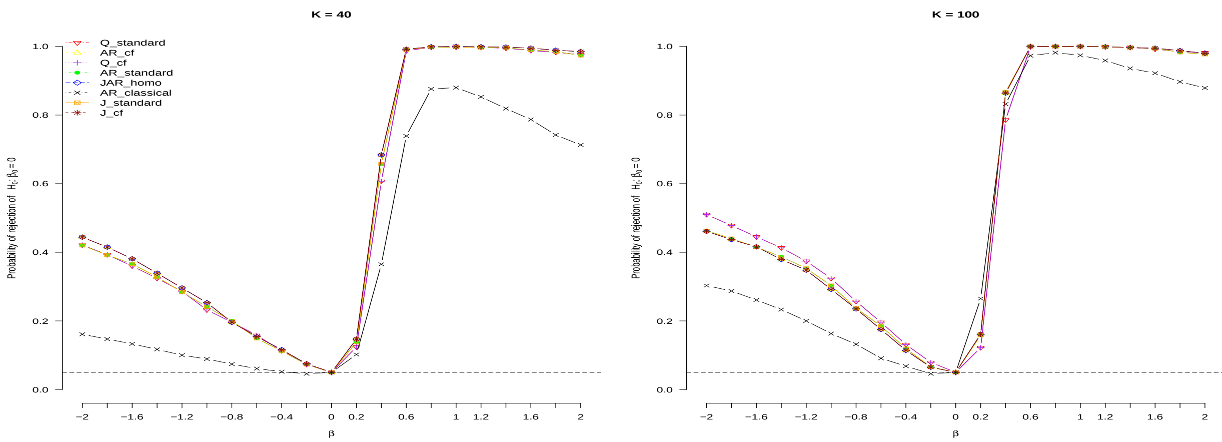

Figure 2.7: Power curve for $K = 40, 100$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

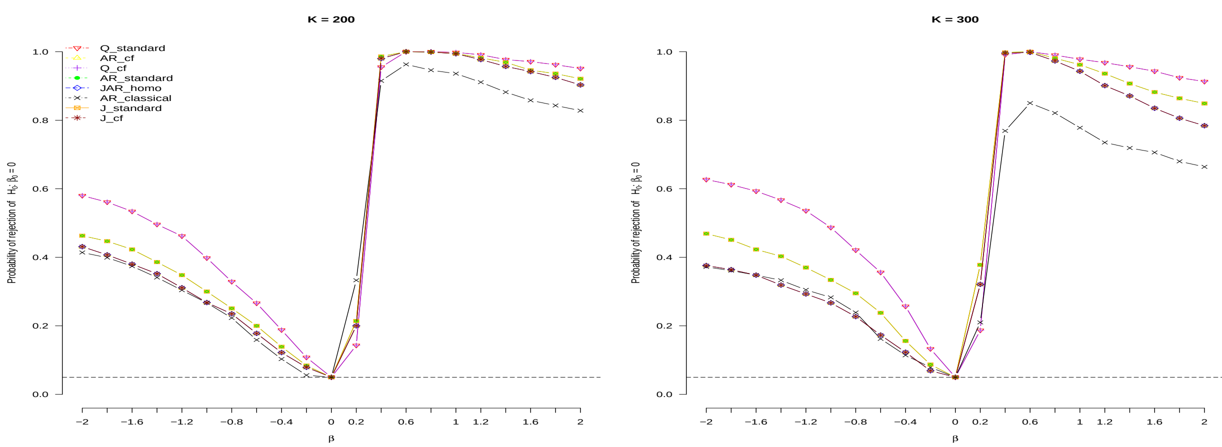

Figure 2.8: Power curve for $K = 200, 300$

Note: The red-line with downward-pointing triangle represents $Q_{standard}$; the yellow-line with a upward-pointing triangle represents AR_{cf}; the purple-line with a cross represents Q_{cf}; the green line with a colored-circle represents $AR_{standard}$; the blue dotted line with diamond represents JAR_{homo}; the black dotted line with an 'x' represents $AR_{classical}$; the orange-line with a colored-square represents $J_{standard}$; the dark-red dotted line with asterisk represents J_{cf}. The horizontal dotted black line represents 5%-level.

Figures 2.2-2.8 plot the size-adjusted power curve for the aforementioned tests; we highlight five observations. First, our four proposed tests $Q_{standard}$, Q_{cf}, $J_{standard}$ and J_{cf} have generally similar power over different number of instruments, which is expected as their rejection rate are asymptotically equal under every alternative. Second, the size-adjusted power of our proposed tests is at least as good as the well-known estimators $AR_{standard}$, AR_{cf}, $AR_{classical}$ and JAR_{homo} over varying numbers of instruments. Third, for moderate to large number of instruments (say $K \geq 6$), the power of the $AR_{classical}$ is comparatively lower than all other tests. Fourth, 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. Finally, for very large instruments ($K = 200, 300$), the size-adjusted power of $Q_{standard}$ and Q_{cf} are approximately equal, and dominates the other tests. The power of $AR_{standard}$ is approximately equal to $J_{standard}$, while the power of AR_{cf} is approximately equal to J_{cf}.

2.6.3 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.[24] 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

$$ln\ W_i = Constant + H_i^\top \zeta + \sum_{c=30}^{38} YOB_{i,c}\xi_c + \sum_{s \neq 56} POB_{i,s}\eta_s + \beta E_i + \gamma_i \qquad (2.6.1)$$

$$E_i = Constant + H_i^\top \lambda + \sum_{c=30}^{38} YOB_{i,c}\mu_c + \sum_{s \neq 56} POB_{i,s}\alpha_s + Z_{i,K} + \varepsilon_i$$

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

where W_i is the weekly wage, E_i is the education of the i-th individual, H_i is a vector of covariates,[25] $YOB_{i,c}$ 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 $s \in \{51 \; states\}$.[26] Both γ_i and ε_i are the error terms. We consider twenty-one varying numbers of instruments; in particular,

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

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

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

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

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

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

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

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

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

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

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

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

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 2.6.1. Specifically, we can write the above model as

$$ln\ W_i = C_i\Gamma + \beta E_i + \gamma_i \tag{2.6.2}$$

$$E_i = C_i\tau + Z_i\Theta + \varepsilon_i, \tag{2.6.3}$$

where C_i is a (329,509×71)-matrix of controls containing the first four terms on the right-hand of (2.6.1). We can then partial out the controls C_i by multiplying each equation (2.6.2) and (2.6.3) by the residual matrix $I - C(C^\top C)^{-1}C^\top$ to obtain a form analogous to that in the main text:

$$\begin{aligned} Y_i &= X_i\beta + e_i, \\ X_i &= \Pi_i + v_i \end{aligned}$$

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 2.6.1). 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 2.2 below.

Table 2.2: **Confidence Interval**

	$AR_{standard}$ (5%)	$\mathbf{Q_{standard}}$ (5%)	$AR_{classical}$ (5%)	JAR_{homo} (5%)	$J_{standard}$ (5%)
$K = 3$	[0.056,0.147]	[0.052,0.151]	[0.053,0.151]	[0.052,0.151]	[0.052,0.15]
$K = 10$	[-0.007,0.16]	[-0.011,0.165]	[-0.011,0.166]	[-0.011,0.165]	[-0.011,0.167]
$K = 20$	[0.017,0.174]	[0.015,0.178]	[0.014,0.18]	[0.014,0.178]	[0.009,0.183]
$K = 30$	[0,0.169]	[-0.002,0.172]	[-0.002,0.177]	[-0.002,0.172]	[-0.004,0.173]
$K = 50$	[0.005,0.183]	[0.002,0.188]	[-0.01,0.188]	[0.002,0.188]	[0.188,0,0.198]
$K = 100$	[0.018,0.2]	[0.017,0.202]	[0.009,0.203]	[0.017,0.202]	[0.013,0.203]
$K = 150$	[0.023,0.208]	[0.022,0.21]	[0.022,0.212]	[0.022,0.21]	[0.021,0.211]
$K = 180$	[0.008,0.201]	[0.007,0.202]	[0.007,0.207]	[0.007,0.202]	[0.005,0.206]
$K = 200$	[-0.216,0.23]	[-0.223,0.233]	[-0.214,0.236]	[-0.224,0.233]	[-0.131,0.252]
$K = 250$	[-0.118,0.258]	[-0.122,0.261]	[-0.111,0.256]	[-0.122,0.261]	[-0.1,0.275]
$K = 300$	[-0.097,0.24]	[-0.1,0.242]	[-0.085,0.238]	[-0.1,0.242]	[-0.092,0.26]
$K = 350$	[-0.107,0.28]	[-0.11,0.283]	[-0.092,0.274]	[-0.11,0.283]	[-0.071,0.273]
$K = 400$	[-0.078,0.305]	[-0.081,0.308]	[-0.058,0.298]	[-0.081,0.308]	[-0.076,0.257]
$K = 450$	[-0.105,0.29]	[-0.107,0.293]	[-0.092,0.281]	[-0.107,0.293]	[-0.047,0.25]
$K = 600$	[-0.018,0.228]	[-0.019,0.229]	[-0.013,0.224]	[-0.019,0.229]	[-0.011,0.231]
$K = 765$	[-0.09,0.192]	[-0.093,0.194]	[-0.125,0.163]	[-0.092,0.194]	[-0.108,0.201]
$K = 918$	[-0.055,0.182]	[-0.058,0.183]	[-0.076,0.157]	[-0.056,0.183]	[-0.064,0.19]
$K = 1071$	[-0.042,0.19]	[-0.044,0.192]	[-0.064,0.168]	[-0.042,0.191]	[-0.05,0.196]
$K = 1224$	[-0.035,0.209]	[-0.036,0.208]	[-0.052,0.186]	[-0.035,0.209]	[-0.042,0.231]
$K = 1377$	[-0.034,0.207]	[-0.036,0.209]	[-0.052,0.186]	[-0.035,0.208]	[-0.042,0.231]
$K = 1530$	[-0.035,0.219]	[-0.036,0.221]	[-0.049,0.206]	[-0.035,0.22]	[-0.038,0.229]

Note: We reject at the 95% confidence-level, i.e. $\alpha = 0.05$

We have omitted AR_{cf}, Q_{cf} and J_{cf} from the Table 2.2 because the confidence interval of these tests are either very similar or exactly the same as $AR_{standard}$, $Q_{standard}$ and $J_{standard}$ respectively. Therefore, we can speak of the confidence interval (C.I) for the aforementioned tests interchangeably (e.g. when we mention the C.I. of AR_{cf}, we also mean the C.I. of

$AR_{standard}$). We now make a few observations, which we discuss in detail. First of all, recall from Table 2.1 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 this 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. In fact, we can also expect the C.I. of $AR_{standard}$ to be close to AR_{cf} over all values of instruments, which holds true. Second, the C.I. of $AR_{classical}$ is quite different from all other statistics for larger instruments, which is to be expected since $AR_{classical}$ is meant for testing under fixed instruments. However, notice that the C.I. of $Q_{standard}$ (and therefore Q_{cf}) is close to $AR_{classical}$ for smaller instruments, while $Q_{standard}$ 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 instruments. For large instruments (say $K \geq 350$), the C.I. of $Q_{standard}$ (and Q_{cf}) 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 that of AR_{cf} 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 instruments, 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[27]

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

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

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

[27]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.2.3

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

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

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

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

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 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.

Appendix A

Technical Results for Chapter 1

A.1 Exogenous Control Variables

Suppose we observe $\{\tilde{Y}_i, \tilde{X}_i, \tilde{Z}_i, W_i\}_{i\in[n]}$, where

$$\tilde{Y}_i = \tilde{X}_i\beta + W_i^\top\gamma + \tilde{e}_i, \quad \tilde{X}_i = \tilde{\Pi}_i + \tilde{V}_i,$$

$\tilde{X}_i \in \Re$, $\tilde{Z}_i \in \Re^K$, $W_i \in \Re^d$, $\tilde{\Pi}_i = \mathbb{E}\tilde{X}_i$, and $(\tilde{Z}_i, W_i)_{i\in[n]}$ are treated as fixed. We allow K to diverge to infinity with n while d is fixed. We then have $\mathbb{E}\tilde{e}_i = \mathbb{E}\tilde{V}_i = 0$. Denote $P_W = W(W^\top W)^{-1}W^\top$ and $M_W = I_n - P_W$ be the projection and residual matrices based on W, respectively, where I_n is the $n \times n$ identity matrix and $W = (W_1, W_2, \cdots, W_n)^\top \in \Re^{n\times d}$. Further denote $\tilde{Y}, \tilde{X}, \tilde{e}, \tilde{\Pi}, \tilde{V}$ as matrices with their ith row being $\tilde{Y}_i, \tilde{X}_i, \tilde{e}_i, \tilde{\Pi}_i, \tilde{V}_i$, respectively. Then, we have

$$Y_i = X_i\beta + e_i, \quad X_i = \Pi_i + V_i,$$

where $Y = M_W\tilde{Y}$, $X = M_W\tilde{X}$, $V = M_W\tilde{V}$, $e = M_W\tilde{e}$, $\Pi = M_W\tilde{\Pi}$, and $Z = M_W\tilde{Z}$. We still denote P as the projection matrix constructed by Z. The next theorem shows Assumption 1 holds.

Theorem A.1.1. *Suppose* $\{\tilde{V}_i, \tilde{e}_i\}_{i\in[n]}$ *are independent,* $\max_i \mathbb{E}\tilde{e}_i^4 + \max_i \mathbb{E}\tilde{V}_i^4 \leq C < \infty$, $\max_i ||W_i||_2 \leq C < \infty$, $\Pi^\top\Pi/K = O(1)$, *and* $0 < c \leq mineig(W^\top W/n) \leq maxeig(W^\top W/n) \leq C < \infty$, *for some constants* c, C. *Then, Assumption 1 holds and* $Q_{e,e} = Q_{\tilde{e},\tilde{e}} + o_P(1)$. *If* *in addition,* $p_n^2 \frac{\Pi^\top \Pi}{K} = o(1)$ *with* $p_n = \max_i P_{ii}$*, then we have* $Q_{X,e} = Q_{\overline{X},\tilde{e}} + o_P(1)$ *and* $Q_{X,X} = Q_{\overline{X},\overline{X}} + o_P(1)$*, where* $\overline{X}_i = \Pi_i + \tilde{V}_i$*.*

Theorem A.1.1 shows Assumption 1 still holds if (Y_i, X_i, Z_i) are defined after partialing out the fixed dimensional control variables W_i. It further provides a sufficient condition under which the effect of partialling-out on the sampling error is asymptotically negligible, i.e., the asymptotic covariance matrix remains the same after partialing out W_i. To interpret the sufficient condition, we consider the balanced design in which p_n is of order K/n. If $K/n = o(1)$ and $\Pi^\top \Pi / n = O(1)$, then the sufficient condition holds because

$$p_n^2 \Pi^\top \Pi / K = O\left(\frac{\Pi^\top \Pi}{n} \frac{K}{n}\right) = o(1).$$

On the other hand, if $K \asymp n$, the sufficient condition requires $\Pi^\top \Pi / K = o(1)$, which can hold under both weak identification ($\Pi^\top \Pi / \sqrt{K} = O(1)$) and strong identification ($\Pi^\top \Pi / \sqrt{K} \to \infty$). We further emphasize that, even if $K \asymp n$ and $\Pi^\top \Pi / K \asymp 1$ so that the sufficient condition does not hold, Assumption 1 still holds. It is just that partialing out the exogenous control variable will have a non-negligible effect on the asymptotic covariance of $(Q_{e,e}, Q_{X,e}, Q_{X,X} - Q_{\Pi,\Pi})$.

A.2 Verifying Assumption 2

A.2.1 Standard Estimators

In this section, we maintain Assumption 10, which is stated below and just Mikusheva and Sun (2022, Assumption 1).

Assumption 10. *Suppose* $\{V_i, e_i\}_{i \in [n]}$ *are independent and* $\mathbb{E} e_i = \mathbb{E} V_i = 0$*. Suppose* P *is an* $n \times n$ *projection matrix of rank* K*,* $K \to \infty$ *as* $n \to \infty$ *and there exists a constant* δ *such that* $P_{ii} \leq \delta < 1$*.*

Following the results in Chao et al. (2012) and Mikusheva and Sun (2022), we can show that under either weak or strong identification, Assumption 1 in the paper holds:

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix} \right), \tag{A.2.1}$$

where $\sigma_i^2 = \mathbb{E}e_i^2$, $\eta_i^2 = \mathbb{E}V_i^2$, $\gamma_i = \mathbb{E}e_iV_i$, $\omega_i = \sum_{j \neq i} P_{ij}\Pi_j$,

$$\Phi_1 = \lim_{n\to\infty} \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \sigma_i^2 \sigma_j^2,$$

$$\Phi_{12} = \lim_{n\to\infty} \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\gamma_j \sigma_i^2 + \gamma_i \sigma_j^2),$$

$$\Phi_{13} = \lim_{n\to\infty} \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \gamma_i \gamma_j,$$

$$\Psi = \lim_{n\to\infty} \left[\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\eta_i^2 \sigma_j^2 + \gamma_i \gamma_j) + \frac{1}{K} \sum_{i\in[n]} \omega_i^2 \sigma_i^2 \right],$$

$$\tau = \lim_{n\to\infty} \left[\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \eta_i^2 \gamma_j + \frac{2}{K} \sum_{i\in[n]} \omega_i^2 \gamma_i \right], \quad \text{and}$$

$$\Upsilon = \lim_{n\to\infty} \left[\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \eta_i^2 \eta_j^2 + \frac{4}{K} \sum_{i\in[n]} \omega_i^2 \eta_i^2 \right].$$

We note that the standard estimators of the above variance components proposed by Crudu et al. (2021) are equal to Chao et al.'s (2012) estimators with their residual $\hat{e}_i$ replaced by $e_i(\beta_0)$. Specifically, let

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

$$\widehat{\Phi}_{12}(\beta_0) = \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (X_j e_j(\beta_0) e_i^2(\beta_0) + X_i e_i(\beta_0) e_j^2(\beta_0)),$$

$$\widehat{\Phi}_{13}(\beta_0) = \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 X_i e_i(\beta_0) X_j e_j(\beta_0),$$

$$\widehat{\Psi}(\beta_0) = \frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i} P_{ij}X_j)^2 e_i^2(\beta_0) + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 X_i e_i(\beta_0) X_j e_j(\beta_0),$$

$$\widehat{\tau}(\beta_0) = \frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i} P_{ij}X_j)^2 X_i e_i(\beta_0) + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 X_i^2 X_j e_j(\beta_0), \quad \text{and}$$

$$\widehat{\Upsilon} = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 X_i^2 X_j^2.$$

Assumption 11. *Suppose* $\max_{i\in[n]}|\Pi_i| \leq C$, $\frac{\Pi^\top\Pi}{K} = o(1)$, *and* $\max_i \mathbb{E}e_i^6 + \max_i \mathbb{E}V_i^6 < \infty$.

Two remarks on Assumption 11 are in order. First, $\max_{i\in[n]}|\Pi_i| \leq C$ is mild because $\Pi_i = \mathbb{E}X_i$. Second, Assumption 11 allows for weak identification when $\Pi^\top\Pi/\sqrt{K} \to c$ for a constant c. It also allows for strong identification when $\Pi^\top\Pi/\sqrt{K} \to \infty$ and $\Pi^\top\Pi/K \to 0$. The restriction that $\Pi^\top\Pi/K \to 0$ is needed because Assumption 2 includes the case of fixed alternatives (i.e., fixed $\Delta \neq 0$), which is not considered in Crudu et al. (2021) and Chao et al. (2012). Furthermore, our results include $\widehat{\tau}(\beta_0)$ and $\widehat{\Upsilon}$, which are not considered in Crudu et al. (2021) and Chao et al. (2012), and the consistency of these terms require $\Pi^\top\Pi/K \to 0$.

Theorem A.2.1. *Suppose Assumptions 10 and 11 hold. Then Assumption 2 holds for Crudu et al.'s (2021) estimators defined above.*

A.2.2 Cross-Fit Estimators

Let $M = I - P$, M_{ij} be the (i,j) element of M, M_i be the ith row of M, and $\widetilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$. Then, Mikusheva and Sun (2022) consider the cross-fit estimators for $\Phi_1(\beta_0)$, $\Psi(\beta_0)$, and Υ defined as

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

$$\widehat{\Psi}(\beta_0) = \frac{1}{K}\left[\sum_{i\in[n]}(\sum_{j\neq i} P_{ij}X_j)^2 \frac{e_i(\beta_0)M_i e(\beta_0)}{M_{ii}} + \sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_i X e_i(\beta_0) M_j X e_j(\beta_0)\right], \quad \text{and}$$

$$\widehat{\Upsilon} = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2[X_i(\beta_0)M_i X][X_j(\beta_0)M_j X],$$

where X and $e(\beta_0)$ are the column vectors that collect all X_i and $e_i(\beta_0)$, respectively. Following their lead, we can construct the cross-fit estimators for the rest three elements in $\gamma(\beta_0)$ as follows:

$$\widehat{\Phi}_{12}(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(M_jXe_j(\beta_0)e_i(\beta_0)M_ie(\beta_0) + M_iXe_i(\beta_0)e_j(\beta_0)M_je(\beta_0)),$$

$$\widehat{\Phi}_{13}(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_iXe_i(\beta_0)M_jXe_j(\beta_0), \quad \text{and}$$

$$\widehat{\tau}(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(X_iM_iX)(M_jXe_j(\beta_0)) + \frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i}P_{ij}X_j)^2\left(\frac{e_i(\beta_0)M_iX}{2M_{ii}} + \frac{X_iM_ie(\beta_0)}{2M_{ii}}\right),$$

Assumption 12. *Suppose Assumption 11 holds. Further suppose that* $\Pi^\top M\Pi \leq \frac{C\Pi^\top\Pi}{K}$ *for some constant* $C > 0$.

Compared with the assumptions in Mikusheva and Sun (2022), Assumption 12 further requires that $\max_{i\in[n]}|\Pi_i| \leq C$. However, for all the above cross-fit estimators to be consistent, we only need $\Pi^\top\Pi/K \to 0$, which is weaker than that assumed in Mikusheva and Sun (2022) (e.g., Theorems 5 in their paper require $\Pi^\top\Pi/K^{2/3} \to 0$).

Lemma A.2.1. *Suppose Assumptions 10 and 12 hold. Then, Lemmas 2, 3, S3.1, S3.2 in Mikusheva and Sun (2022) hold.*

Theorem A.2.2. *Suppose Assumptions 10 and 12 hold. Then, Assumption 2 holds for Mikusheva and Sun's (2022) cross-fit estimators defined above.*

A.3 Details for Simulations Based on Calibrated Data

The DGP contains only the intercept as the control variable. Therefore, we implement our jackknife CLC test on the demeaned version of $(\tilde{y}_i, \tilde{s}_i, \tilde{Z}_i)$. The parameter space is $\mathcal{B} = [-0.5, 0.5]$. We test the null hypothesis that $\beta = \beta_0$ for $\beta_0 = 0.1$ while varying the true value β over 31 equal-spaced grids over $\mathcal{B}$. The grids for δ is the grid for β minus β_0. We generate grids of (a_1, a_2) as $a_1 = \sin^2(t_1)$ and $a_2 = \cos^2(t_1)\sin^2(t_2)$ with t_1 taking values over 16 equal-spaced grids over $[\underline{a}^{1/2}(f_s(\widehat{D}, \widehat{\gamma}(\beta_0)), \pi/2]$ and t_2 taking values over 16 equal-spaced grids over $[0, \pi/2]$. We gauge $\mathbb{E}^*\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$ via a Monte Carlo integration with

$R = 2000$ draws of independent standard normal random variables. In practice, it is rare but possible that $\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))$ defined in (1.3.8) is not unique. To increase numerical stability, we follow I.Andrews (2016) and allow for some slackness in the minimization. Let $\mathcal{G}_a$ be the grid of (a_1, a_2) mentioned above, $\widehat{Q}(a_1, a_2) = \sup_{\delta \in \mathcal{D}(\beta_0)}(\mathcal{P}_{\delta,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}^* \phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0)))$, $\widehat{Q}_{\min} = \min_{(a_1,a_2)\in\mathcal{G}_a} \widehat{Q}(a_1, a_2) + 1/n$, where n is the sample size, and

$$\Xi = \{(a_1, a_2) \in \mathcal{G}_a : \widehat{Q}(a) \leq \widehat{Q}_{\min} + (\widehat{Q}_{\min}(1 - \widehat{Q}_{\min}))^{1/2}(2\log(\log(R)))^{1/2}R^{-1/2}\}.$$

The slackness term in the definition of Ξ is due to the law of the iterated logarithm for sum of Bernoulli random variables and captures the randomness of the Monte Carlo integration. Suppose there are L elements in Ξ with an ascending order w.r.t. (t_1, t_2), which are denoted as $\{(a_{1,l}, a_{2,l})\}_{l=1}^{L}$. We then define $\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))$ as $(a_{1,\lfloor L/2 \rfloor}, a_{2,\lfloor L/2 \rfloor})$. We use the cross-fit estimators defined in Section A.2.2 throughout the simulation.

A.4 Details for Empirical Application

We consider the 1980s census of 329,509 men born in 1930-1939 based on Angrist and Krueger's (1991) dataset. The model for **180 instruments** follows Mikusheva and Sun (2022), which can be written explicitly as

$$\begin{aligned}
ln\ W_i &= Constant + H_i^\top \zeta + \sum_{c=30}^{38} YOB_{i,c}\xi_c + \sum_{s\neq 56} POB_{i,s}\eta_s + \beta E_i + \gamma_i \\
E_i &= Constant + H_i^\top \lambda + \sum_{c=30}^{38} YOB_{i,c}\mu_c + \sum_{s\neq 56} POB_{i,s}\alpha_s \\
&\quad + \sum_{j=1}^{3}\sum_{s\neq 56} QOB_{i,j}POB_{i,s}\delta_{c,s} + \sum_{j=1}^{3}\sum_{c=30}^{39} QOB_{i,j}YOB_{i,c}\theta_{j,c} + \varepsilon_i,
\end{aligned}$$

where W_i is the weekly wage, E_i is the education of the i-th individual, H_i is a vector of covariates,[1] $YOB_{i,c}$ 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

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

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 $s \in \{51\ states\}$.[2] Both γ_i and ε_i are the error terms. The coefficient β is the return to education. We vary this β across 10,000 equidistant grid-points from -0.5 to 0.5 (i.e., $\beta \in \{-0.5, -4.9999, -4.9998, ..., 0, ..., 4.9999, 0.5\}$) and solve for the range of β where the null hypothesis cannot be rejected. Specifically, we can write the above model as

$$\begin{aligned} ln\ W_i &= C_i\Gamma + \beta E_i + \gamma_i \\ E_i &= C_i\tau + Z_i\Theta + \varepsilon_i, \end{aligned}$$

where C_i is a (329,509×71)-matrix of controls containing the first four terms on the right-hand of the first equation, while Z_i is the (329,509×180)-matrix of instruments containing the first two terms in the third line. We can then partial out the controls C_i by multiplying each equation by the residual matrix $I - C(C^\top C)^{-1}C^\top$ to obtain a form analogous to that in the main text:

$$\begin{aligned} Y_i &= X_i\beta + e_i, \\ X_i &= \Pi_i + v_i. \end{aligned}$$

Then, at each grid-point we take $\beta_0 = \beta$ and compute $AR(\beta_0)$, $LM(\beta_0)$, $Wald(\beta_0)$, $\widehat{\phi}_{\mathcal{A}_{pp}(\widehat{D},\widehat{\gamma}(\beta_0))}$ and $\widehat{\phi}_{\mathcal{A}_{krs}(\widehat{D},\widehat{\gamma}(\beta_0))}$. We reject the chosen value of β_0 for $AR(\beta_0)$ if it exceeds the one-sided 5%-quantile of the standard normal (i.e., reject if $AR(\beta_0) > z_{0.05}$). If $LM(\beta_0)^2 > \mathbb{C}_{0.05}$, we reject the chosen β_0 for Jackknife LM. If $Wald(\beta_0) > \mathbb{C}_{0.05}$, we reject for JIVE-t. If $\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} > \mathbb{C}_{0.05}(\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0)); \widehat{\rho}(\beta_0))$ for $s \in \{pp, krs\}$, we reject accordingly. The two-step procedure depends on the value of $\widetilde{F}$. If $\widetilde{F} > 9.98$, we reject if $Wald(\beta_0) > \mathbb{C}_{0.02}$; otherwise if $\widetilde{F} \leq 9.98$, we reject if $AR(\beta_0) > z_{0.02}$.

The model for **1,530 instruments** can be written explicitly as

$$ln\ W_i = Constant + H_i^\top\zeta + \sum_{c=30}^{38} YOB_{i,c}\xi_c + \sum_{s\neq 56} POB_{i,s}\eta_s + \beta E_i + \gamma_i.$$

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

$$
\begin{aligned}
E_i = Constant + H_i^\top \lambda + \sum_{c=30}^{38} YOB_{i,c}\mu_c + \sum_{s\neq 56} POB_{i,s}\alpha_s \\
+ \sum_{j=1}^{3}\sum_{c=30}^{39}\sum_{s\in\{51\ states\}} QOB_{i,j}YOB_{i,c}POB_{i,s}\delta_{j,c,s}.
\end{aligned}
$$

The main difference between this 1,530-instrument specification and the 180-instrument one is that we now have QOB-YOB-POB interactions as our instruments, compared with QOB-YOB and QOB-POB interactions in the case of 180 instruments. Note that in both cases, only quarter-of-birth 1–3 are used; quarter 4 is omitted in order to avoid multicollinearity.

A.5 Proof of Lemma 1.2.1

Under strong identification, by (1.2.3) and Assumption 2, we have

$$
\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d_n \end{pmatrix} \begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \widetilde{\mathcal{C}} \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & 0 \\ \Phi_{12} & \Psi & 0 \\ 0 & 0 & 0 \end{pmatrix} \right),
$$

In addition, we note that $e_i(\beta_0) = e_i + X_i\Delta$ with $\Delta = d_n\widetilde{\Delta} \to 0$. Therefore, we have

$$
\begin{aligned}
Q_{e(\beta_0),e(\beta_0)} &= Q_{e,e} + 2\Delta Q_{X,e} + \Delta^2 Q_{X,X} = Q_{e,e} + o_p(1), \\
Q_{X,e(\beta_0)} &= Q_{X,e} + \Delta Q_{X,X} = Q_{X,e} + \widetilde{\mathcal{C}}\widetilde{\Delta} + o_p(1), \\
\widehat{\Phi}_1^{1/2}(\beta_0) &\xrightarrow{p} \Phi_1^{1/2}, \quad \text{and} \quad \widehat{\Psi}^{1/2}(\beta_0) \xrightarrow{p} \Psi^{1/2}.
\end{aligned}
$$

This implies

$$
\begin{pmatrix} AR(\beta_0) \\ LM(\beta_0) \end{pmatrix} = \begin{pmatrix} Q_{e(\beta_0),e(\beta_0)}/\widehat{\Phi}_1^{1/2}(\beta_0) \\ Q_{X,e(\beta_0)}/\widehat{\Psi}^{1/2}(\beta_0) \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\widetilde{\mathcal{C}}\widetilde{\Delta}}{\Psi^{1/2}} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).
$$

A.6 Proof of Lemma 1.2.2

Recall $\mathcal{N}_2^* = (1-\rho^2)^{-1/2}(\mathcal{N}_2 - \rho\mathcal{N}_1)$ and

$$\begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\theta}{(1-\rho^2)^{1/2}} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Because ρ is known, it suffices to construct the uniformly most powerful invariant test based on observations $(\mathcal{N}_1, \mathcal{N}_2^*)$. As the null and alternative are invariant to sign changes, the maximum invariant is $(\mathcal{N}_1, \mathcal{N}_2^{*2})$. Then, Lehmann and Romano (2006, Theorem 6.2.1) implies the invariant test should be based on the maximum invariant. Note $(\mathcal{N}_1, \mathcal{N}_2^{*2})$ are independent, $\mathcal{N}_1$ follows a standard normal distribution, and $\mathcal{N}_2^*$ follows a noncentral chi-square distribution with one degree of freedom and noncentrality parameter $\lambda = \frac{\theta^2}{1-\rho^2}$. Therefore, by the Neyman-Pearson's Lemma (Lehmann and Romano (2006, Theorem 3.2.1)), the most powerful test based on observations $(\mathcal{N}_1, \mathcal{N}_2^{*2})$ is the likelihood ratio test where the likelihood ratio function evaluated at $(\mathcal{N}_1 = \ell_1, \mathcal{N}_2^{*2} = \ell_2)$ depends on ℓ_2 only and can be written as

$$LR(\ell_2; \lambda) = -\frac{\lambda}{2} + \log\left(\frac{\exp(\sqrt{\lambda \ell_2}) + \exp(-\sqrt{\lambda \ell_2})}{2} \right)$$

In addition, we note that $LR(\ell_2; \lambda)$ is monotone increasing in ℓ_2 for any $\lambda \geq 0$ and $\ell_2 \geq 0$. Therefore, Lehmann and Romano (2006, Theorem 3.4.1) implies the likelihood ratio test is equivalent to $1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$, which is uniformly most powerful among tests for $\lambda = 0$ v.s. $\lambda > 0$ and based on observations $(\mathcal{N}_1, \mathcal{N}_2^{*2})$ only. This means it is also the uniformly most powerful test that is invariant to sign changes.

In addition, the joint density of $(\mathcal{N}_1, \mathcal{N}_2)$ is

$$(2\pi)^{-1}(1-\rho^2)^{-1/2} \exp\left(-\frac{1}{2} \left(\frac{\mathcal{N}_1^2}{1-\rho^2} - \frac{2\rho\mathcal{N}_1\mathcal{N}_2}{1-\rho^2} + \frac{\mathcal{N}_2^2}{1-\rho^2} \right) \right) \exp\left(\theta \frac{\rho\mathcal{N}_1 - \mathcal{N}_2}{1-\rho^2} \right) \exp\left(\frac{\theta^2}{1-\rho^2} \right)$$
$$\equiv C(\theta) \exp(\theta \mathcal{N}_2^*) h(\mathcal{N}_1, \mathcal{N}_2),$$

where $C(\theta) = (2\pi)^{-1}(1-\rho^2)^{-1/2} \exp\left(\frac{\theta^2}{1-\rho^2} \right)$ and $h(\mathcal{N}_1, \mathcal{N}_2) = \exp\left(-\frac{1}{2} \left(\frac{\mathcal{N}_1^2}{1-\rho^2} - \frac{2\rho\mathcal{N}_1\mathcal{N}_2}{1-\rho^2} + \frac{\mathcal{N}_2^2}{1-\rho^2} \right) \right)$. Note that $\mathcal{N}_2^*$ is symmetric around 0 under the null. By Lehmann and Romano (2006, Section 4.2), $1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}$ is the UMP unbiased level-α test.

A.7 Proof of Lemma 1.2.3

Under strong identification and fixed alternatives, because $(Q_{e(\beta_0),e(\beta_0)} - \Delta^2\mathcal{C}, Q_{X,e(\beta_0)} - \Delta\mathcal{C}, Q_{X,X} - \mathcal{C})^\top = O_p(1)$, we have

$$\begin{pmatrix} d_n AR(\beta_0) \\ d_n LM(\beta_0) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \frac{\Delta^2\widetilde{\mathcal{C}}}{\Phi_1^{1/2}(\beta_0)} \\ \frac{\Delta\widetilde{\mathcal{C}}}{\Psi^{1/2}(\beta_0)} \end{pmatrix}.$$

This implies

$$d_n LM^*(\beta_0) \xrightarrow{p} \frac{1}{(1-\rho^2(\beta_0))^{1/2}} \left(\frac{\Delta\widetilde{\mathcal{C}}}{\Psi^{1/2}(\beta_0)} - \frac{\rho(\beta_0)\Delta^2\widetilde{\mathcal{C}}}{\Phi_1^{1/2}(\beta_0)} \right),$$

which leads to the desired result.

A.8 Proof of Lemma 1.2.4

Under weak identification, (1.2.3) implies

$$\begin{pmatrix} Q_{e(\beta_0),e(\beta_0)} \\ Q_{X,e(\beta_0)} \end{pmatrix} = \begin{pmatrix} Q_{e,e} + 2\Delta Q_{X,e} + \Delta^2 Q_{X,X} \\ Q_{X,e} + \Delta Q_{X,X} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} \Delta^2\widetilde{\mathcal{C}} \\ \Delta\widetilde{\mathcal{C}} \end{pmatrix}, \begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix} \right),$$

which leads to the first result.

For the second result, it is obvious that $m_1(\Delta) \to \widetilde{\mathcal{C}}\Upsilon^{-1/2}$. In addition, we have

$$\begin{aligned} m_2(\Delta) &= \frac{\widetilde{\mathcal{C}}\left(\Delta\Phi_1(\beta_0) - \Delta^2\Phi_{12}(\beta_0)\right)}{(\Phi_1(\beta_0)(\Phi_1(\beta_0)\Psi(\beta_0) - \Phi_{12}^2(\beta_0)))^{1/2}} \\ &\to \frac{\tau\widetilde{\mathcal{C}}}{(\Upsilon(\Upsilon\Psi - \tau^2))^{1/2}} = \frac{\widetilde{\mathcal{C}}}{\Upsilon^{1/2}} \frac{\rho_{23}}{(1-\rho_{23}^2)^{1/2}}, \end{aligned}$$

where we use the fact that

$$\begin{aligned} &\Phi_1(\beta_0)/\Delta^4 \to \Upsilon, \\ &(\Phi_1(\beta_0)\Psi(\beta_0) - \Phi_{12}^2(\beta_0))/\Delta^4 \to \Upsilon\Psi - \tau^2, \end{aligned}$$

$$\frac{\Phi_1(\beta_0) - \Delta\Phi_{12}(\beta_0)}{\Delta^3} \to \tau.$$

A.9 Proof of Theorem 1.2.1

The first statement in Theorem 1.2.1(i) is a direct consequence of Marden (1982, Theorem 2.1) because the acceptance region $\mathcal{A} = \{(A, B) : \nu_1 A^2 + \nu_2 B^2 \leq \mathbb{C}_\alpha(a_1, a_2; \rho(\beta_0))\}$ is closed, convex, and monotone decreasing in the sense that if $(A, B) \in \mathcal{A}$ and $A' \leq A$, $B' \leq B$, then $(A', B') \in \mathcal{A}$. The second statement in Theorem 1.2.1(i) follows Andrews (2016, Theorem 2.1), which is a direct consequence of results in Monti and Sen (1976) and Koziol and Perlman (1978).

For Theorem 1.2.1(ii), we note that $\tilde{\rho} = \rho$ under local alternatives and

$$\phi_{a_1,a_2,\infty} = 1\left\{(a_1 + a_2\rho^2)\mathcal{N}_1^2 + 2a_2\rho(1-\rho^2)^{1/2}\mathcal{N}_1\mathcal{N}_2^* + (1 - a_1 - a_2\rho^2)\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha(a_1, a_2; \rho)\right\}.$$

The "if" part of Theorem 1.2.1(ii) is a direct consequence of Lemma 1.2.2. The "only if" part of Theorem 1.2.1(ii) is a direct consequence of the necessary part of Lehmann and Romano (2006, Theorem 3.2.1). Specifically, given $\mathcal{N}_1$ and $\mathcal{N}_2^*$ are independent, the "only if" part requires $a_1 + a_2\rho^2 = 0$, which implies $a_1 = 0$ and $a_2\rho = 0$.

For Theorem 1.2.1(iii), we consider two cases of fixed alternatives: (1) $\Delta \neq \Phi_1^{1/2}(\beta_0)\Psi^{-1/2}(\beta_0)\rho^{-1}(\beta_0)$ and (2) $\Delta = \Phi_1^{1/2}(\beta_0)\Psi^{-1/2}(\beta_0)\rho^{-1}(\beta_0)$. In Case (1), by Lemma 1.2.3, the limits of $d_n^2 AR^2(\beta_0)$, $d_n^2 LM^2(\beta_0)$, $d_n^2 LM^{*2}(\beta_0)$ are all positive, which implies that for all $(a_{1,n}, a_{2,n}) \in \mathbb{A}_0$,

$$1\{a_{1,n}AR^2(\beta_0) + a_{2,n}LM^2(\beta_0) + (1 - a_{1,n} - a_{2,n})LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha(a_{1,n}, a_{2,n}; \widehat{\rho}(\beta_0))\} \xrightarrow{p} 1.$$

In Case (2), we have

$$\begin{aligned}
&\mathbb{P}\left(a_{1,n}AR^2(\beta_0) + a_{2,n}LM^2(\beta_0) + (1 - a_{1,n} - a_{2,n})LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha(a_{1,n}, a_{2,n}; \widehat{\rho}(\beta_0))\right) \\
&\geq \mathbb{P}\left(\frac{\tilde{q}\Psi^2(\beta_0)\rho^4(\beta_0)}{\widetilde{\mathcal{C}}^2\Phi_1(\beta_0)} d_n^2 AR^2(\beta_0) \geq \mathbb{C}_\alpha(a_{1,n}, a_{2,n}; \widehat{\rho}(\beta_0)\right) \\
&\geq \mathbb{P}\left(\tilde{q} + o_p(1) \geq \mathbb{C}_{\alpha,\max}(\rho(\beta_0))\right) \to 1,
\end{aligned}$$

where the first inequality follows from the restriction on $a_{1.n}$ and the facts that $LM^2(\beta_0) \geq 0$

and $LM^{*2}(\beta_0) \geq 0$, the second inequality follows from $d_n^2 AR^2(\beta_0) \xrightarrow{p} \Phi_1^{-1}(\beta_0)\Delta_*^4(\beta_0)\tilde{\mathcal{C}}^2$ (by Lemma 1.2.3) and $\widehat{\rho}(\beta_0) \xrightarrow{p} \rho(\beta_0)$, and the last convergence follows from the fact that $\tilde{q} > \mathbb{C}_{\alpha,\max}(\rho(\beta_0))$. This concludes the proof.

A.10 Proof of Theorem 1.4.1

We are under weak identification. By Lemma 1.2.4 and Assumption 2, we have

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \\ \widehat{D} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \\ \mu_D \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma_D^2 \end{pmatrix}\right).$$

This implies $(AR(\beta_0), LM^*(\beta_0), \widehat{D})$ are asymptotically independent. By Assumption 3, we have

$$(AR^2(\beta_0), LM^{*2}(\beta_0), \mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))) \rightsquigarrow (\mathcal{Z}^2(m_1(\Delta)), \mathcal{Z}^2(m_2(\Delta)), \mathcal{A}_s(D, \gamma(\beta_0)))$$

where the two normal random variables are independent and independent of D, and by definition, $\mathcal{A}_s(D,\gamma(\beta_0))) = (a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)), a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)))$. In addition, we have $\widehat{\rho}(\beta_0) \xrightarrow{p} \rho(\beta_0)$. By the bounded convergence theorem, this further implies

$$\mathbb{E}\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} \to \mathbb{E}\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(\Delta, \mu_D, \gamma(\beta_0)). \tag{A.10.1}$$

In addition, suppose the null holds so that $\Delta = 0$. This implies $m_1(\Delta) = m_2(\Delta) = 0$. Then, we have

$$(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} - \alpha)f(\widehat{D}) \rightsquigarrow (\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(0, \mu_D, \gamma(\beta_0)) - \alpha)f(D),$$

where

$$\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(0, \mu_D, \gamma(\beta_0))$$

$$= 1\left\{\begin{array}{c} a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0))\mathcal{Z}_1^2 + a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0))(\rho(\beta_0)\mathcal{Z}_1 + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2) \\ (1 - a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)) - a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)))\mathcal{Z}_2^2 \\ \geq \mathbb{C}_\alpha(a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)), a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)); \rho(\beta_0)) \end{array}\right\},$$

$\mathcal{Z}_1$ and $\mathcal{Z}_2$ are independent standard normals, and they are independent of D. Then, by the definition of $\mathbb{C}_\alpha(\cdot)$, we have

$$\mathbb{E}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))} - \alpha)h(\widehat{D}) \to \mathbb{E}\left[\mathbb{E}\left(\phi_{a(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(0,\mu_D,\gamma(\beta_0)) - \alpha|D\right)h(D)\right] = 0.$$

A.11 Proof of Corollary 1.4.1

By the continuous mapping theorem, we have

$$\lim_{n\to\infty}\frac{\mathbb{E}\hat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}1\{|\widehat{D}-d|\leq\varepsilon\}}{\mathbb{E}1\{|\widehat{D}-d|\leq\varepsilon\}} = \frac{\mathbb{E}(\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}1\{|D-d|\leq\varepsilon\})}{\mathbb{E}1\{|D-d|\leq\varepsilon)\}},$$

and

$$\lim_{\varepsilon\to 0}\frac{\mathbb{E}(\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}1\{|D-d|\leq\varepsilon\})}{\mathbb{E}1\{|D-d|\leq\varepsilon)\}}$$
$$= \mathbb{E}(\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}|D=d),$$

where, by construction, we have

$$\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}$$
$$= 1\{\nu_{1,s}(D,\gamma(\beta_0))\widetilde{\mathcal{N}}_1^2 + \nu_{2,s}(D,\gamma(\beta_0))\widetilde{\mathcal{N}}_2^2 \geq \widetilde{\mathbb{C}}_\alpha(\nu_{1,s}(D,\gamma(\beta_0)),\nu_{2,s}(D,\gamma(\beta_0)))\}$$

and

$$(\widetilde{\mathcal{N}}_1,\widetilde{\mathcal{N}}_2) = (\mathcal{Z}_1(m_1(\Delta)),\mathcal{Z}_2(m_2(\Delta)))\mathcal{U}_s(D,\gamma(\beta_0)).$$

Similarly, we can show

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\mathbb{E}\tilde{\phi}(\widetilde{AR}_s^2(\beta_0),\widetilde{LM}_s^{*2}(\beta_0),\widehat{D},\widehat{\gamma}(\beta_0))1\{|\widehat{D}-d|\leq\varepsilon\}}{\mathbb{E}1\{|\widehat{D}-d|\leq\varepsilon\}} = \mathbb{E}(\tilde{\phi}(\widetilde{\mathcal{N}}_1^2,\widetilde{\mathcal{N}}_2^2,D,\gamma(\beta_0))|D=d).$$

Therefore, conditional on $D = d$, $\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}$ is a linear combination of $(\widetilde{\mathcal{N}}_1^2, \widetilde{\mathcal{N}}_2^2)$ with weights $(\nu_{1,s}(d,\gamma(\beta_0)), \nu_{2,s}(d,\gamma(\beta_0)))$, and $\widetilde{\mathcal{N}}_1$ and $\widetilde{\mathcal{N}}_2$ are two independent normal random variables with unit variance and expectations θ_1 and θ_2, respectively. Under the null, we have $(\theta_1,\theta_2) = (0,0)$, which, by definition of $\tilde{\phi}(\cdot)$, implies

$$\mathbb{E}(\tilde{\phi}(\widetilde{\mathcal{N}}_1^2, \widetilde{\mathcal{N}}_2^2, D, \gamma(\beta_0))|D = d) \leq \alpha.$$

Therefore, $\tilde{\phi}(\widetilde{\mathcal{N}}_1^2, \widetilde{\mathcal{N}}_2^2, D, \gamma(\beta_0))$ is a level-α test. Then, the two optimality results follow Theorem 1.2.1(i).

A.12 Proof of Theorem 1.4.2

Denote $c_{\mathcal{B}} = c_{\mathcal{B}}(\beta)$ and $\Delta_* = \Delta_*(\beta)$. By Assumption 2, $\Phi_1 > 0$, which implies $|\Delta_*| > 0$. Under strong identification and local alternatives, we have $\Delta \to 0$, $c_{\mathcal{B}}(\beta_0) \to c_{\mathcal{B}}$, $\Delta_*(\beta_0) \to \Delta_*$, $\mathbb{C}_{\alpha,\max}(\rho(\beta_0)) \to \mathbb{C}_{\alpha,\max}(\rho)$, and

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \\ d_n\widehat{D} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{((1-\rho^2)\Psi)^{1/2}} \\ \widetilde{\mathcal{C}} \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right).$$

This implies $d_n\widehat{\sigma}_D\sqrt{\widehat{r}} = d_n\widehat{D} \xrightarrow{p} \widetilde{\mathcal{C}}$, which further implies $d_n f_{pp}(\widehat{D}, \widehat{\gamma}(\beta_0)) \xrightarrow{p} \widetilde{\mathcal{C}}$. For $f_{krs}(\widehat{D}, \widehat{\gamma}(\beta_0))$, we note that

$$\max(\widehat{r} - 1, 0) \leq \widehat{r}_{krs} \leq \widehat{r}.$$

Therefore, we also have $f_{krs}(\widehat{D}, \widehat{\gamma}(\beta_0))d_n \xrightarrow{p} \widetilde{\mathcal{C}}$. Let $\mathcal{E}_n(\varepsilon) = \{||\widehat{\gamma}(\beta_0) - \gamma(\beta_0)|| + |\delta_n\widehat{D} - \widetilde{\mathcal{C}}| \leq \varepsilon\}$. Then, for an arbitrary $\varepsilon > 0$, we have $\mathbb{P}(\mathcal{E}_n(\varepsilon)) \geq 1 - \varepsilon$ when n is sufficiently large.

Denote $\delta = d_n\widetilde{\delta}$. We have

$$\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0)) \in \operatorname*{arg\,min}_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))} \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} \left(\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta}, \widehat{D}, \widehat{\gamma}(\beta_0))\right),$$

where $\widetilde{\mathcal{D}}_n = \{\widetilde{\delta} : d_n\widetilde{\delta} \in \mathcal{D}(\beta_0)\}$. Let

$$\begin{aligned}
&Q_n(a_1, a_2, \widetilde{\delta}) = \mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta}, \widehat{D}, \widehat{\gamma}(\beta_0)) \quad \text{and} \\
&Q(a_1, a_2, \widetilde{\delta}) = \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\} \\
&- \mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array}\right\},
\end{aligned}$$

where $\mathcal{Z}_1$ is standard normal, $\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})$ is normal with mean $(1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}$ and unit variance, and $\mathcal{Z}_1$ and $\mathcal{Z}_2(\cdot)$ are independent. Then, we aim to show that

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_n}\left|Q_n(a_1, a_2, \widetilde{\delta}) - Q(a_1, a_2, \widetilde{\delta})\right| \xrightarrow{p} 0. \tag{A.12.1}$$

We divide $\widetilde{\mathcal{D}}_n$ into three parts:

$$\begin{aligned}
&\widetilde{\mathcal{D}}_{n,1}(\varepsilon) = \{\widetilde{\delta} \in \widetilde{\mathcal{D}}_n, |\widetilde{\delta}| \leq M_1(\varepsilon)\}, \\
&\widetilde{\mathcal{D}}_{n,2}(\varepsilon) = \left\{\widetilde{\delta} \in \widetilde{\mathcal{D}}_n, \left|\frac{d_n\widetilde{\delta}}{\widehat{\Delta}_*(\beta_0)} - 1\right| \leq \varepsilon\right\}, \quad \text{and} \\
&\widetilde{\mathcal{D}}_{n,3}(\varepsilon) = \widetilde{\mathcal{D}}_n \cap \widetilde{\mathcal{D}}_{n,1}^c(\varepsilon) \cap \widetilde{\mathcal{D}}_{n,2}^c(\varepsilon),
\end{aligned}$$

where $M_1(\varepsilon)$ is a large constant so that

$$\mathbb{P}\left((1-\overline{a})\mathcal{Z}^2\left(\frac{M_1(\varepsilon)\varepsilon|\widetilde{\mathcal{C}}|}{(2(1-\rho^2)\Psi c_{\mathcal{B}})^{1/2}}\right) \geq \mathbb{C}_{\alpha,\max}(\rho) + 1\right) = 1 - \varepsilon. \tag{A.12.2}$$

When n is sufficiently large and ε is sufficiently small, on $\mathcal{E}_n(\varepsilon)$, there exists a constant c such that

$$\begin{aligned}
&|\widehat{\Delta}_*(\beta_0) - \Delta_*| \leq c\varepsilon, \quad \inf_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}|d_n\widetilde{\delta}| \geq (1-\varepsilon)(|\Delta_*| - c\varepsilon), \\
&|\widehat{\Phi}_1(\beta_0) - \Phi_1| \leq c\varepsilon, \quad |d_n^2 f_s^2(\widehat{D}, \widehat{\gamma}(\beta_0)) - \widetilde{\mathcal{C}}^2| \leq c\varepsilon, \\
&\sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[1 - (d_n^2\widetilde{\delta}^2, d_n\widetilde{\delta})\left(\begin{pmatrix}\widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0)\end{pmatrix}^{-1}\begin{pmatrix}\widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0)\end{pmatrix}\right)\right]^2
\end{aligned}$$

$$\leq \left[1 - (\Delta_*^2, \Delta_*) \left(\begin{pmatrix} \Phi_1 & \Phi_{12} \\ \Phi_{12} & \Psi \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13} \\ \tau \end{pmatrix} \right) \right]^2 + c\varepsilon \leq c_{\mathcal{B}} + c\varepsilon,$$

$$|\widehat{c}_{\mathcal{B}}(\beta_0) - c_{\mathcal{B}}| \leq c\varepsilon. \tag{A.12.3}$$

This further implies

$$\widetilde{\mathcal{D}}_{n,1}(\varepsilon) \cap \widetilde{\mathcal{D}}_{n,2}(\varepsilon) = \emptyset.$$

Recall $\phi_{a_1,a_2,s}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$ defined in (1.3.7). With δ replaced by $d_n\widetilde{\delta}$ and when $\widetilde{\delta} \in \widetilde{\mathcal{D}}_{n,1}(\varepsilon)$, we have

$$\begin{pmatrix} d_n^{-1}\widehat{C}_1(d_n\widetilde{\delta}) \\ d_n^{-1}\widehat{C}_2(d_n\widetilde{\delta}) \end{pmatrix} (d_n f_s(\widehat{D}, \widehat{\gamma}(\beta_0))) \xrightarrow{p} \begin{pmatrix} 0 \\ (1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}} \end{pmatrix},$$

Therefore, uniformly over $(a_1, a_2) \in \mathbb{A}_0$ and $\widetilde{\delta} \in \widetilde{\mathcal{D}}_{n,1}(\varepsilon)$ and conditional on data, we have

$$\phi_{a_1,a_2,s}(d_n\widetilde{\delta}, \widehat{D}, \widehat{\gamma}(\beta_0)) \rightsquigarrow 1 \left\{ \begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array} \right\}.$$

This implies

$$\sup_{(a_1,a_2)\in\mathbb{A}_0, \widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} \Bigg| \mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta}, \widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}1 \left\{ \begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array} \right\} \Bigg| \xrightarrow{p} 0.$$

In addition, by Lemma 1.2.2, for any $\widetilde{\delta}$, $\mathbb{E}1 \left\{ \begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array} \right\}$ is maximized at $a_1 = 0$ and $a_2\rho = 0$. This implies

$$\sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} |\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\}|$$

$$= \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} | \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))} \mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0)) - \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\}|$$

$$\leq \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} \left| \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))} \mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho) \end{array}\right\}\right.$$
$$\left. - \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\}\right| + o_p(1),$$
$$\leq \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} \left| \sup_{(a_1,a_2)\in\mathbb{A}_0} \mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho) \end{array}\right\}\right.$$
$$\left. - \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\}\right| + o_p(1) = o_p(1),$$

where the second inequality is due to the facts that $\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)) = o_p(1)$ under strong identification and $\mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho) \end{array}\right\}$ is continuous at $a_1 = 0$ uniformly over $|\widetilde{\delta}| \leq M_1(\varepsilon)$. Therefore, we have

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} \left|Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})\right| \xrightarrow{p} 0. \tag{A.12.4}$$

Next, we consider the case in which $\widetilde{\delta} \in \widetilde{\mathcal{D}}_{n,2}(\varepsilon)$. We have

$$\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))$$
$$= 1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2(\widehat{C}_1(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0))) \\ +a_2\left(\widehat{\rho}(\beta_0)\mathcal{Z}_1(\widehat{C}_1(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0))) + (1-\widehat{\rho}^2(\beta_0))^{1/2}\mathcal{Z}_2(\widehat{C}_2(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0)))\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(\widehat{C}_2(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0))) \geq \mathbb{C}_\alpha(a_1,a_2;\widehat{\rho}(\beta_0)) \end{array}\right\}$$
$$\geq 1\left\{\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))\mathcal{Z}_1^2(\widehat{C}_1(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0))) \geq \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\right\}.$$

By (A.12.3), on $\mathcal{E}_n(\varepsilon)$, there exists a constant $c > 0$ such that

$$\widehat{C}_1^2(d_n\widetilde{\delta})(d_nf_s(\widehat{D},\widehat{\gamma}(\beta_0)))^2$$

$$= \frac{\widehat{\Phi}_1^{-1}(\beta_0)(d_n\widetilde{\delta})^4(d_n f_s(\widehat{D},\widehat{\gamma}(\beta_0)))^2}{\left[1-(d_n^2\widetilde{\delta}^2, d_n\widetilde{\delta})\left(\begin{pmatrix}\widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0)\\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0)\end{pmatrix}^{-1}\begin{pmatrix}\widehat{\Phi}_{13}(\beta_0)\\ \widehat{\tau}(\beta_0)\end{pmatrix}\right)\right]^2}$$
$$\geq \frac{(\Phi_1(\beta_0)+c\varepsilon)^{-1}(1-\varepsilon)^4(|\Delta_*|-c\varepsilon)^4(\widetilde{\mathcal{C}}^2-c\varepsilon)}{c_{\mathcal{B}}+c\varepsilon} \geq c$$

and

$$\begin{aligned}
&\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))\widehat{C}_1^2(d_n\widetilde{\delta})f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))\\
&\geq \frac{p_2\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0)d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))}\widehat{C}_1^2(d_n\widetilde{\delta})(d_n f_s(\widehat{D},\widehat{\gamma}(\beta_0)))^2\\
&\geq \frac{p_2\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))(\Phi_1-c\varepsilon)(c_{\mathcal{B}}-c\varepsilon)}{(|\Delta_*|+c\varepsilon)^4(\widetilde{\mathcal{C}}^2+c\varepsilon)}\frac{(\Phi_1(\beta_0)+c\varepsilon)^{-1}(1-\varepsilon)^4(|\Delta_*|-c\varepsilon)^4(\widetilde{\mathcal{C}}^2-c\varepsilon)}{c_{\mathcal{B}}+c\varepsilon}\\
&\geq (p_2-c\varepsilon)\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)),
\end{aligned}$$

where the last inequality holds because ε can be arbitrarily small. This means, on $\mathcal{E}_n(\varepsilon)$ and when $\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)$,

$$\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0)) \geq \mathbb{P}^*(o_p(1)+(p_2-c\varepsilon)\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)) \geq \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))) \to 1.$$

As $\mathbb{P}(\mathcal{E}_n(\varepsilon))\to 1$, we have

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[1-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right]\xrightarrow{p}0,$$

and thus,

$$\begin{aligned}
&\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D},\widehat{\gamma}(\beta_0))-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right]\\
&\leq \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[1-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right]\xrightarrow{p}0. \qquad \text{(A.12.5)}
\end{aligned}$$

Furthermore, note that $a_1+a_2\leq\overline{a}<1$ and when $\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)$, on $\mathcal{E}_n(\varepsilon)$, (A.12.3) implies

$\widetilde{\delta}^2 \to \infty$. Therefore, we have

$$a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 + (1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})$$
$$\geq (1-\overline{a})\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) = \frac{(1-\overline{a})\widetilde{\delta}^2\widetilde{\mathcal{C}}^2}{(1-\rho^2)\Psi}(1+o_p(1)) \to \infty,$$

which further implies

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[1-\mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\right] -$$

and

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\Bigg[\mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\}$$
$$-\mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\Bigg] \xrightarrow{p} 0. \qquad (A.12.6)$$

Combining (A.12.5) and (A.12.6), we have

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left|Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})\right| \to 0. \qquad (A.12.7)$$

Last, we consider the case in which $\widetilde{\delta} \in \widetilde{\mathcal{D}}_{n,3}(\varepsilon)$. On $\mathcal{E}_n(\varepsilon)$, (A.12.3) implies

$$\widehat{C}_2^2(d_n\widetilde{\delta})f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))$$
$$= \frac{\widetilde{\delta}^2(1-\frac{d_n\widetilde{\delta}}{\widehat{\Delta}_*(\beta_0)})^2}{(1-\widehat{\rho}^2(\beta_0))\widehat{\Psi}(\beta_0)}\frac{d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))}{\left[1-(d_n^2\widetilde{\delta}^2, d_n\widetilde{\delta})\left(\begin{pmatrix}\widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0)\\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0)\end{pmatrix}^{-1}\begin{pmatrix}\widehat{\Phi}_{13}(\beta_0)\\ \widehat{\tau}(\beta_0)\end{pmatrix}\right)\right]^2}$$
$$\geq \frac{(1-c\varepsilon)M_1^2(\varepsilon)\varepsilon^2(\widetilde{\mathcal{C}}^2 - c\varepsilon)}{(1-\rho^2)\Psi c_{\mathcal{B}}}$$
$$\geq \frac{M_1^2(\varepsilon)\varepsilon^2\widetilde{\mathcal{C}}^2}{2(1-\rho^2)\Psi c_{\mathcal{B}}},$$

where the second inequality holds when ε is sufficiently small. In this case,

$$\begin{aligned}\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0)) &\geq \mathbb{P}^*((1-\overline{a})\mathcal{Z}_2^2(\widehat{C}_2(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0)))\geq \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)))\\ &\geq \mathbb{P}^*\left((1-\overline{a})\mathcal{Z}_2^2\left(\frac{M_1(\varepsilon)\varepsilon|\widetilde{\mathcal{C}}|}{(2(1-\rho^2)\Psi c_{\mathcal{B}})^{1/2}}\right)\geq \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\right)\\ &\geq \mathbb{P}^*\left((1-\overline{a})\mathcal{Z}_2^2\left(\frac{M_1(\varepsilon)\varepsilon|\widetilde{\mathcal{C}}|}{(2(1-\rho^2)\Psi c_{\mathcal{B}})^{1/2}}\right)\geq \mathbb{C}_{\alpha,\max}(\rho)+c\varepsilon\right)-\varepsilon\geq 1-2\varepsilon,\end{aligned}$$

where the second inequality is by the fact that the CDF (survival function) of $\mathcal{Z}^2(\lambda)$ is monotone decreasing (increasing) in $|\lambda|$ and the last equality is by the definition of $M_1(\varepsilon)$ in (A.12.2) and the fact that $\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\xrightarrow{p}\mathbb{C}_{\alpha,\max}(\rho)$. This implies, on $\mathcal{E}_n(\varepsilon)$,

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)}\left[\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D},\widehat{\gamma}(\beta_0))-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right]\leq 2\varepsilon. \tag{A.12.8}$$

In addition, we note that $(1-\rho^2)^{-1}\Psi^{-1}\widetilde{\delta}^2\widetilde{\mathcal{C}}^2$ satisfies

$$(1-\rho^2)^{-1}\Psi^{-1}\widetilde{\delta}^2\widetilde{\mathcal{C}}^2\geq\frac{M_1^2(\varepsilon)\varepsilon^2\widetilde{\mathcal{C}}^2}{2(1-\rho^2)\Psi c_{\mathcal{B}}},$$

where we use the facts that $\widetilde{\delta}^2 \geq M_1^2(\varepsilon)$, $c_{\mathcal{B}}\geq 1$, and $\varepsilon<1$. Therefore, by the same argument, we have

$$\mathbb{E}1\left\{\begin{array}{c}a_1\mathcal{Z}_1^2+a_2\left(\rho\mathcal{Z}_1+(1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2\\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\geq\mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\geq 1-\varepsilon$$

and

$$\begin{aligned}\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)}&\Bigg[\mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\geq\mathbb{C}_\alpha\}\\ -\mathbb{E}1&\left\{\begin{array}{c}a_1\mathcal{Z}_1^2+a_2\left(\rho\mathcal{Z}_1+(1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2\\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\geq\mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\Bigg]\leq\varepsilon.\end{aligned} \tag{A.12.9}$$

Combining (A.12.8) and (A.12.9), we have, on $\mathcal{E}_n(\varepsilon)$,

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)} \left| Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta}) \right| \leq 3\varepsilon. \tag{A.12.10}$$

Combining (A.12.4), (A.12.7), and (A.12.10), we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} |Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})| > 5\varepsilon \right) \\
&\leq \mathbb{P}\left(\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} |Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})| > \varepsilon, \mathcal{E}_n(\varepsilon) \right) \\
&+ \mathbb{P}\left(\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)} |Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})| > \varepsilon, \mathcal{E}_n(\varepsilon) \right) \\
&+ \mathbb{P}\left(\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)} |Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})| > 3\varepsilon, \mathcal{E}_n(\varepsilon) \right) + \mathbb{P}\left(\mathcal{E}_n^c(\varepsilon)\right) \\
&\leq o(1) + \varepsilon.
\end{aligned}$$

Since ε is arbitrary, we have

$$\omega_n \equiv \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} |Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})| \xrightarrow{p} 0.$$

Then we have

$$\begin{aligned}
0 &\leq \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q_n(\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),0,\widetilde{\delta}) - \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q_n(\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widetilde{\delta}) \\
&\leq \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q(\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),0,\widetilde{\delta}) - \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q(\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widetilde{\delta}) + 2\omega_n \\
&= o_p(1) - \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q(\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widetilde{\delta}) + 2\omega_n,
\end{aligned}$$

where the equality holds because (1) $\sup_{\widetilde{\delta}\in\Re} Q(a_1,0,\widetilde{\delta})$ is continuous at $a_1 = 0$ as shown in the proof of I.Andrews (2016, Theorem 5), (2) $\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)) = o_p(1)$ under strong identification, and (3) $\sup_{\widetilde{\delta}\in\Re} Q(0,0,\widetilde{\delta}) = 0$ by construction.

Furthermore, we have

$$
\begin{aligned}
Q(a_1, a_2, \widetilde{\delta}) &= \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\} \\
&- \mathbb{E}1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array}\right\} \\
&= \mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\} \\
&- \mathbb{E}1\left\{\begin{array}{c} (a_1 + a_2\rho^2)\mathcal{Z}_1^2 + a_2\rho(1-\rho^2)^{1/2}\mathcal{Z}_1\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \\ +(1-a_1-a_2\rho^2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1, a_2; \rho) \end{array}\right\}
\end{aligned}
$$

Note that $a_1 = 0$ and $a_2\rho = 0$ if and only if $a_1 + a_2\rho^2 = 0$, given that a_1 and a_2 are nonnegative. Therefore, Theorem 1.2.1(ii) implies, for any constant $C > 0$, there exists a constant $c > 0$ such that

$$
\inf_{(a_1,a_2)\in\mathbb{A}_0, a_1+a_2\rho^2 \geq C} \sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_n} Q(a_1, a_2, \widetilde{\delta}) \geq c > 0.
$$

Therefore,

$$
\mathbb{P}\left(\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) + \mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0))\rho^2 \geq C > 0\right) \leq \mathbb{P}\left(c \leq o_p(1) + 2\omega_n\right) \to 0.
$$

This implies $\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) \xrightarrow{p} 0$ and $\mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0))\rho \xrightarrow{p} 0$.

To see the optimality result, note that

$$
(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}, \phi(AR(\beta_0), LM(\beta_0))) \rightsquigarrow (1\{\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha\}, \phi(\mathcal{N}_1, \mathcal{N}_2)),
$$

where $(\mathcal{N}_1, \mathcal{N}_2)$ is defined above Theorem 1.4.2 and $\mathcal{N}_2^* = (1-\rho^2)^{-1/2}(\mathcal{N}_2 - \rho\mathcal{N}_1)$. Then, the result holds by Theorem 1.2.1(ii).

A.13 Proof of Theorem 1.4.3

We prove the result that $\limsup_{n\to\infty}\sup_{\lambda_n\in\Lambda_n}\mathbb{E}_\lambda(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}) = \alpha$. The other one can be proved in the same manner. Throughout the proof, we are under the null, i.e., $\beta_0 = \beta$. We start by proving the result for the full sequence $\{n\}$, rather than a subsequence $\{n_k\}$ of $\{n\}$.

Then, we note that the same proof goes through with n_k in place of n.

We consider two cases: sequences λ_n for which $\mathcal{C}_n$ converges to a constant and those for which it diverges to infinity. First, let us consider the case where $\mathcal{C}_n \to \widetilde{\mathcal{C}}$ for some fixed constant $\widetilde{\mathcal{C}} \in \Re$. For this case, it is established in Theorem 1.4.1 that under $\beta_0 = \beta$,

$$(AR^2(\beta_0), LM^{*2}(\beta_0), \mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))) \rightsquigarrow (\mathcal{Z}_1^2, \mathcal{Z}_2^2, \mathcal{A}_s(D, \gamma)),$$

where the two normal random variables are independent from each other and independent of D, and furthermore (by letting $h(\cdot)$ in Theorem 1.4.1 be an identity function),

$$\lim_{n\to\infty} \mathbb{E}_{\lambda_n}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}) = \alpha.$$

Second, let us consider the case where $\mathcal{C}_n$ diverges to infinity. Then, by Theorem 1.4.2, we have

$$\lim_{n\to\infty} \mathbb{E}_{\lambda_n}(\widehat{\phi}_{\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0))}) = \mathbb{P}(\mathcal{Z}_2^2 \geq \mathbb{C}_\alpha) = \alpha.$$

To complete the proof, we note that the above argument verifies Assumption B* in Andrews et al. (2020a) and then we can establish the result by using Corollary 2.1 in their paper.

A.14 Proof of Theorem 1.4.4

We consider strong identification with fixed alternatives. By construction, we have $\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) \geq \frac{p_2 \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))}$. By Theorem 1.2.1(iii), it suffices to show that, w.p.a.1,

$$\frac{p_2 \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))} \geq \frac{\tilde{q}\Psi^2(\beta_0)\rho^4(\beta_0)}{\mathcal{C}^2\Phi_1(\beta_0)},$$

or equivalently,

$$\frac{p_2 \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))} \geq \frac{\tilde{q}\Psi^2(\beta_0)\rho^4(\beta_0)}{\widetilde{\mathcal{C}}^2\Phi_1(\beta_0)} = \frac{\tilde{q}\Phi_1(\beta_0)}{\widetilde{\mathcal{C}}^2\Delta_*^4(\beta_0)}, \tag{A.14.1}$$

for some constant $\tilde{q} > \mathbb{C}_{\alpha,max}(\rho(\beta_0))$. Under strong identification and fixed alternatives, we have

$$d_n \widehat{D} = d_n \left(Q_{X,X} - (Q_{e(\beta_0),e(\beta_0)}, Q_{X,e(\beta_0)}) \begin{pmatrix} \widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0) \\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Phi}_{13}(\beta_0) \\ \widehat{\tau}(\beta_0) \end{pmatrix} \right)$$
$$\xrightarrow{p} \left[1 - (\Delta^2, \Delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right] \widetilde{\mathcal{C}}.$$

Therefore, we have

$$d_n f_s(\widehat{D}, \widehat{\gamma}(\beta_0)) = d_n \widehat{D} + o_p(1) \xrightarrow{p} \left[1 - (\Delta^2, \Delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right] \widetilde{\mathcal{C}}$$

for $s \in \{pp, krs\}$. This means for any $\varepsilon > 0$, w.p.a.1,

$$d_n^2 f_s^2(\widehat{D}, \widehat{\gamma}(\beta_0)) \leq (c_{\mathcal{B}}(\beta_0) + \varepsilon)\widetilde{\mathcal{C}}^2.$$

In addition, we have $\widehat{c}_{\mathcal{B}}(\beta_0) \xrightarrow{p} c_{\mathcal{B}}(\beta_0) \geq 1$, $\widehat{\Delta}_*(\beta_0) \xrightarrow{p} \Delta_*(\beta_0)$, $\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)) \xrightarrow{p} \mathbb{C}_{\alpha,\max}(\rho(\beta_0))$, and $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0) > 0$, which imply $\widehat{c}_{\mathcal{B}}(\beta_0) \geq c_{\mathcal{B}}(\beta_0) - c\varepsilon$, $\widehat{\Phi}_1(\beta_0) \geq \Phi_1(\beta_0) - c\varepsilon$, $\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)) \geq \mathbb{C}_{\alpha,\max}(\rho(\beta_0)) - c\varepsilon$, and $\widehat{\Delta}_*^4(\beta_0) \leq \Delta_*^4(\beta_0) + c\varepsilon$, w.p.a.1. Therefore, we have, w.p.a.1,

$$\frac{p_2 \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) d_n^2 f_s^2(\widehat{D}, \widehat{\gamma}(\beta_0))} \geq \frac{p_2(\mathbb{C}_{\alpha,\max}(\rho(\beta_0)) - c\varepsilon)(c_{\mathcal{B}}(\beta_0) - c\varepsilon)(\Phi_1(\beta_0) - c\varepsilon)}{(\Delta_*^4(\beta_0) + c\varepsilon)(c_{\mathcal{B}}(\beta_0) + \varepsilon)\widetilde{\mathcal{C}}^2}$$
$$\geq \frac{(p_2 - c\varepsilon)\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)}{\Delta_*^4(\beta_0)\widetilde{\mathcal{C}}^2},$$

where the second inequality holds because ε can be arbitrarily small. Then, we can let $\tilde{q}$ in (A.14.1) be $(p_2 - c\varepsilon)\mathbb{C}_{\alpha,\max}(\rho(\beta_0))$ which is greater than $\mathbb{C}_{\alpha,\max}(\rho(\beta_0))$. This concludes the proof.

A.15 Proof of Theorem A.1.1

We first extend our notation. For $a_i \in \Re^{d_1 \times 1}$ and $b_j \in \Re^{d_2 \times 1}$, we write $Q_{a,b}$ as $\sum_{i \in [n]} \sum_{j \neq i} a_i P_{ij} b_j^\top / \sqrt{K}$. Let $\hat{\gamma}_e = (W^\top W)^{-1}(W^\top \tilde{e})$ and $\hat{\gamma}_V = (W^\top W)^{-1}(W^\top \tilde{V})$. Then, we have $e_i = \tilde{e}_i - W_i^\top \hat{\gamma}_e$, $V_i = \tilde{V}_i - W_i^\top \hat{\gamma}_V$, and $X_i = \Pi_i + V_i = \Pi_i + \tilde{V}_i - W_i^\top \hat{\gamma}_V$. By Lemma A.19.1, we have

$$Q_{e,e} = Q_{\tilde{e} - W\hat{\gamma}_e, \tilde{e} - W\hat{\gamma}_e} = Q_{\tilde{e},\tilde{e}} - 2Q_{\tilde{e},W}\hat{\gamma}_e + \hat{\gamma}_e^\top Q_{W,W}\hat{\gamma}_e = Q_{\tilde{e},\tilde{e}} + o_P(1).$$

In addition, let $\overline{X} = \Pi + \tilde{V}$. Then, we have $X = \overline{X} - W\hat{\gamma}_V$ and

$$\begin{aligned}
Q_{X,e} &= Q_{\overline{X} - W\hat{\gamma}_V, \tilde{e} - W\hat{\gamma}_e} \\
&= Q_{\overline{X},\tilde{e}} - Q_{\tilde{e},W}\hat{\gamma}_V - Q_{\overline{X},W}\hat{\gamma}_e + \hat{\gamma}_V^\top Q_{W,W}\hat{\gamma}_e \\
&= Q_{\overline{X},\tilde{e}} - Q_{\overline{X},W}\hat{\gamma}_e + o_P(1) \\
&= Q_{\overline{X},\tilde{e}} - Q_{\Pi,W}\hat{\gamma}_e + o_P(1) \\
&= Q_{\overline{X},\tilde{e}} + \sum_{i \in [n]} \Pi_i P_{ii} W_i^\top \hat{\gamma}_e / \sqrt{K} + o_P(1),
\end{aligned}$$

where the last equality holds because

$$Q_{\Pi,W} = \sum_{i \in [n]} \Pi_i (\sum_{j \neq i} P_{ij} W_j^\top) / \sqrt{K} = -\sum_{i \in [n]} \Pi_i P_{ii} W_i^\top / \sqrt{K}.$$

Denote $G_i = (\sum_{i \in [n]} \Pi_i P_{ii} W_i^\top)(\sum_{i \in [n]} W_i W_i^\top)^{-1} W_i$. Then, we have

$$\begin{aligned}
Q_{X,e} &= Q_{\tilde{V},\tilde{e}} + Q_{\Pi,\tilde{e}} + \sum_{i \in [n]} G_i \tilde{e}_i / \sqrt{K} + o_P(1) \\
&= \frac{\sum_{i \in [n]} \sum_{j \neq i} \tilde{V}_i P_{ij} \tilde{e}_j}{\sqrt{K}} + \sum_{i \in [n]} \frac{(G_i + \omega_i)}{\sqrt{K}} \tilde{e}_i + o_P(1),
\end{aligned}$$

where $\omega_i = \sum_{j \neq i} P_{ij} \Pi_j$.

Similarly, we have

$$\begin{aligned}
Q_{X,X} &= Q_{\overline{X} - W\hat{\gamma}_V, \overline{X} - W\hat{\gamma}_V} \\
&= Q_{\overline{X},\overline{X}} - 2Q_{\overline{X},W}\hat{\gamma}_V + \hat{\gamma}_V^\top Q_{W,W}\hat{\gamma}_V
\end{aligned}$$

$$= Q_{\Pi,\Pi} + 2Q_{\Pi,\tilde{V}} + Q_{\tilde{V},\tilde{V}} - 2Q_{\Pi,W}\hat{\gamma}_V + o_P(1)$$
$$= Q_{\Pi,\Pi} + \frac{\sum_{i\in[n]}\sum_{j\neq i}\tilde{V}_i P_{ij}\tilde{V}_j}{\sqrt{K}} + 2\sum_{i\in[n]}\frac{\omega_i + G_i}{\sqrt{K}}\tilde{V}_i + o_P(1).$$

Given $\{\tilde{e}_i, \tilde{V}_i\}_{i\in[n]}$ are independent, we can follow the same argument in the proof of Chao et al. (2012, Lemma 2) and show the joint asymptotic normality of

$$\left(\frac{\sum_{i\in[n]}\sum_{j\neq i}\tilde{e}_i P_{ij}\tilde{e}_j}{\sqrt{K}}, \frac{\sum_{i\in[n]}\sum_{j\neq i}\tilde{V}_i P_{ij}\tilde{e}_j}{\sqrt{K}}, \frac{\sum_{i\in[n]}\sum_{j\neq i}\tilde{V}_i P_{ij}\tilde{V}_j}{\sqrt{K}}, \sum_{i\in[n]}\frac{(G_i+\omega_i)}{\sqrt{K}}\tilde{e}_i, \sum_{i\in[n]}\frac{(G_i+\omega_i)}{\sqrt{K}}\tilde{V}_i\right).$$

In particular, we see that

$$\begin{aligned}
Var\left(\sum_{i\in[n]}\frac{(G_i+\omega_i)\tilde{e}_i}{\sqrt{K}}\right) &= \sum_{i\in[n]}\frac{(G_i+\omega_i)^2\tilde{\sigma}_i^2}{K} \\
&\leq C\sum_{i\in[n]}\frac{(G_i+\omega_i)^2}{K} \\
&\leq C\left[\frac{(\sum_{i\in[n]}\Pi_i P_{ii} W_i^\top)(\sum_{i\in[n]} W_i W_i^\top)^{-1}(\sum_{i\in[n]}\Pi_i P_{ii} W_i)}{K} + \frac{\Pi^\top\Pi}{K}\right] \\
&\leq C\left[p_n^2\frac{\Pi^\top\Pi}{K} + \frac{\Pi^\top\Pi}{K}\right] = O(1)
\end{aligned}$$

and the same result for $Var(\sum_{i\in[n]}\frac{(G_i+\omega_i)\tilde{V}_i}{\sqrt{K}})$. This implies the joint asymptotic normality of

$$(Q_{e,e}, Q_{X,e}, Q_{X,X} - Q_{\Pi,\Pi}),$$

and thus, verifying Assumption 1.

To see the second result in Theorem A.1.1, we note that

$$\mathbb{E}\left(\sum_{i\in[n]} G_i\tilde{e}_i/\sqrt{K}\right)^2 \leq C\sum_{i\in[n]} G_i^2/K$$

$$
\begin{aligned}
&= C(\sum_{i\in[n]} \Pi_i P_{ii} W_i^\top)(\sum_{i\in[n]} W_i W_i^\top)^{-1}(\sum_{i\in[n]} \Pi_i P_{ii} W_i)/K \\
&\leq C \sum_{i\in[n]} \Pi_i^2 P_{ii}^2 / K \\
&\leq C \Pi^\top \Pi p_n^2 / K.
\end{aligned}
$$

If $\Pi^\top \Pi p_n^2/K = o(1)$, then we have $\sum_{i\in[n]} G_i \tilde{e}_i/\sqrt{K} = o_P(1)$. Similarly, we can show that, if $\Pi^\top \Pi p_n^2/K = o(1)$, $\sum_{i\in[n]} G_i \tilde{V}_i/\sqrt{K} = o_P(1)$. These imply $Q_{\overline{X},W}\hat{\gamma}_e = o_P(1)$ and $Q_{\overline{X},W}\hat{\gamma}_V = o_P(1)$, which further imply that

$$
Q_{X,e} = Q_{\overline{X},\tilde{e}} + o_P(1) \quad \text{and} \quad Q_{X,X} = Q_{\overline{X},\overline{X}} + o_P(1).
$$

A.16 Proof of Theorem A.2.1

We focus on the consistency of $\widehat{\Phi}_1(\beta_0)$ and $\widehat{\Psi}(\beta_0)$. The consistency of the rest four estimators can be established in the same manner. We have $e_i(\beta_0) = e_i + \Delta X_i = V_i(\Delta) + \Delta \Pi_i$, where $V_i(\Delta) = e_i + \Delta V_i$. Therefore,

$$
\begin{aligned}
\widehat{\Phi}_1(\beta_0) &= \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 e_i^2(\beta_0) e_j^2(\beta_0) \\
&= \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\Delta^2 \Pi_i^2 + 2\Delta \Pi_i V_i(\Delta) + V_i^2(\Delta))(\Delta^2 \Pi_j^2 + 2\Delta \Pi_j U_j(\Delta) + U_j^2(\Delta)) \\
&= \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 V_i^2(\Delta) U_j^2(\Delta) + \Delta \frac{4}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\Pi_i V_i(\Delta) U_j^2(\Delta) + \Pi_j U_j(\Delta) V_i^2(\Delta)) \\
&+ \Delta^2 \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\Pi_i^2 U_j^2(\Delta) + \Pi_j^2 V_i^2(\Delta) + 4\Pi_i \Pi_j V_i(\Delta) U_j(\Delta)) \\
&+ \Delta^3 \frac{4}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\Pi_i^2 \Pi_j U_j(\Delta) + \Pi_j^2 \Pi_i V_i(\Delta)) + \Delta^4 \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \Pi_i^2 \Pi_j^2 \\
&\equiv \sum_{l=0}^{4} \Delta^l T_l.
\end{aligned}
$$

We first note that $\frac{1}{K}\sum_{i\in[n]} \omega_i^2 \sigma_i^2 = o(1)$, $\frac{1}{K}\sum_{i\in[n]} \omega_i^2 \gamma_i = o(1)$, and $\frac{1}{K}\sum_{i\in[n]} \omega_i^2 \eta_i^2 = o(1)$.

To see this, note that

$$\frac{1}{K}\sum_{i\in[n]}\omega_i^2\sigma_i^2 \le \frac{C}{K}\sum_{i\in[n]}\omega_i^2 = \frac{C}{K}\sum_{i\in[n]}(P_i\Pi - P_{ii}\Pi_i)^2$$
$$\le \frac{C}{K}(2\Pi^\top P^2\Pi + 2\sum_{i\in[n]}P_{ii}^2\Pi_i^2) \le C\frac{\Pi^\top\Pi}{K} = o(1),$$

where the second and third inequalities are shown in the Proof of Mikusheva and Sun (2022, Lemma S1.4). The results for $\frac{1}{K}\sum_{i\in[n]}\omega_i^2\gamma_i = o(1)$ and $\frac{1}{K}\sum_{i\in[n]}\omega_i^2\eta_i^2 = o(1)$ can be established in the same manner.

We first consider T_0. Denote $\xi_{ij} = V_i^2(\Delta)U_j^2(\Delta) - \mathbb{E}V_i^2(\Delta)U_j^2(\Delta)$. We want to show that

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

Note that

$$\mathbb{E}\left[\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\xi_{ij}\right]^2 = \frac{1}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^4\mathbb{E}\xi_{ij}^2 + \frac{4}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{i'\neq i,j}P_{ij}^2P_{ii'}^2\mathbb{E}\xi_{ij}\xi_{ii'}.$$

As both $\mathbb{E}\xi_{ij}^2$ and $|\mathbb{E}\xi_{ij}\xi_{ii'}|$ are bounded, we have

$$\frac{1}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^4\mathbb{E}\xi_{ij}^2 \le \frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \le \frac{C}{K} = o(1)$$

and

$$\left|\frac{1}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{i'\neq i,j}P_{ij}^2P_{ii'}^2\mathbb{E}\xi_{ij}\xi_{ii'}\right| \le \frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}\sum_{i'\neq i,j}P_{ij}^2P_{ii'}^2 \le \frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2P_{ii} = o(1).$$

Therefore, we have

$$T_0 = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(V_i^2(\Delta)U_j^2(\Delta)) + o_p(1)$$

$$= \Delta^4 \frac{2}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \eta_i^2 \eta_j^2 + \Delta^3 \frac{4}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 (\eta_i^2 \gamma_j + \eta_j^2 \gamma_i) + \Delta^2 \frac{2}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 (\eta_i^2 \sigma_j^2 + \eta_j^2 \sigma_i^2 + 4\gamma_i \gamma_j)$$
$$+ \Delta \frac{4}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 (\gamma_i \sigma_j^2 + \gamma_j \sigma_i^2) + \frac{2}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \sigma_i^2 \sigma_j^2 + o_p(1)$$
$$= \Phi_1(\beta_0) + o_p(1).$$

By the same argument above, we have

$$T_1 = \mathbb{E}T_1 + o_p(1) = o_p(1)$$

because $\mathbb{E}T_1 = 0$. Similarly, we have $\mathbb{E}T_3 = 0$ and $T_3 = o_p(1)$. Next, we have

$$T_2 = \mathbb{E}T_2 + o_P(1) \leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \Pi_i^2 + o_p(1) \leq \frac{C p_n \Pi^\top \Pi}{K} + o_p(1) = o_p(1).$$

Last, we have

$$T_4 \leq \frac{C}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \Pi_i^2 = o(1),$$

where the first inequality is by $\max_{i \in [n]} |\Pi_i| < C$. This implies

$$\widehat{\Phi}_1(\beta_0) - \Phi_1(\beta_0) = o_p(1).$$

Next, we consider the consistency of $\widehat{\Psi}(\beta_0)$. By the similar argument above, we have

$$\frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 X_i e_i(\beta_0) X_j e_j(\beta_0))$$
$$= \frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \Pi_i e_i(\beta_0) \Pi_j e_j(\beta_0)) + \frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 \Pi_i e_i(\beta_0) V_j e_j(\beta_0))$$
$$+ \frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 V_i e_i(\beta_0) \Pi_j e_j(\beta_0)) + \frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 V_i e_i(\beta_0) V_j e_j(\beta_0))$$
$$= \frac{1}{K} \sum_{i \in [n]} \sum_{j \neq i} P_{ij}^2 (\gamma_i + \Delta \eta_i^2)(\gamma_j + \Delta \eta_j^2) + o_p(1). \tag{A.16.1}$$

In addition, we have

$$
\begin{aligned}
&\frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i}P_{ij}X_j)^2e_i^2(\beta_0)\\
&=\frac{1}{K}\sum_{i\in[n]}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2e_i^2(\beta_0)\\
&=\frac{1}{K}\sum_{i\in[n]}\omega_i^2\mathbb{E}e_i^2(\beta_0)+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2\mathbb{E}e_i^2(\beta_0)+o_p(1)\\
&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2(\sigma_i^2+2\gamma_i\Delta+\Delta^2\eta_i^2)+o_p(1),
\end{aligned}
\tag{A.16.2}
$$

where the second equality is due to Mikusheva and Sun (2022, proof of statement (a) in Lemma S3.2), and the third equality is due to $\frac{1}{K}\sum_{i\in[n]}\omega_i^2\sigma_i^2=o(1)$. In the next section, we show the same results hold under Assumption 11. Combining (A.16.1) and (A.16.2), we have

$$
\begin{aligned}
\widehat{\Psi}(\beta_0)&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\gamma_i+\Delta\eta_i^2)(\gamma_j+\Delta\eta_j^2)+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2(\sigma_i^2+2\gamma_i\Delta+\Delta^2\eta_i^2)+o_p(1)\\
&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\gamma_i\gamma_j+\sigma_i^2\eta_j^2)+\frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\gamma_j+\frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\eta_j^2+o_p(1)\\
&=\Psi(\beta_0)+o_p(1).
\end{aligned}
$$

A.17 Proof of Theorem A.2.2

Given Lemma A.2.1, Lemmas 2 and 3 in Mikusheva and Sun (2022) hold under Assumptions 10 and 12. Therefore, Mikusheva and Sun (2022, Theorem 3) shows that

$$
\widehat{\Phi}_1(\beta_0)-\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}V_i^2(\Delta)\mathbb{E}U_j^2(\Delta)=o_p(1).
$$

In addition, the proof of Theorem A.2.1 shows that

$$
\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}V_i^2(\Delta)\mathbb{E}U_j^2(\Delta)=\Phi_1(\beta_0)+o(1),
$$

which implies the consistency of $\widehat{\Phi}_1(\beta_0)$.

Similarly, given Lemma A.2.1, Lemma S3.1 in Mikusheva and Sun (2022) holds under Assumptions 10 and 12, so that the consistency of $\widehat{\Upsilon}$ to Υ is also shown by using their argument. In addition, we use the same argument in the proof of Mikusheva and Sun (2022, Theorem 5) to show that

$$\begin{aligned}
\widehat{\Psi}(\beta_0) = &\left\{ \frac{1}{K} \sum_{i\in[n]} (\sum_{j\neq i} P_{ij} X_j)^2 \frac{e_i M_i e}{M_{ii}} + \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_i X e_i M_j X e_j \right\} \\
&+ \Delta \left\{ \frac{1}{K} \sum_{i\in[n]} (\sum_{j\neq i} P_{ij} X_j)^2 \left(\frac{e_i M_i X}{M_{ii}} + \frac{X_i M_i e}{M_{ii}} \right) + \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_i X e_i M_j X X_j \right\} \\
&+ \Delta^2 \left\{ \frac{1}{K} \sum_{i\in[n]} (\sum_{j\neq i} P_{ij} X_j)^2 \frac{X_i M_i X}{M_{ii}} + \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_i X X_i M_j X X_j \right\} \\
&= \Psi + 2\Delta\tau + \Delta^2 \Upsilon + o_p(1) = \Psi(\beta_0) + o_p(1),
\end{aligned}$$

where the second equality also follows from Lemma S3.1 in Mikusheva and Sun (2022).

Next for $\widehat{\Phi}_{12}(\beta_0)$, we have

$$\begin{aligned}
&\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_j X e_j(\beta_0) e_i(\beta_0) M_i e(\beta_0) \\
&= \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_j X e_j e_i M_i e \\
&+ \Delta \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 \left(M_j X X_j e_i M_i e + M_j X e_j X_i M_i e + M_j X e_j e_i M_i X \right) \\
&+ \Delta^2 \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 \left(M_j X X_j X_i M_i e + M_j X X_j e_i M_i X + M_j X e_j X_i M_i X \right) \\
&+ \Delta^3 \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_j X X_j X_i M_i X.
\end{aligned}$$

Note that $\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 M_j X e_j e_i M_i e = \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 (M_j V + \lambda_i) e_j e_i M_i e$, where

$\lambda_i = M_i\Pi$. Then, by Lemma A.2.1 and Lemma 3 of Mikusheva and Sun (2022),

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X e_j e_i M_i e - \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j V e_j e_i M_i e = o_p(1).$$

Furthermore, by Lemma A.2.1 and Lemma 2 of Mikusheva and Sun (2022),

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j V e_j e_i M_i e - \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \gamma_j \sigma_i^2 = o_p(1).$$

By using similar arguments, we find that

$$\begin{aligned}
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X X_j e_i M_i e &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \eta_j^2 \sigma_i^2 + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X e_j X_i M_i e &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \gamma_j \gamma_i + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X e_j e_i M_i X &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \gamma_j \gamma_i + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X X_j X_i M_i e &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \eta_j^2 \gamma_i + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X X_j e_i M_i X &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \eta_j^2 \gamma_i + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X e_j X_i M_i X &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \gamma_j \eta_i^2 + o_p(1),\\
\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_j X X_j X_i M_i X &= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 \eta_j^2 \eta_i^2 + o_p(1).
\end{aligned}$$

Putting these results together, we obtain

$$\widehat{\Phi}_{12}(\beta_0) = \Phi_{12} + \Delta(2\Psi + \Phi_{13}) + 3\Delta^2\tau + \Delta^3\Upsilon + o_p(1) = \Phi_{12}(\beta_0) + o_p(1).$$

We use similar arguments to prove the results for $\widehat{\Psi}_{13}(\beta_0)$ and $\widehat{\tau}(\beta_0)$. For $\widehat{\Phi}_{13}(\beta_0)$, notice that

$$
\begin{aligned}
&\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_iXe_i(\beta_0)M_jXe_j(\beta_0)\\
&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_iXe_iM_jXe_j\\
&+\Delta\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(M_iXe_iM_jXX_j+M_iXX_iM_jXe_j)\\
&+\Delta^2\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_iXX_iM_jXX_j\\
&=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\gamma_i\gamma_j+\Delta\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\gamma_i\eta_j^2+\eta_i^2\gamma_j)+\Delta^2\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\eta_j^2+o_p(1),
\end{aligned}
$$

which implies that

$$\widehat{\Phi}_{13}(\beta_0)=\Phi_{13}+2\Delta\tau+\Delta^2\Upsilon+o_p(1)=\Phi_{13}(\beta_0)+o_p(1).$$

Finally, for $\widehat{\tau}(\beta_0)$, notice that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 X_iM_iXM_jXe_j(\beta_0)=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\gamma_j+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\eta_j^2\Delta+o_p(1),$$

$$\frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i}P_{ij}X_j)^2\left(\frac{e_i(\beta_0)M_iX}{2M_{ii}}+\frac{X_iM_ie(\beta_0)}{2M_{ii}}\right)=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\gamma_j+\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_i^2\eta_j^2\Delta+o_p(1),$$

which implies that

$$\widehat{\tau}(\beta_0)=\tau+\Delta\Upsilon+o_p(1)=\tau(\beta_0)+o_p(1).$$

This completes the proof of the theorem.

A.18 Proof of Lemma A.2.1

Let $p_n = \max_i P_{ii}$. We first give some useful bounds, which is similar to Lemma S1.4 in Mikusheva and Sun (2022):

$$\sum_{i\in[n]} \omega_i^2 = \sum_{i\in[n]} (P_i\Pi - P_{ii}\Pi_i)^2 \leq 2\Pi' P^2 \Pi + 2\sum_{i\in[n]} P_{ii}^2 \Pi^2 \leq C\Pi^\top \Pi,$$
$$\max_{i\in[n]} \omega_i^2 = \max_{i\in[n]} (\sum_{j\neq i} P_{ij}\Pi_j)^2 \leq \max_{i\in[n]} (\sum_{j\neq i} P_{ij}^2)\Pi^\top \Pi \leq p_n \Pi^\top \Pi,$$

which imply

$$\sum_{i\in[n]} \omega_i^4 \leq \max_{i\in[n]} \omega_i^2 (\sum_{i\in[n]} \omega_i^2) \leq C p_n (\Pi^\top \Pi)^2.$$

First, we show that Mikusheva and Sun (2022, Lemma S2.1) hold under our conditions following the lines of argument in their proof. More specifically, we notice that to show $\Delta^2 |\mathbb{E} A_2| = o(1)$, where A_2 is defined in the proof of Mikusheva and Sun (2022, Lemma S2.1), it suffices to show the following terms are o(1):

$$\frac{C\Delta^2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 |\lambda_i||\Pi_j| \leq \frac{C\Delta^2}{K} \left(\sum_{i\in[n]} P_{ii}\lambda_i^2\right)^{1/2} \left(\sum_{j\in[n]} P_{jj}\Pi_j^2\right)^{1/2} \leq \frac{C\Delta^2}{K} p_n \left(\lambda^\top \lambda\right)^{1/2} \left(\Pi^\top \Pi\right)^{1/2}$$
$$\leq \frac{C\Delta^2}{K^{3/2}} p_n \left(\Pi^\top \Pi\right) = o(1) \text{ by } \lambda^\top \lambda \leq C\frac{\Pi^\top \Pi}{K},$$
$$\frac{C\Delta^2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 |\Pi_i||\Pi_j| \leq \frac{C\Delta^2}{K} \left(\sum_{i\in[n]} P_{ii}\Pi_i^2\right)^{1/2} \left(\sum_{j\in[n]} P_{jj}\Pi_j^2\right)^{1/2} \leq \frac{C\Delta^2}{K} p_n \left(\Pi^\top \Pi\right) = o(1).$$

Then, we prove the variance of $\Delta^2 A_2 = o(1)$ by showing that

$$\frac{C\Delta^4}{K^2} \sum_{i\in[n]} \sum_{j\in[n]} P_{ij}^4 \lambda_i^2 \lambda_j^2 \leq \frac{C\Delta^4}{K^2} p_n^2 \left(\lambda^\top \lambda\right)^2 \leq \frac{C\Delta^4}{K^2} p_n^2 \left(\frac{\Pi^\top \Pi}{K}\right)^2 = o(1) \text{ by } P_{ij}^2 \leq P_{ii},$$
$$\frac{C\Delta^4}{K^2} \left(\sum_{i\in[n]} \lambda_i^2 \left(\sum_{j\in[n]} P_{ij}^2\right) \Pi^\top \Pi + \lambda^\top \lambda \left(\sum_{j\in[n]} P_{jj}|\Pi_j|\right)^2\right) \leq \frac{C\Delta^4}{K^2} \left(p_n(\lambda^\top \lambda)(\Pi^\top \Pi) + (\lambda^\top \lambda)(p_n K)(\Pi^\top \Pi)\right)$$

$$\le \frac{C\Delta^4}{K^3}\left(p_n(\Pi^\top\Pi)^2 + p_nK(\Pi^\top\Pi)^2\right) = o(1) \text{ by } \sum_{j\in[n]} P_{jj}^2 \le p_n K,$$

$$\frac{C\Delta^4}{K^2}\left(\sum_{i\in[n]}\sum_{j\in[n]} P_{ij}^2|\Pi_i\Pi_j|\right)^2 \le \frac{C\Delta^4}{K^2}\left(\sum_{i\in[n]} P_{ii}\Pi_i^2\right)\left(\sum_{j\in[n]}P_{jj}\Pi_j^2\right) \le \frac{C\Delta^4}{K^2}p_n^2(\Pi^\top\Pi)^2 = o(1),$$

and

$$\frac{C\Delta^4}{K^2}\sum_{j\in[n]}\sum_{k\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\lambda_i\Pi_i M_{jk}|\right)^2 = \frac{C\Delta^4}{K^2}\sum_{j\in[n]}\sum_{k\neq j}\left(\sum_{i\in[n]}P_{ij}^2|\lambda_i\Pi_i M_{jk}|\right)^2 + \frac{C\Delta^4}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\lambda_i\Pi_i M\right.$$

$$\le \frac{C\Delta^4}{K^2}\sum_{j\in[n]}\sum_{k\neq j}M_{jk}^2\left(\sum_{i\in[n]}P_{ii}\lambda_i^2\right)\left(\sum_{i\in[n]}P_{ii}\Pi_i^2\right) + \frac{C\Delta^4}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\lambda_i|\right)^2$$

$$\le \frac{C\Delta^4}{K^2}Kp_n^2(\lambda^\top\lambda)(\Pi^\top\Pi) + \frac{C\Delta^4}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^4\right)\lambda^\top\lambda$$

$$\le \frac{C\Delta^4Kp_n^2(\Pi^\top\Pi)^2}{K^2} + \frac{C\Delta^4p_nK\Pi^\top\Pi}{K^2} = o(1) \text{ by } \sum_{j\in[n]}\sum_{k\neq j}M_{jk}^2 = \sum_{j\in[n]}\sum_{k\neq j}P_{jk}^2 \le K \text{ and } P_{ij}^2 \le P_{ii} \le p_n.$$

Second, we show that Mikusheva and Sun (2022, Lemma S2.2) holds under our conditions. Notice that $|\Delta\mathbb{E}A_1| = o(1)$ by

$$\frac{C|\Delta|}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2|\Pi_i| \le \frac{C|\Delta|}{K}\left(\sum_{i\in[n]}P_{ii}^2\right)^{1/2}(\Pi^\top\Pi)^{1/2} \le \frac{C|\Delta|}{K}(p_nK)^{1/2}(\Pi^\top\Pi)^{1/2} = o(1),$$

Then, we show that the variance of ΔA_1 is $o(1)$ by showing the following terms are $o(1)$:

$$\frac{C\Delta^2}{K^2}\left(\sum_{i\in[n]}\left(\sum_{j\in[n]}P_{ij}^2\right)\lambda_i^2 + \sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2|\lambda_i||\lambda_j|\right) \le \frac{C\Delta^2}{K^2}\left(p_n(\lambda^\top\lambda) + p_n(\lambda^\top\lambda)\right) = o(1),$$

$$\frac{C\Delta^2}{K^2}\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^4(\lambda_i^2 + |\lambda_i||\lambda_j|) + \sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\lambda_i^2\right) \le \frac{C\Delta^2}{K^2}\left(p_n^2(\lambda^\top\lambda) + p_n^2(\lambda^\top\lambda) + p_n(\lambda^\top\lambda)\right) = o(1),$$

$$\frac{C\Delta^2}{K^2}\left(\sum_{i\in[n]}\sum_{k\in[n]}P_{ik}^2|\lambda_i||\lambda_k|+\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2|\lambda_i||\lambda_j|\right)\leq\frac{C\Delta^2}{K^2}\left(p_n(\lambda^\top\lambda)+p_n(\lambda^\top\lambda)\right)=o(1),$$

$$\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\lambda_i|\right)^2\leq\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^4\right)(\lambda^\top\lambda)\leq\frac{C\Delta^2}{K^2}(p_nK)(\lambda^\top\lambda)=o(1),$$

$$\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_i|\right)^2\leq\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^4\right)(\Pi^\top\Pi)\leq\frac{C\Delta^2}{K^2}(p_nK)(\Pi^\top\Pi)=o(1),$$

$$\begin{aligned}
&\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\sum_{k\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_iM_{ik}M_{jk}|\right)^2\\
&=\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\sum_{k\neq j}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_iM_{ik}M_{jk}|\right)^2+\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_iM_{ij}M_{jj}|\right)^2\\
&\leq\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\sum_{k\neq j}M_{jk}^2\left(\sum_{i\in[n]}P_{ij}^4\right)\Pi^\top\Pi+\frac{C\Delta^2}{K^2}\left(\sum_{j\in[n]}\sum_{i\in[n]}P_{ij}^4\right)\Pi^\top\Pi\\
&\leq\frac{C\Delta^2}{K^2}Kp_n^2(\Pi^\top\Pi)+\frac{C\Delta^2}{K^2}Kp_n(\Pi^\top\Pi)=o(1),
\end{aligned}$$

$$\frac{C\Delta^2}{K^2}\sum_{j\in[n]}\sum_{k\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_iM_{ik}M_{jk}|\right)\left(\sum_{i\in[n]}P_{ik}^2|\Pi_iM_{ij}M_{jk}|\right)\leq\frac{C\Delta^2}{K^2}Kp_n(\Pi^\top\Pi)=o(1),$$

$$\frac{C\Delta^2}{K^2}\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2|\Pi_i|\right)^2\leq\frac{C\Delta^2}{K^2}(p_nK)(\Pi^\top\Pi)=o(1).$$

Then, to show that Mikusheva and Sun (2022, Lemma 3) holds under our conditions, we show the following terms are o(1):

$$\begin{aligned}
&\frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2|\Pi_i\lambda_i\Pi_j\lambda_j|\leq\frac{C}{K}\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\Pi_i^2\Pi_j^2\right)^{1/2}\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\lambda_i^2\lambda_j^2\right)^{1/2}\\
&\leq\frac{C}{K}p_n\left(\Pi^\top\Pi\right)\left(\lambda^\top\lambda\right)\leq\frac{C}{K^2}p_n\left(\Pi^\top\Pi\right)^2=o(1),
\end{aligned}$$

$$\frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\in[n]}P_{ij}^2|\Pi_i||\lambda_i|\right)^2\lambda_j^2 \le \frac{C}{K^2}\sum_{j\in[n]}\left(p_n\sum_{i\in[n]}|\Pi_i||\lambda_j|\right)^2\lambda_j^2 \le \frac{C}{K^2}p_n^2\left(\Pi^\top\Pi\right)\left(\frac{\Pi^\top\Pi}{K}\right)^2 = o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\in[n]}\sum_{j\in[n]}\sum_{j'\in[n]}P_{ij}^2|\Pi_i\lambda_i\Pi_j|P_{i'j'}^2|\Pi_{i'}\lambda_{i'}\Pi_{j'}|\sum_{k\in[n]}|M_{jk}M_{j'k}|$$

$$\le \frac{C}{K^2}\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\Pi_i^2\lambda_i^2\right)\left(\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\Pi_j^2\right) \le \frac{C}{K^2}p_n^2(\Pi^\top\Pi)(\lambda^\top\lambda) \le \frac{C}{K^3}p_n^2\left(\Pi^\top\Pi\right)^2 = o(1),$$

where $\sum_{k\in[n]}|M_{jk}M_{j'k}| \le 1$ by Mikusheva and Sun (2022, Lemma S1.1(ii)).

Now we show that Mikusheva and Sun (2022, Lemma S3.2) holds under our conditions, i.e.,

$$(a)\quad \frac{1}{K}\sum_{i=1}^{n}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2\mathcal{V}_i - \left(\frac{1}{K}\sum_{i=1}^{n}\omega_i^2\mathbb{E}[\mathcal{V}_i] + \frac{1}{K}\sum_{i,j\neq i}P_{ij}^2\mathbb{E}[\mathcal{V}_i]\eta_j^2\right) \xrightarrow{p} 0,$$

$$(b)\quad \frac{1}{K}\sum_{i=1}^{n}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2\frac{\xi_{1,i}}{M_{ii}}\sum_{k\neq j}P_{ik}\xi_{2,k} \xrightarrow{p} 0,$$

$$(c)\quad \frac{1}{K}\sum_{i=1}^{n}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2a_i\xi_{1,i} \xrightarrow{p} 0,$$

$$(d)\quad \frac{1}{K}\sum_{i=1}^{n}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2\frac{a_i}{M_{ii}}\sum_{k\neq i}P_{ik}\xi_{1,k} - \frac{2}{K}\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^2\omega_i\frac{a_i}{M_{ii}}\mathbb{E}[V_j\xi_{1,j}] \xrightarrow{p} 0,$$

$$(e)\quad \frac{1}{K}\sum_{i=1}^{n}(\omega_i+\sum_{j\neq i}P_{ij}V_j)^2\Pi_i\frac{\lambda_i}{M_{ii}} \xrightarrow{p} 0,$$

where $\xi_{1,i},\xi_{2,i}$ stay for either e_i or V_i, $\mathcal{V}_i$ stay for e_i^2, e_iV_i, or V_i^2, and a_i stay for either Π_i or $\frac{\lambda_i}{M_{ii}}$.

To prove statement (a), following the arguments in Mikusheva and Sun (2022), we just need to show the following terms are $o(1)$:

$$\mathbb{E}\left[\frac{1}{K}\sum_{i\in[n]}\omega_i^2\mathcal{V}_i\right]^2 \le \frac{C}{K^2}\sum_{i\in[n]}\omega_i^4 \le \frac{C}{K^2}\max_{i\in[n]}\omega_i^2\left(\sum_{i\in[n]}\omega_i^2\right) \le \frac{C}{K^2}p_n\left(\Pi^\top\Pi\right)^2 = o(1),$$

$$\frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\omega_i^2+|\omega_i||\omega_j|)\leq\frac{C}{K}\left(\sum_{i\in[n]}P_{ii}\omega_i^2+\left(\sum_{i\in[n]}P_{ii}\omega_i^2\right)^{1/2}\left(\sum_{j\in[n]}P_{jj}\omega_j^2\right)^{1/2}\right)\leq\frac{C}{K}p_n\Pi^\top\Pi=o(1$$

where we have used $\max_{i\in[n]}\omega_i^2\leq p_n\Pi^\top\Pi$, $\sum_{i\in[n]}\omega_i^2\leq C\Pi^\top\Pi$, and Mikusheva and Sun (2022, Lemma S1.3(b)).

To prove statement (b), we show that

$$\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}(P_{ij}^2\omega_i^4+P_{ij}^2w_i^2w_j^2+P_{ij}^4w_i^2+P_{ij}^4|\omega_i\omega_j|)$$

$$\leq\frac{C}{K^2}\left(p_n\sum_{i\in[n]}\omega_i^4+\left(\sum_{i\in[n]}P_{ii}\omega_i^4\right)^{1/2}\left(\sum_{j\in[n]}P_{jj}\omega_j^4\right)^{1/2}+\sum_{i\in[n]}P_{ii}\omega_i^2p_n+p_n\left(\sum_{i\in[n]}P_{ii}\omega_i^2\right)^{1/2}\left(\sum_{j\in[n]}P_{jj}\omega_j\right.$$

$$\leq\frac{C}{K^2}\left(p_n^2(\Pi^\top\Pi)^2+p_n^2(\Pi^\top\Pi)^2+p_n^2(\Pi^\top\Pi)+p_n^2(\Pi^\top\Pi)\right)=o(1),$$

$$\frac{C}{K^2}\left(\sum_{i\in[n]}\omega_i^2+\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2|\omega_i\omega_j|\right)\leq\frac{C}{K^2}\left(\Pi^\top\Pi+p_n\Pi^\top\Pi\right)=o(1),$$

where we have used $\sum_{i\in[n]}\omega_i^2\leq C\Pi^\top\Pi$ and $\sum_{i\in[n]}\omega_i^4\leq Cp_n(\Pi^\top\Pi)^2$.

To prove statement (c), we show that, for $a_i=\Pi_i$ or λ_i/M_{ii},

$$\frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}^2a_i^2+\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\,|a_ia_j|\right)\leq\frac{C}{K^2}\left(p_n^2a^\top a+p_na^\top a\right)=o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\omega_i^4\frac{\lambda_i^2}{M_{ii}^2}\leq\frac{C}{K^2}\left(\max_{i\in[n]}\omega_i^2\right)^2\sum_{i\in[n]}\lambda_i^2\leq Cp_n^2\left(\frac{\Pi^\top\Pi}{K}\right)^3=o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\omega_i^4\Pi_i^2\leq\frac{C}{K^2}\sum_{i\in[n]}\omega_i^4\leq\frac{C}{K^2}p_n\left(\Pi^\top\Pi\right)^2=o(1),\ \text{where we have used}\ \max_{i\in[n]}|\Pi_i|\leq C,$$

$$\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^4\left(a_i^2+|a_i|\,|a_j|\right)\leq\frac{C}{K^2}\left(p_n^2a^\top a+p_n^2a^\top a\right)=o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(\omega_i^2a_i^2+|\omega_ia_i||\omega_ja_j|)\leq\frac{C}{K^2}\left(p_n^2(\Pi^\top\Pi)(a^\top a)+p_n(\Pi^\top\Pi)(a^\top a)\right)=o(1).$$

To prove statement (d), we first show that

$$\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\,|a_i|\right)^2+\left(\sum_{i\in[n]}|\omega_i a_i|\right)^2\right)=o(1).$$

In particular, when $a_i=\Pi_i$, we have

$$\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\,|\Pi_i|\right)^2+\left(\sum_{i\in[n]}|\omega_i\Pi_i|\right)^2\right)\le\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\right)^2+\left(\sum_{i\in[n]}|\omega_i\Pi_i|\right)^2\right)$$
$$\le\frac{C}{K^2}\left(\left(\Pi^\top\Pi\right)^2+\left(\sum_{i\in[n]}\omega_i^2\right)\left(\Pi^\top\Pi\right)\right)\le\frac{C}{K^2}\left((\Pi^\top\Pi)^2+(\Pi^\top\Pi)^2\right)=o(1),$$

When $a_i=\frac{\lambda_i}{M_{ii}}$, we have

$$\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\left|\frac{\lambda_i}{M_{ii}}\right|\right)^2+\left(\sum_{i\in[n]}\left|\omega_i\frac{\lambda_i}{M_{ii}}\right|\right)^2\right)\le\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^4\right)(\lambda^\top\lambda)+\left(\sum_{i\in[n]}\omega_i^2\right)(\lambda^\top\lambda)\right)$$
$$\le\frac{C}{K^2}\left(p_n(\Pi^\top\Pi)^2(\lambda^\top\lambda)+(\Pi^\top\Pi)(\lambda^\top\lambda)\right)=o(1).$$

Furthermore, we can show that

$$\frac{C}{K^2}\left(\sum_{i\in[n]}|\omega_i a_i|\right)^2\le\frac{C}{K^2}(\Pi^\top\Pi)(a^\top a)=o(1),$$

$$\frac{C}{K}\sum_{i\in[n]}P_{ii}\,|a_i|\le\frac{C}{K}\left(\sum_{i\in[n]}P_{ii}^2\right)^{1/2}\left(a^\top a\right)^{1/2}\le\frac{C}{K}(p_nK)^{1/2}\left(a^\top a\right)^{1/2}=o(1),$$

$$\frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}\,|a_i|\right)^2\le\frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}^2\right)\left(a^\top a\right)\le\frac{C}{K^2}p_nK\left(a^\top a\right)=o(1).$$

To prove statement (e), we show that

$$\left|\frac{C}{K}\sum_{i\in[n]}\omega_i^2\Pi_i\frac{\lambda_i}{M_{ii}}\right| \leq \frac{C}{K}\sum_{i\in[n]}\omega_i^2\left|\frac{\lambda_i}{M_{ii}}\right| \leq \frac{C}{K}\left(\sum_{i\in[n]}\omega_i^4\right)^{1/2}\left(\lambda^\top\lambda\right)^{1/2} \leq \frac{C}{K}p_n^{1/2}(\Pi^\top\Pi)(\lambda^\top\lambda)^{1/2} = o(1),$$

$$\frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\neq j}P_{ij}\omega_i\Pi_i\frac{\lambda_i}{M_{ii}}\right)^2 \leq \frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\neq j}|P_{ij}||\omega_i||\lambda_i|\right)^2 \leq \frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\neq j}\omega_i^2\right)\left(\sum_{i\neq j}P_{ij}^2\lambda_i^2\right)$$
$$\leq \frac{CKp_n^{1/2}\Pi^\top\Pi\lambda^\top\lambda}{K^2} = o(1),$$

$$\frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\neq j}P_{ij}^2\Pi_i\frac{\lambda_i}{M_{ii}}\right)^2 \leq \frac{C}{K^2}\sum_{j\in[n]}\left(\sum_{i\neq j}P_{ij}^2|\lambda_i|\right)^2 \leq \frac{CKp_n\lambda^\top\lambda}{K^2} = o(1),$$

$$\frac{C}{K}\sum_{j\in[n]}\sum_{i\neq j}P_{ij}^2\left|\Pi_i\frac{\lambda_i}{M_{ii}}\right| \leq \frac{C}{K}\sum_{i\in n}\sum_{j\in[n]}P_{ij}^2|\Pi_i\lambda_i| \leq \frac{C}{K}p_n(\Pi^\top\Pi)^{1/2}(\lambda^\top\lambda)^{1/2} = o(1),$$

$$\frac{C}{K^2}\sum_{j\in[n]}\sum_{k\neq j}\left(\sum_{i\neq j,k}P_{ij}^2P_{ik}^2\Pi_i\frac{\lambda_i}{M_{ii}}\right)^2 \leq \frac{C}{K^2}\sum_{j\in[n]}\sum_{k\neq j}\left(\sum_{i\neq j,k}P_{ij}^2P_{ik}^2|\lambda_i|\right)^2$$
$$\leq \frac{C}{K^2}\left(\sum_{j\in[n]}\sum_{k\neq j}\sum_{i\neq j,k}P_{ij}^4P_{ik}^4\right)\lambda^\top\lambda \leq \frac{Cp_n^3K\lambda^\top\lambda}{K^2} = o(1),$$

where we have used Mikusheva and Sun (2022, Lemma S1.1(ii)).

Finally, we can show that Mikusheva and Sun (2022, Lemma S3.1) also holds under our conditions by using similar arguments. We omit the details for brevity.

A.19 Lemma A.19.1 and Its Proof

Lemma A.19.1. *Suppose assumptions in Theorem A.1.1 hold. Then, we have*

$$\hat{\gamma}_e = O_P(n^{-1/2}), \quad \hat{\gamma}_V = O_P(n^{-1/2}), \quad Q_{\tilde{e},W} = O_P(1), \quad Q_{\tilde{V},W} = O_P(1),$$
$$\hat{\gamma}_V^\top Q_{W,W^\top}\hat{\gamma}_V = o_P(1), \quad \hat{\gamma}_e^\top Q_{W,W^\top}\hat{\gamma}_e = o_P(1), \quad \textit{and} \quad \hat{\gamma}_e^\top Q_{W,W^\top}\hat{\gamma}_V = o_P(1).$$

Proof. We have $\hat{\gamma}_e = O_P(n^{-1/2})$ because $\mathbb{E}\tilde{e}_i = 0$ and $\text{mineig}(W^\top W/n) \geq c > 0$. Similarly, we have $\hat{\gamma}_V = O_P(n^{-1/2})$. To see that $Q_{\tilde{e},W} = O_P(1)$, we note that $\mathbb{E}Q_{\tilde{e},W} = 0$ and

$$\mathbb{E}Q_{\tilde{e},W}Q_{\tilde{e},W}^\top \leq C\sum_{i\in[n]}(\sum_{j\neq i}P_{ij}W_j)^\top(\sum_{j\neq i}P_{ij}W_j)/K = C\sum_{i\in[n]}P_{ii}^2W_i^\top W_i/K \leq C,$$

where we use the fact that $\sum_{j\neq i}P_{ij}W_j = -P_{ii}W_i$ since P_{ij} is the ij-th element of $P = Z(Z^\top Z)^{-1}Z^\top$. Similarly, we have $Q_{\tilde{V},W} = O_P(1)$.

To see $\hat{\gamma}_V^\top Q_{W,W^\top}\hat{\gamma}_V = o_P(1)$, we note that

$$\left|\hat{\gamma}_V^\top Q_{W,W^\top}\hat{\gamma}_V\right| \leq \sum_{i\in[n]}(W_i^\top\hat{\gamma}_V)^2/\sqrt{K} = o_P(1),$$

where we use the fact that $\sum_{i\in[n]}W_iW_i^\top/n = O_P(1)$ and $\hat{\gamma}_V = O_P(n^{-1/2})$, so that

$$\sum_{i\in[n]}(W_i^\top\hat{\gamma}_V)^2 = O_P(1).$$

Similarly, we can show that

$$\hat{\gamma}_e^\top Q_{W,W^\top}\hat{\gamma}_e = o_P(1), \quad \text{and} \quad \hat{\gamma}_e^\top Q_{W,W^\top}\hat{\gamma}_V = o_P(1).$$

□

A.20 Comparison with HLIM Estimator under Strong Identification

We consider the model in Section A.1 and the HLIM estimator proposed by Hausman et al. (2012). Specifically, Hausman et al. (2012) estimate (β,γ) by $(\hat{\beta}^{HLIM},\hat{\gamma}^{HLIM})$ defined as

$$(\hat{\beta}^{HLIM},\hat{\gamma}^{HLIM}) = \arg\min_{b,r}\mathcal{Q}(b,r), \quad \mathcal{Q}(b,r) = \frac{\sum_{i\in[n]}\sum_{j\neq i}(\tilde{Y}_i - \tilde{X}_ib - W_i^\top r)\tilde{P}_{ij}(\tilde{Y}_i - \tilde{X}_ib - W_i^\top r)}{\sum_{i\in[n]}(\tilde{Y}_i - \tilde{X}_ib - W_i^\top r)^2},$$

where $\tilde{P}_{ij}$ is the projection matrix constructed by $(W_i^\top, \tilde{Z}_i^\top)^\top$. Following Hausman et al. (2012), we let $\tilde{\Pi}_i = \mu_n \tilde{\pi}_i/\sqrt{n}$ such that $\sum_{i\in[n]} \tilde{\pi}_i^2/n \geq c > 0$ for some constant c. As explained in the paper, under strong identification, we have $\mu_n^2/\sqrt{K} \to \infty$. In both cases considered in Hausman et al. (2012, Assumption 6), the convergence rate can be unified as $\sqrt{K}/\mu_n^2$. Then, the Wald statistic can be written as

$$W_h(\beta_0) = \frac{\mu_n^2(\hat{\beta}^{HLIM} - \beta_0)/\sqrt{K}}{\hat{\Phi}_h^{1/2}},$$

where $\hat{\Phi}_h$ is a consistent estimator of Φ_h, and Φ_h is the asymptotic variance of $\hat{\beta}^{HLIM}$. To study the behaviour of $W_h(\beta_0)$ under strong identification and local alternatives, we let β_0 denote the local alternative in the sense that $\beta_0 = \beta + \frac{\widetilde{\Delta}}{\mu_n^2/\sqrt{K}}$. We will provide the expression for Φ_h later. We also note that the notation in Hausman et al. (2012) and our paper is different. Specifically, their δ_0 is our $(\gamma^\top, \beta_0)^\top$, their $\hat{\delta}$ is our $((\hat{\gamma}^{HLIM})^\top, \hat{\beta}^{HLIM})^\top$, their X_i is our $(W_i^\top, \tilde{X}_i)^\top$, their Z_i is our $(W_i^\top, \tilde{Z}_i^\top)^\top$, and thus their projection matrix P is our $\tilde{P}$, which is the one based on $(W_i^\top, \tilde{Z}_i^\top)^\top$. We use P and P_W to denote the projection matrices based on our Z_i and W_i, respectively, where $Z_i = ([M_W]_{i\cdot}\tilde{Z})^\top$, $[M_W]_{i\cdot}$ is the ith row of M_W, and $M_W = I_n - P_W$.

Further denote L as a matrix that selects the last element of $\hat{\delta} = ((\hat{\gamma}^{HLIM})^\top, \hat{\beta}^{HLIM})^\top$ and

$$S_n = \begin{pmatrix} I_d & 0 \\ \pi_x^\top & 1 \end{pmatrix} \operatorname{diag}(\sqrt{n}, \cdots, \sqrt{n}, \mu_n),$$

where $\pi_x = (W^\top W)^{-1} W^\top \tilde{\Pi}$ is the projection coefficient of $\tilde{\Pi}$ on W. Then, the corresponding definition of $\hat{D}(\delta_0)$ in Hausman et al. (2012, p.235) under our notation is as follows:

$$\hat{D}(\delta_0) = \frac{\sum_{i\in[n]} \sum_{j\neq i} \left[\mathbb{W}_i \tilde{P}_{ij} \overline{e}_j(\beta_0) - \overline{e}_i(\beta_0) \tilde{P}_{ij} \overline{e}_j(\beta_0) \frac{\mathbb{W}^\top \overline{e}(\beta_0)}{\overline{e}^\top(\beta_0)\overline{e}(\beta_0)}\right]}{\sqrt{K}},$$

where $\mathbb{W}_i = (W_i^\top, \tilde{X}_i)^\top$, $\mathbb{W}$ is a $n \times (d+1)$ matrix with its ith row being $\mathbb{W}_i^\top$ where d is the dimension of W_i, and $\overline{e}_i(\beta_0) = \tilde{e}_j - \tilde{X}_j(\beta_0 - \beta)$. In addition, we note that $\overline{X}_i = \tilde{X}_i - W_i^\top \pi_x = \Pi_i + \tilde{V}_i$ as defined in Theorem A.1.1, $X_i = \overline{X}_i - W_i^\top \hat{\gamma}_V$, $\overline{e}_i(\beta_0) = e_i(\beta_0) + W_i^\top \hat{\gamma}_e - W_i^\top \hat{\pi}_x(\beta_0 - \beta)$, where $\pi_x = (W^\top W)^{-1}(W^\top \tilde{\Pi})$, $\hat{\pi}_x = (W^\top W)^{-1}(W^\top \tilde{X}) = \pi_x + \hat{\gamma}_V$,

$\hat{\gamma}_V = (W^\top W)^{-1}(W^\top \tilde{V})$, and $\hat{\gamma}_e = (W^\top W)^{-1}(W^\top \tilde{e})$. Further let $\overline{\delta}$ be between $\delta = (\gamma^\top, \beta)^\top$ and δ_0.

Then, following the argument in the proof of Hausman et al. (2012, Theorem 2), we have

$$
\begin{aligned}
&(\mu_n^2/\sqrt{K})(\hat{\beta}^{HLIM} - \beta_0) \\
&= (\mu_n^2/\sqrt{K})L(\hat{\delta} - \delta_0) \\
&= -(\mu_n^2/\sqrt{K})L\left(\frac{\partial \hat{D}(\overline{\delta})}{\partial \delta}\right)^{-1}\hat{D}(\delta_0) \\
&= -(\mu_n^2/\sqrt{K})L(S_n^\top)^{-1}\left(S_n^{-1}\frac{\partial \hat{D}(\overline{\delta})}{\partial \delta}(S_n^\top)^{-1}\right)^{-1} S_n^{-1}\hat{D}(\delta_0) \\
&= -(\mu_n^2/\sqrt{K})(0, 1/\mu_n)(H^{-1} + o_P(1))\mathrm{diag}(1/\sqrt{n}, \cdots, 1/\sqrt{n}, 1/\mu_n)\begin{pmatrix} I_d & 0 \\ -\pi_x^\top & 1 \end{pmatrix}\hat{D}(\delta_0) \\
&= -\frac{\mu_n}{\sqrt{K}}\left(\left((H^{21} + o_P(1))/\sqrt{n} - \pi_x^\top(H^{22} + o_P(1))/\mu_n, (H^{22} + o_P(1))/\mu_n\right)\right)\hat{D}(\delta_0) \\
&= (H^{22} + o_P(1))(-\pi_x^\top, 1)\hat{D}(\delta_0)/\sqrt{K} \\
&= (H^{22} + o_P(1))\frac{\sum_{i\in[n]}\sum_{j\neq i}\left[\overline{X}_i\tilde{P}_{ij}\overline{e}_j(\beta_0) - \overline{e}_i(\beta_0)\tilde{P}_{ij}\overline{e}_j(\beta_0)\frac{\overline{X}^\top\overline{e}(\beta_0)}{\overline{e}^\top(\beta_0)\overline{e}(\beta_0)}\right]}{\sqrt{K}},
\end{aligned}
$$

where by Hausman et al. (2012, Lemma A7), $S_n^{-1}\frac{\partial \hat{D}(\overline{\delta})}{\partial \delta}(S_n^\top)^{-1} \xrightarrow{p} H$, and we denote $H^{-1} = \begin{pmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{pmatrix}$.

Following the same argument in the proof of Lemma A.19.1, we can show that

$$
\begin{aligned}
&\frac{\sum_{i\in[n]}\sum_{j\neq i}\hat{\gamma}_V^\top W_i\tilde{P}_{ij}\overline{e}_j(\beta_0)}{\sqrt{K}} = o_P(1), \quad \frac{\sum_{i\in[n]}\sum_{j\neq i}\overline{X}_i\tilde{P}_{ij}W_i^\top(\hat{\gamma}_e - \hat{\pi}_x(\beta_0 - \beta))}{\sqrt{K}} = o_P(1) \\
&\frac{\sum_{i\in[n]}\sum_{j\neq i}\overline{e}_i(\beta_0)\tilde{P}_{ij}W_i^\top(\hat{\gamma}_e - \hat{\pi}_x(\beta_0 - \beta))}{\sqrt{K}} = o_P(1), \quad \text{and} \\
&\frac{\sum_{i\in[n]}\sum_{j\neq i}(\hat{\gamma}_e - \hat{\pi}_x(\beta_0 - \beta))^\top W_i\tilde{P}_{ij}W_i^\top(\hat{\gamma}_e - \hat{\pi}_x(\beta_0 - \beta))}{\sqrt{K}} = o_P(1).
\end{aligned}
$$

In addition, we have $\overline{X}^\top \overline{e}(\beta_0)/\overline{e}^\top(\beta_0)\overline{e}(\beta_0) \xrightarrow{p} \tilde{\rho}$. Then, we have

$$\mu_n^2(\hat{\beta}^{HLIM} - \beta_0)/\sqrt{K} = H^{22} \frac{\sum_{i\in[n]}\sum_{j\neq i}\left[X_i\tilde{P}_{ij}e_j(\beta_0) - e_i(\beta_0)\tilde{P}_{ij}e_j(\beta_0)\tilde{\rho}\right]}{\sqrt{K}} + o_P(1).$$

Because $X^\top W = 0$ and $e^\top W = 0$, we have $X^\top \tilde{P}e(\beta_0) = X^\top Pe(\beta_0)$ and $e(\beta_0)^\top \tilde{P}e(\beta_0) = e(\beta_0)^\top Pe(\beta_0)$. Therefore, we have

$$\begin{aligned}
\frac{\sum_{i\in[n]}\sum_{j\neq i} X_i\tilde{P}_{ij}e_j(\beta_0)}{\sqrt{K}} &= \frac{X^\top Pe(\beta_0) - \sum_{i\in[n]} X_i\tilde{P}_{ii}e_i(\beta_0)}{\sqrt{K}} \\
&= \frac{\sum_{i\in[n]}\sum_{j\neq i} X_iP_{ij}e_j(\beta_0) + \sum_{i\in[n]} X_ie_i(\beta_0)(P_{ii} - \tilde{P}_{ii})}{\sqrt{K}} \\
&= Q_{X,e(\beta_0)} - \frac{\sum_{i\in[n]} X_ie_i(\beta_0)P_{W,ii}}{\sqrt{K}} \\
&= Q_{X,e(\beta_0)} + o_P(1),
\end{aligned}$$

where we use the facts that $\tilde{P}_{ii} = P_{ii} + P_{W,ii}$ and

$$\sum_{i\in[n]} X_ie_i(\beta_0)P_{W,ii} = \frac{1}{n}\sum_{i\in[n]} X_ie_i(\beta_0)W_i^\top \left(W^\top W/n\right)^{-1} W_i = O_P(1).$$

Similarly, we have

$$\frac{\sum_{i\in[n]}\sum_{j\neq i} e_i(\beta_0)\tilde{P}_{ij}e_j(\beta_0)}{\sqrt{K}} = Q_{e(\beta_0),e(\beta_0)} + o_P(1),$$

and thus,

$$\mu_n^2(\hat{\beta}^{HLIM} - \beta_0)/\sqrt{K} = H^{22}(Q_{X,e(\beta_0)} - \tilde{\rho}Q_{e(\beta_0),e(\beta_0)}) + o_P(1).$$

In order for the HLIM based Wald test to have a pivotal standard normal distribution in the limit, the asymptotic variance Φ_h must be

$$\Phi_h = (H^{22})^2(\Psi - 2\tilde{\rho}\Phi_{12} + \tilde{\rho}^2\Phi_1),$$

which means the Wald statistic satisfies $W_h(\beta) = \frac{Q_{X,e(\beta_0)} - \tilde{\rho} Q_{e(\beta_0),e(\beta_0)}}{(\Psi - 2\tilde{\rho}\Phi_{12} + \tilde{\rho}^2 \Phi_1)^{1/2}} + o_P(1)$.

A.21 Additional Simulation Results

A.21.1 Additional Simulation Results Based on the Limit Problem

In this section, we present further simulation results for the power behavior of tests under the limit problem described in Section 1.2.

For Figures A.1–A.32, all the settings remain the same as those in Section 1.5.1 in the main paper except we use alternative values of the tuning parameters for (1.3.5). Specifically, for the values of p_1 and p_2 in

$$\underline{a}(\mu_D, \gamma(\beta_0)) = \min \left(p_1, \frac{p_2 \mathbb{C}_{\alpha,\max}(\rho(\beta_0)) \Phi_1(\beta_0) c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0) \mu_D^2} \right),$$

we use $(p_1, p_2) = (0.01, 1.5)$, $(0.01, 2)$, $(0.001, 1.1)$, $(0.001, 1.5)$, $(0.001, 2)$, $(0.1, 1.1)$, $(0.1, 1.5)$, or $(0.1, 2)$, instead of $(0.01, 1.1)$ in Section 1.5. Specifically, Figures A.1–A.4 report the results for $(0.01, 1.5)$, Figures A.5–A.8 report those for $(0.01, 2)$, Figures A.9–A.12 report those for $(0.001, 1.1)$, Figures A.13–A.16 report those for $(0.001, 1.5)$, Figures A.17–A.20 report those for $(0.001, 2)$, Figures A.21–A.24 report those for $(0.1, 1.1)$, Figures A.25–A.28 report those for $(0.1, 1.5)$, and Figures A.29–A.32 report those for $(0.1, 2)$, respectively. We find the results are very similar to those reported in the main paper.

Furthermore, Figures A.33–A.36 present the power curves in the cases with stronger identification ($\mathcal{C} = 9$ or 12). The overall patterns are very similar to those for $\mathcal{C} = 6$. For Figures A.33–A.36, the tuning parameters are set as $(p_1, p_2) = (0.01, 1.1)$, which are same as those in Section 1.5 of the main text. The results for other values of p_1 and p_2 remain very similar and are thus omitted for brevity.

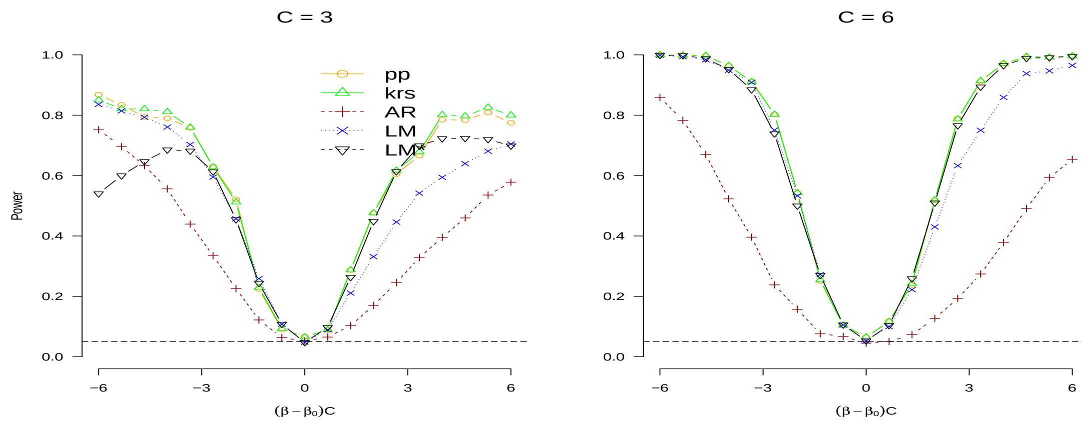

Figure A.1: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.01, 1.5)$

Note: The orange line with circle represents pp, which is the probability of rejection by using the test $\phi_{a_1,a_2,pp}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$; the green line with upward-pointing triangle represents krs, which is the probability of rejection by using the test $\phi_{a_1,a_2,krs}(\delta, \widehat{D}, \widehat{\gamma}(\beta_0))$; the brown dash line with additive sign represents AR test given in (1.2.5); the blue dotted line with cross represents LM test given in (1.2.6); the dark dash line with downward-pointing triangle represents LM^* test defined just above (1.2.7).

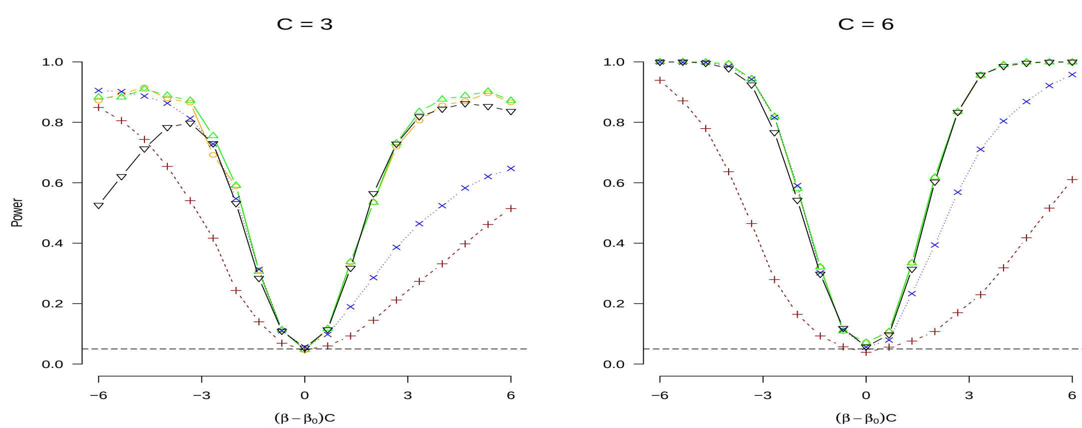

Figure A.2: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.01, 1.5)$

Note: The lines are explained under Figure A.1.

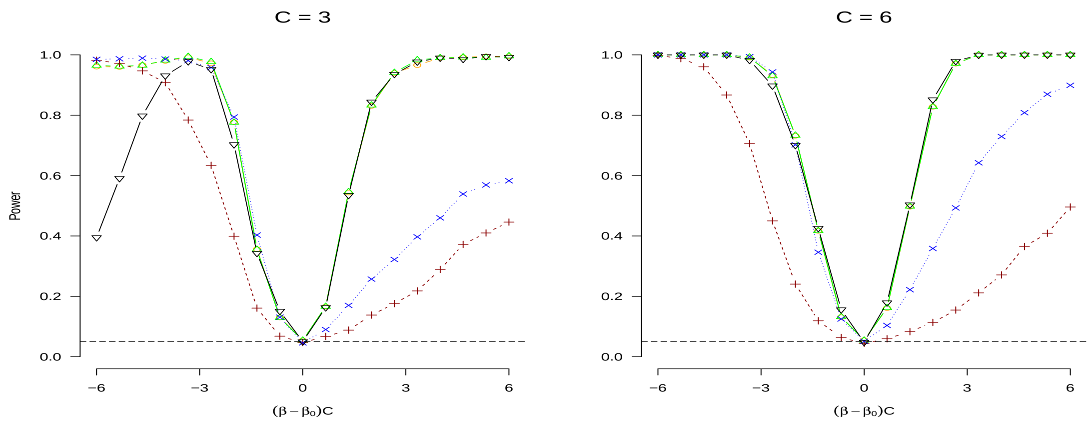

Figure A.3: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.01, 1.5)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

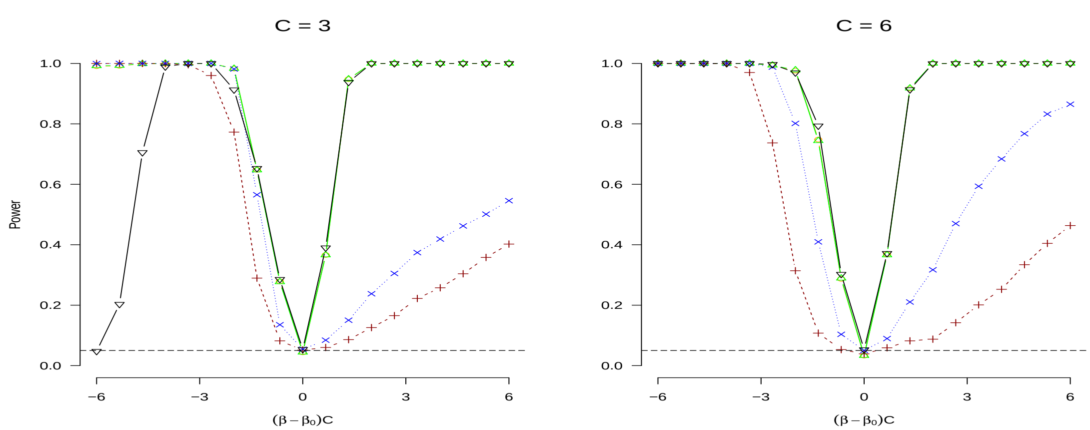

Figure A.4: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.01, 1.5)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

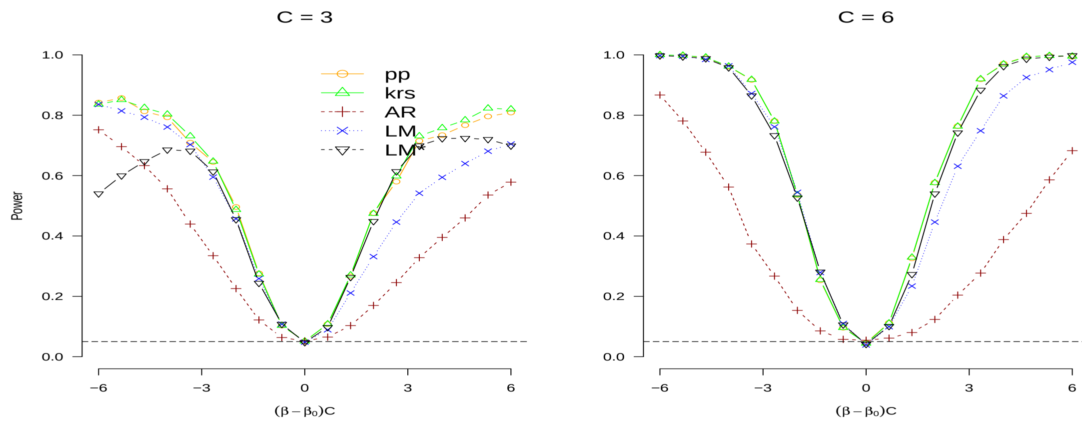

Figure A.5: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.01, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

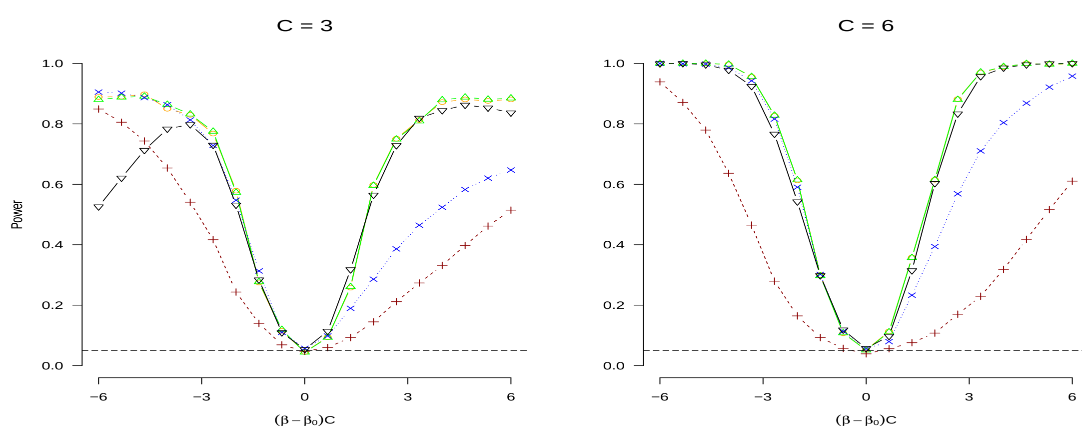

Figure A.6: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.01, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

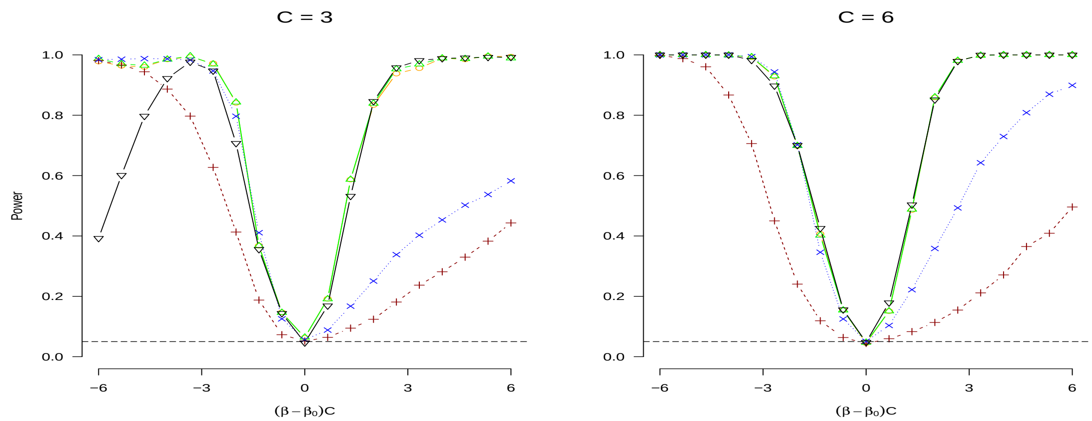

Figure A.7: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.01, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

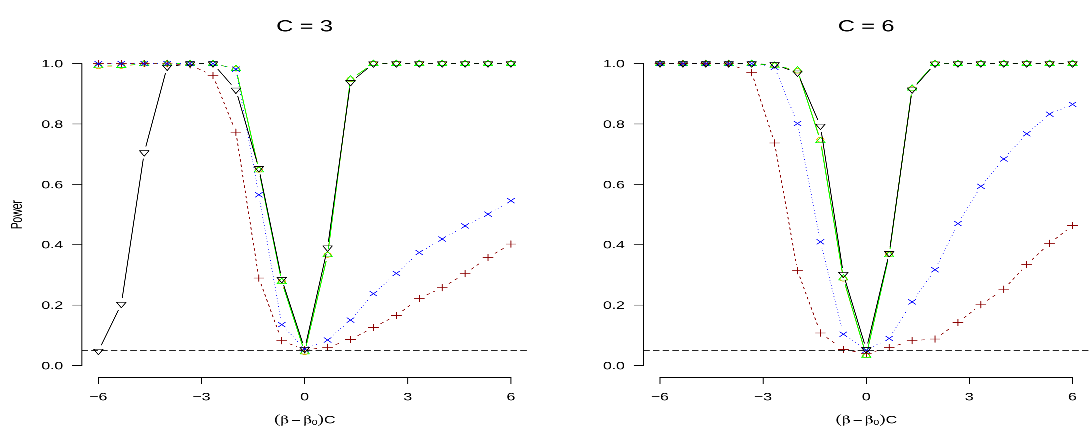

Figure A.8: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.01, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

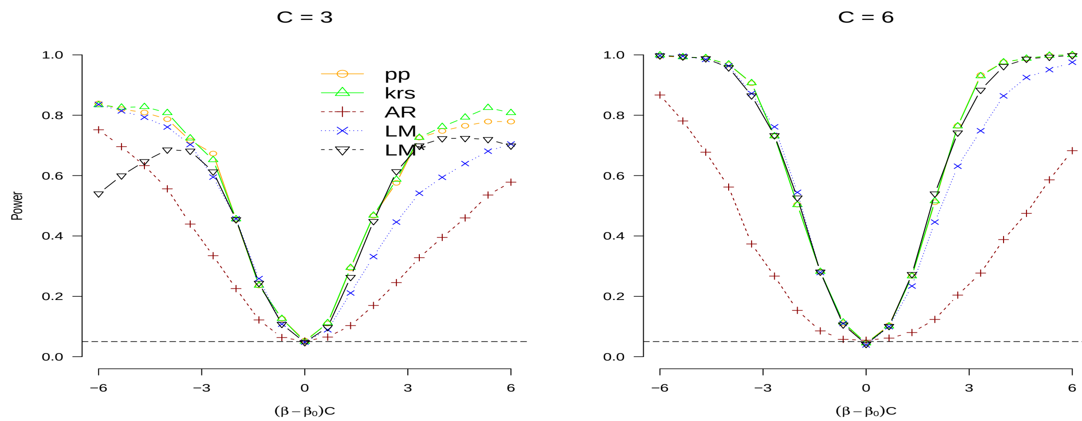

Figure A.9: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.001, 1.1)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

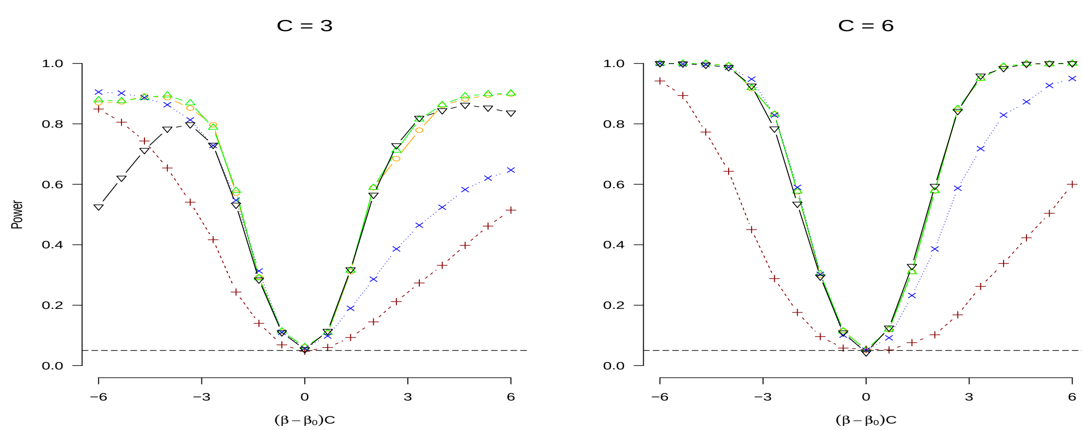

Figure A.10: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.001, 1.1)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

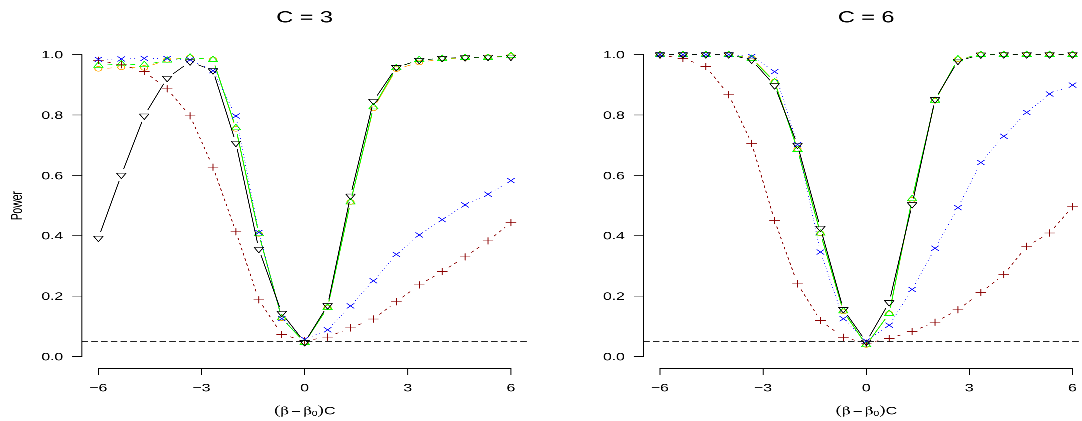

Figure A.11: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.001, 1.1)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

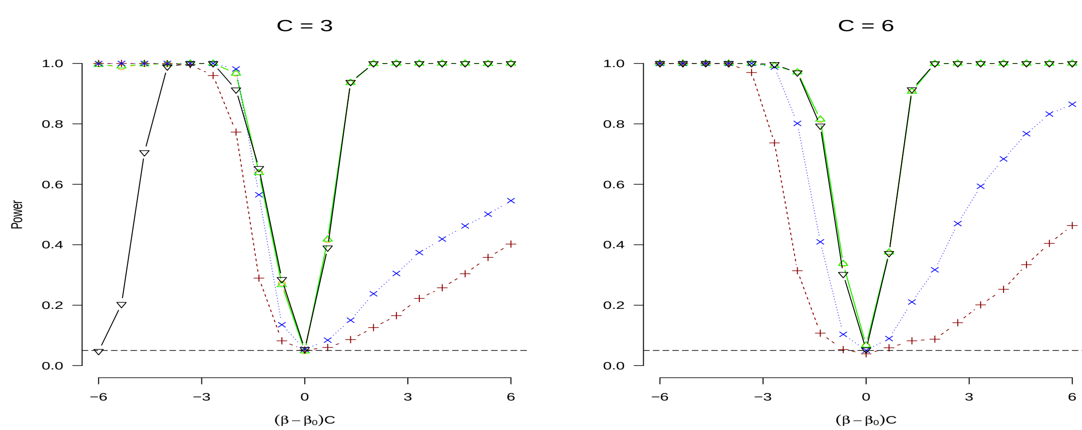

Figure A.12: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.001, 1.1)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

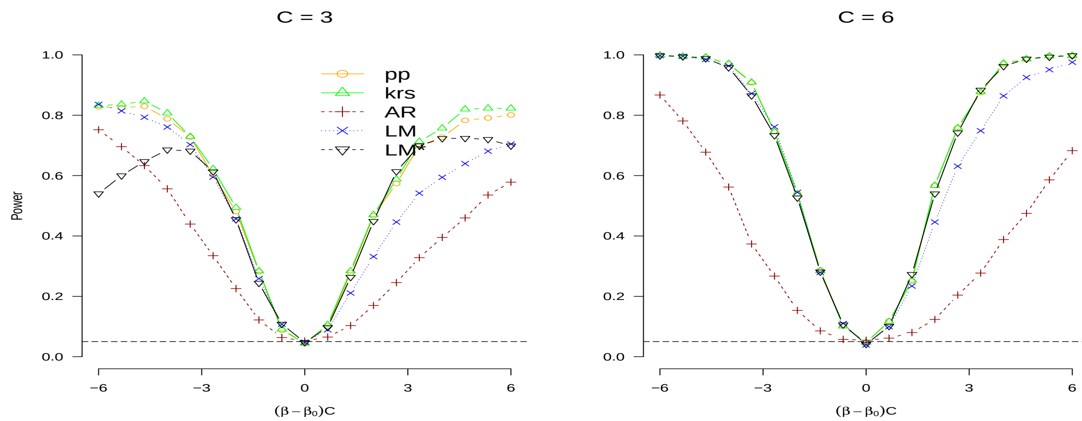

Figure A.13: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.001, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

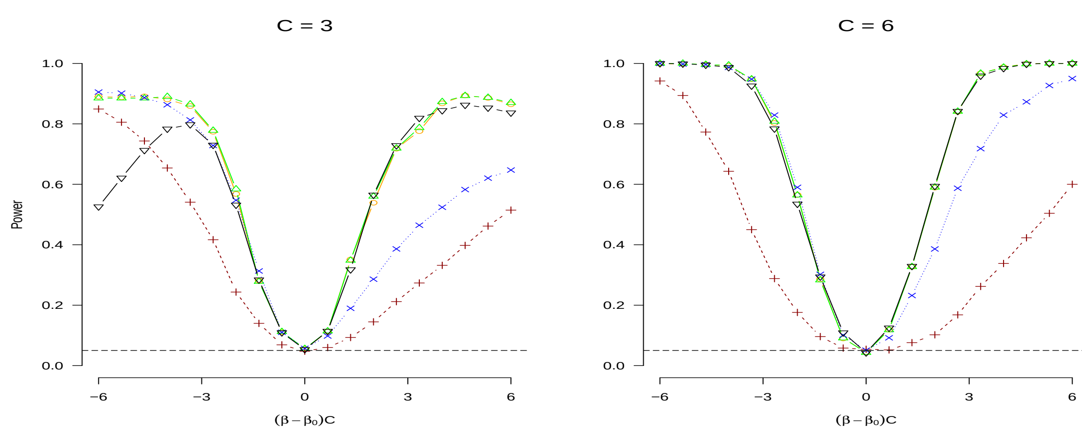

Figure A.14: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.001, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

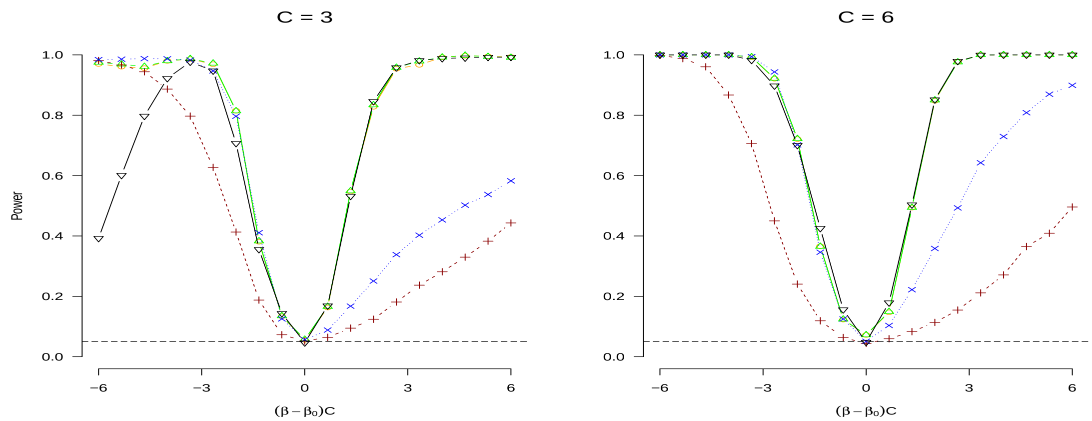

Figure A.15: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.001, 1.5)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

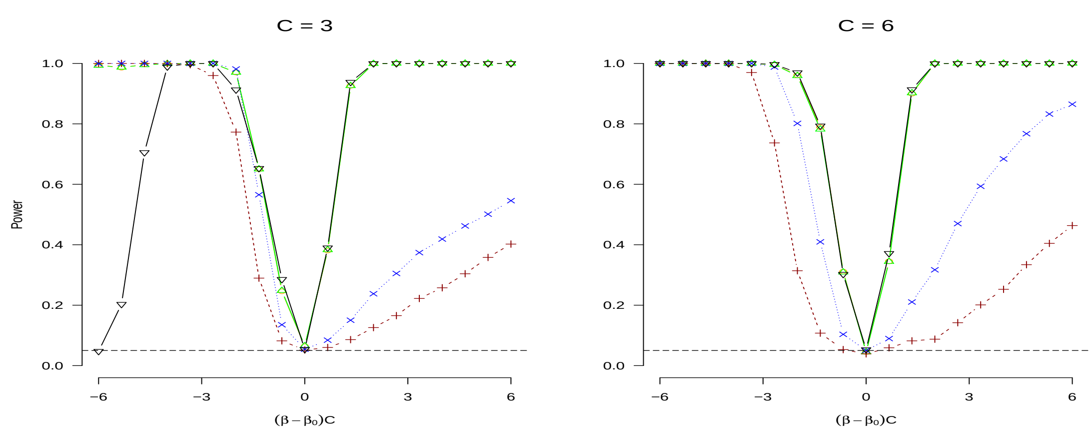

Figure A.16: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.001, 1.5)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

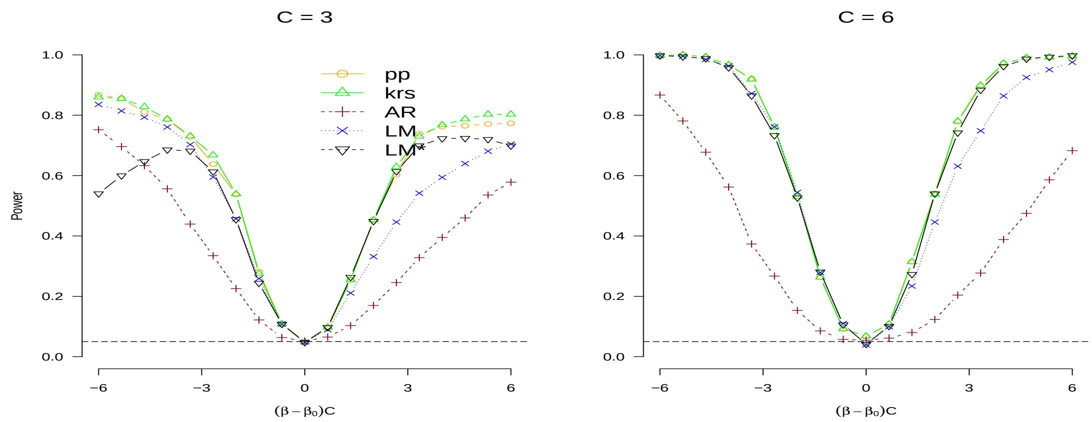

Figure A.17: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.001, 2)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

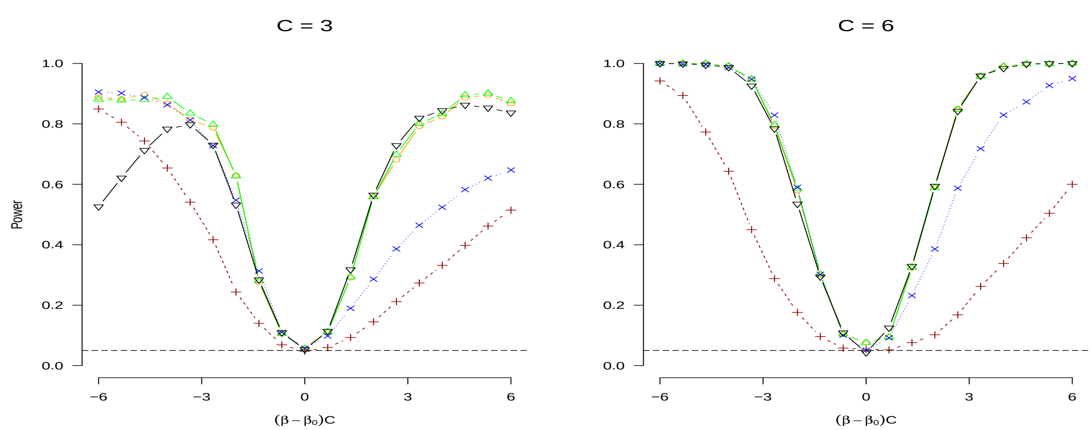

Figure A.18: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.001, 2)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

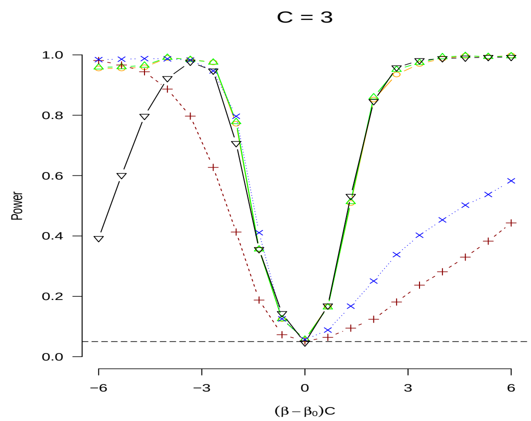

Figure A.19: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.001, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

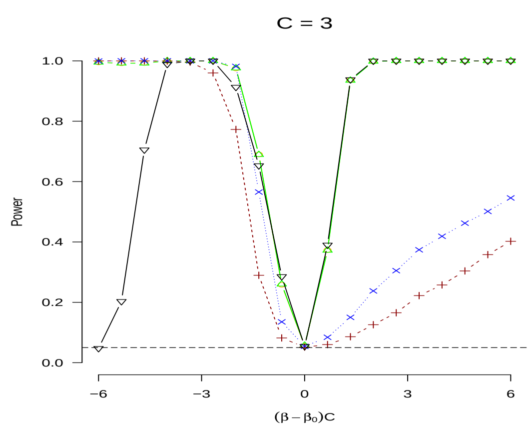

Figure A.20: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.001, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

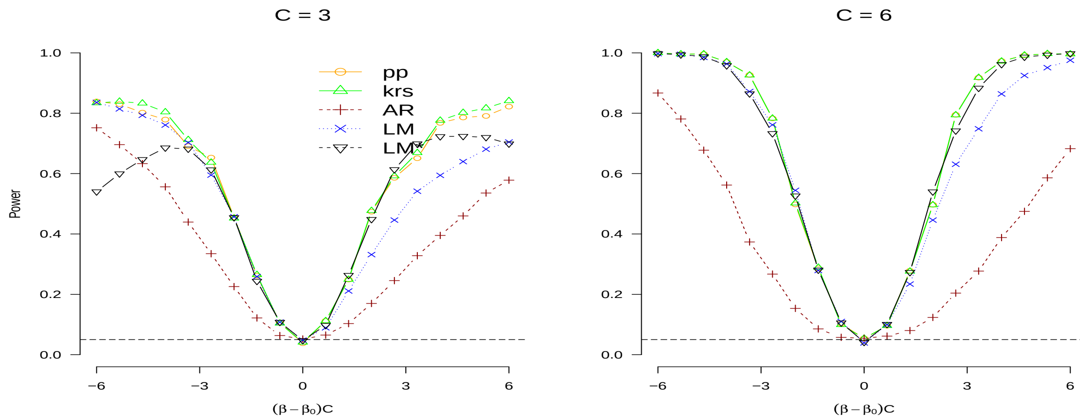

Figure A.21: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.1, 1.1)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

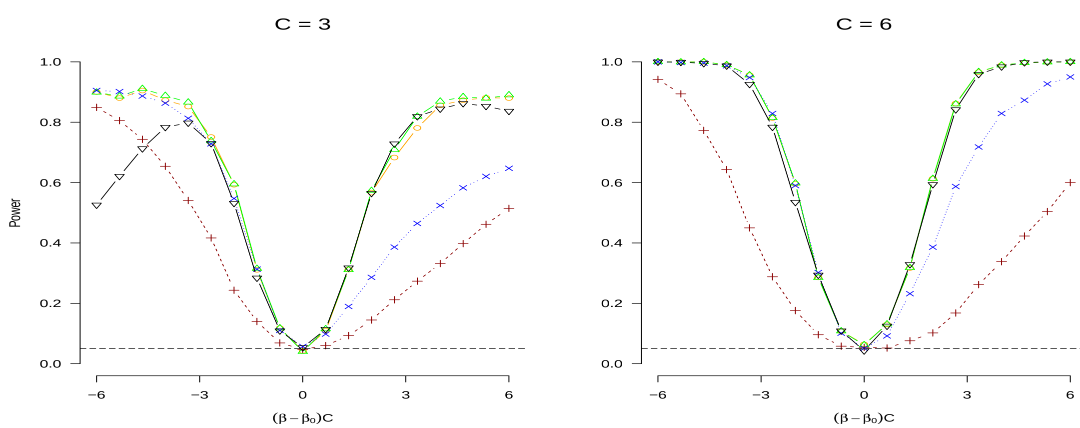

Figure A.22: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.1, 1.1)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

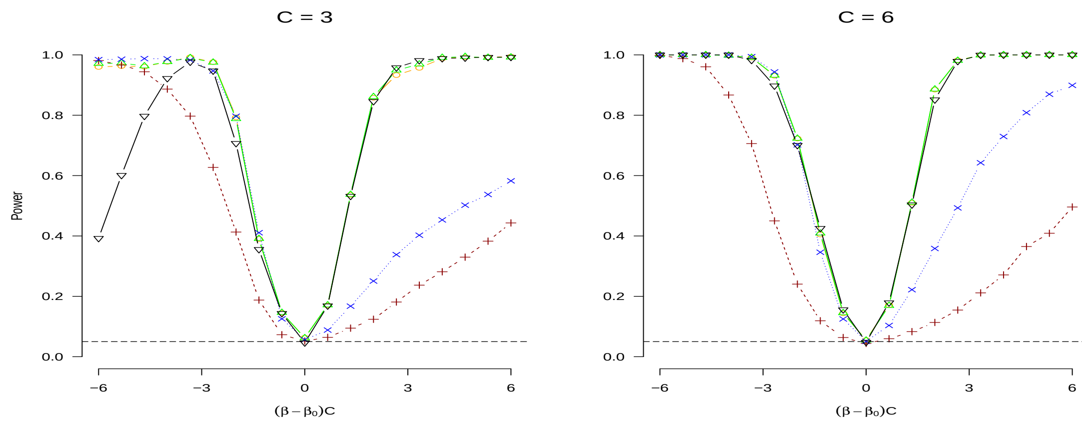

Figure A.23: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.1, 1.1)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

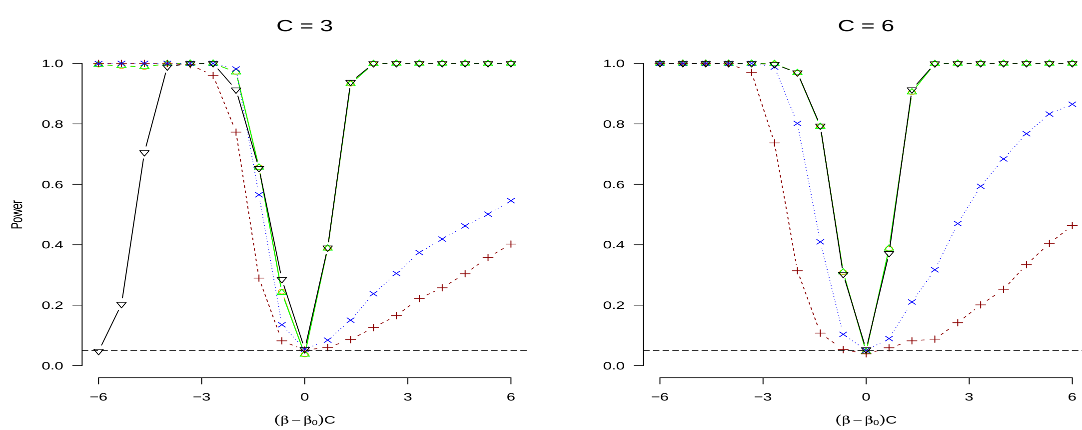

Figure A.24: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.1, 1.1)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

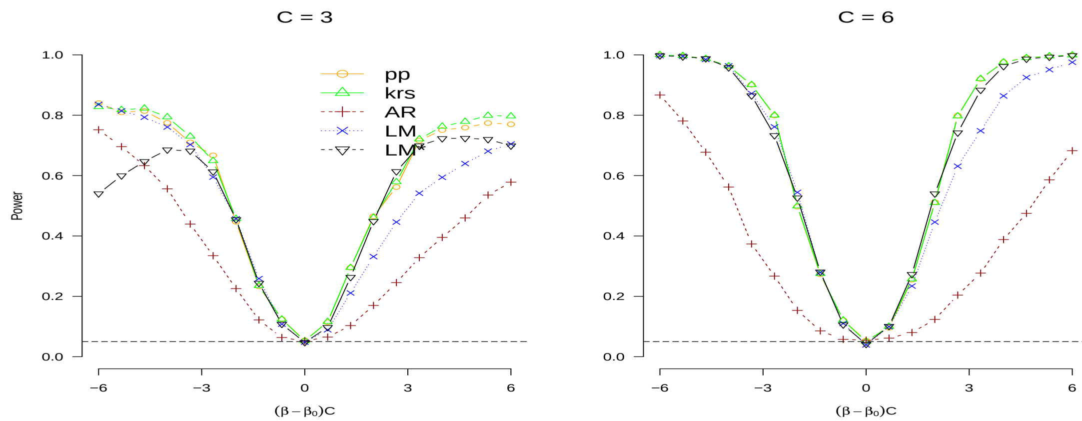

Figure A.25: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.1, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

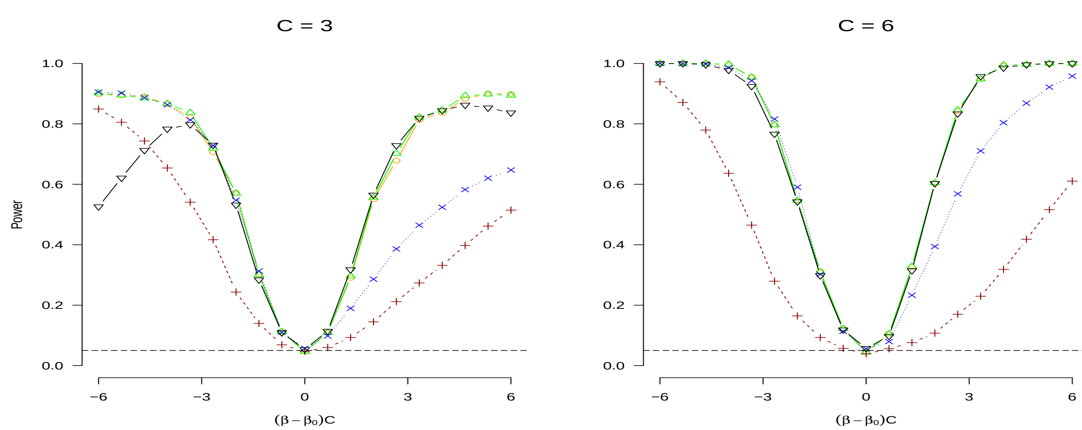

Figure A.26: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.1, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

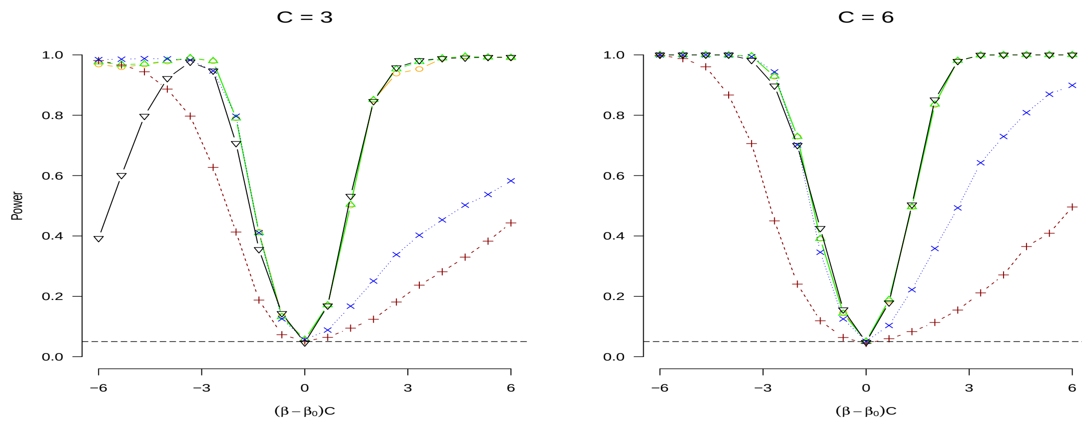

Figure A.27: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.1, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

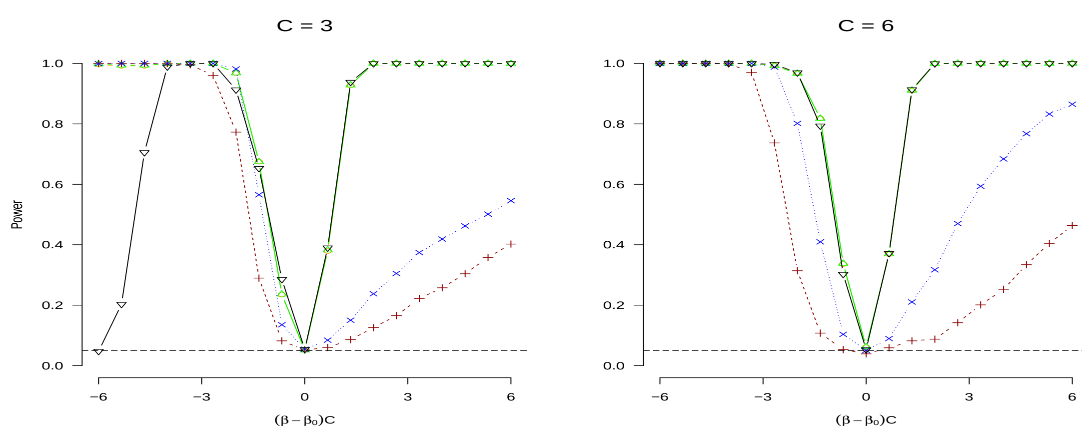

Figure A.28: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.1, 1.5)$ with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

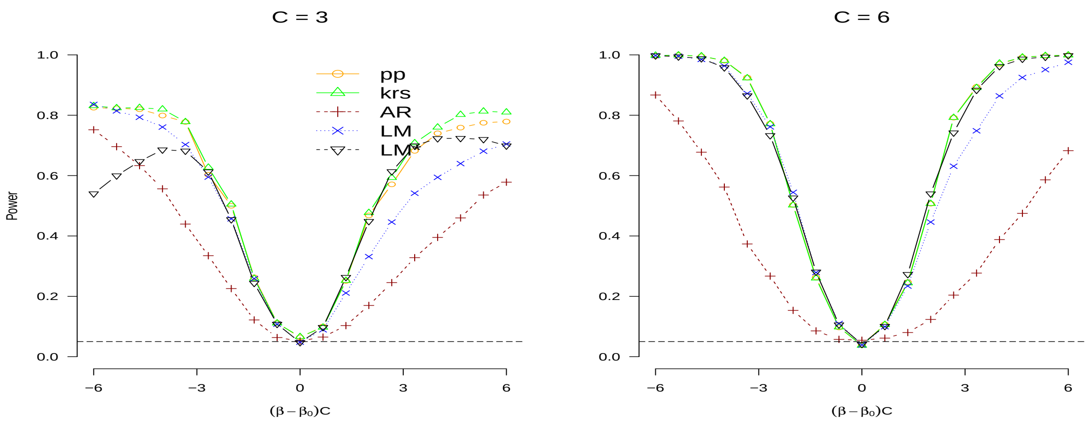

Figure A.29: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.1, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

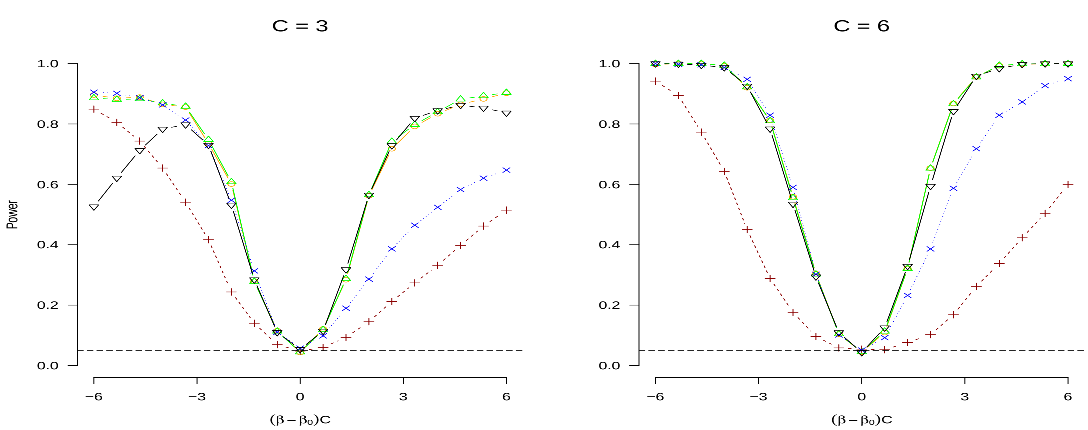

Figure A.30: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.1, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

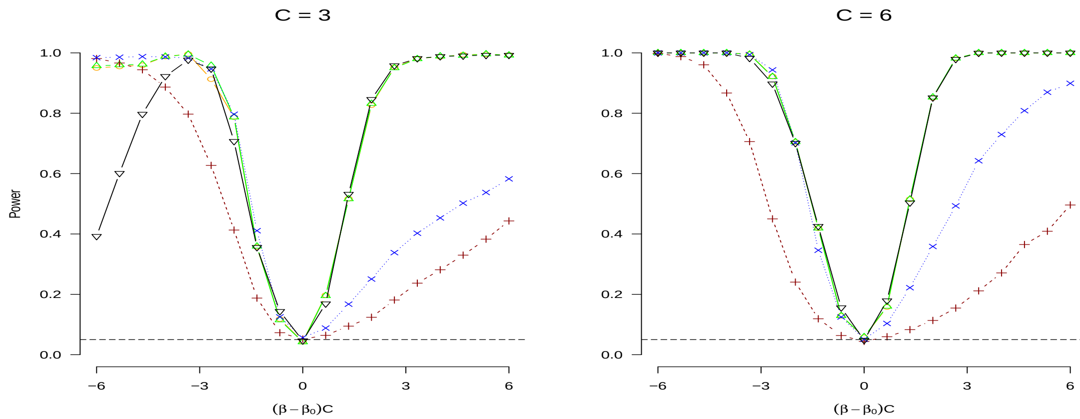

Figure A.31: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.1, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

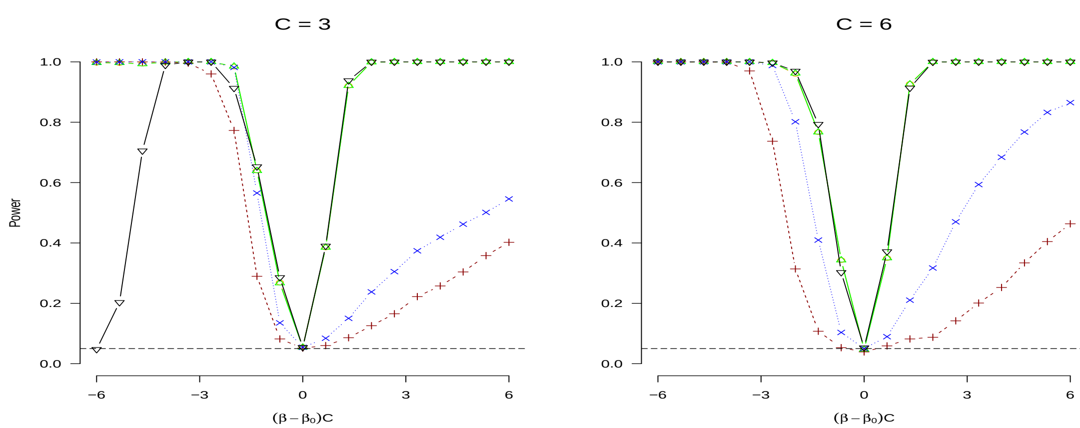

Figure A.32: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.1, 2)$ with 5% nominal size represented by the horizontal dotted line

Note: The lines are explained under Figure A.1.

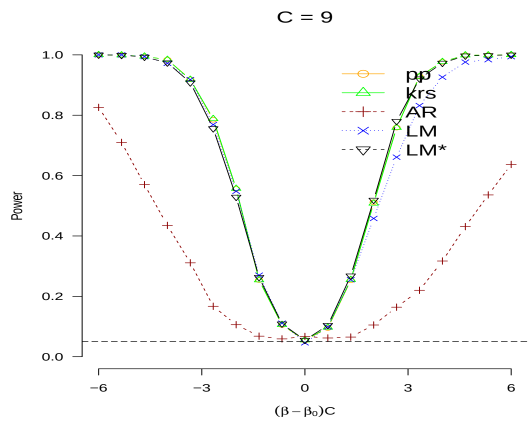

Figure A.33: Power Curve for $\rho = 0.2$ and $(p_1, p_2) = (0.01, 1.1)$, $C = 9$ or 12 with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

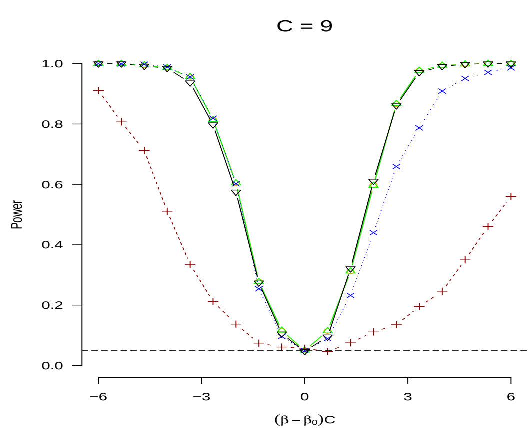

Figure A.34: Power Curve for $\rho = 0.4$ and $(p_1, p_2) = (0.01, 1.1)$, $C = 9$ or 12 with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

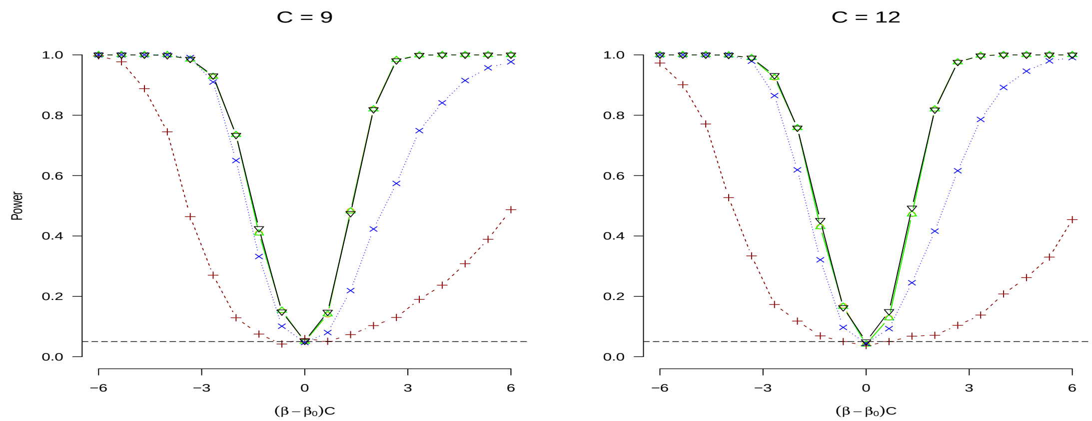

Figure A.35: Power Curve for $\rho = 0.7$ and $(p_1, p_2) = (0.01, 1.1)$, $C = 9$ or 12 with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

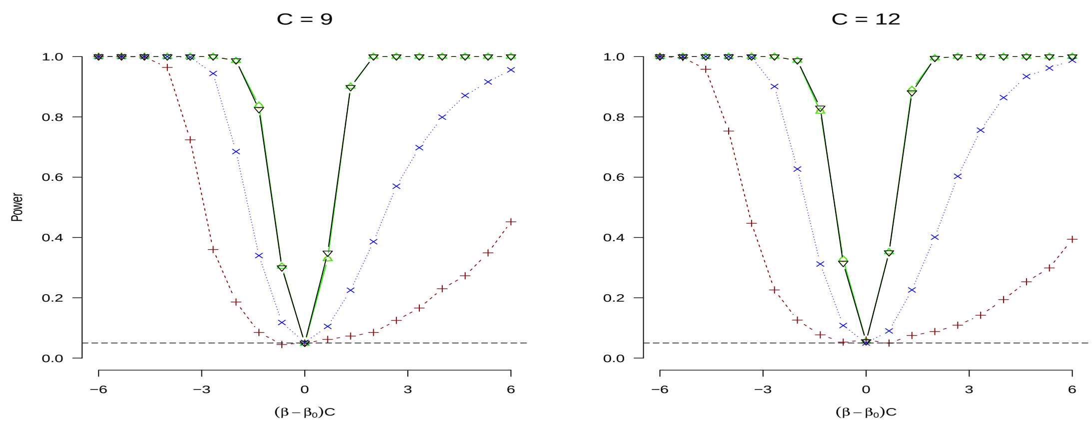

Figure A.36: Power Curve for $\rho = 0.9$ and $(p_1, p_2) = (0.01, 1.1)$, $C = 9$ or 12 with 5% nominal size represented by the horizontal dotted line
Note: The lines are explained under Figure A.1.

A.21.2 Additional Simulation Results Based on the Calibrated Data

We run two sets of robustness checks for the calibrated data provided in Section 1.5.2. For the first set, we retained the parameter space of $\mathcal{B} = [-0.5, 0.5]$ and used 16 grid-points in total over this space, instead of 31 grid-points used in the main text. As in the previous section, we vary over (p_1, p_2) equals $(0.001, 1.1)$, $(0.001, 1.5)$, $(0.001, 2)$, $(0.01, 1.5)$, $(0.01, 2)$, $(0.1, 1.1)$, $(0.1, 1.5)$, and $(0.1, 2)$. Figures A.37–A.44 are results for DGP 1, while Figures A.45–A.52 are results for DGP 2. We find that our results are very similar to the main text's specification, i.e. $(p_1, p_2) = (0.01, 1.1)$.

For the second set of robustness checks, we fix $(p_1, p_2) = (0.01, 1.1)$ as in the main text and vary the parameter space as $\mathcal{B}_2 = [-0.25, 0.25]$ and $\mathcal{B}_3 = [-1, 1]$ over 21 equally-sized grid-points. This is done in order to capture the null of $H_0 : \beta = 0.1$. DGP 1 is reported in Figures A.53 and A.54, while DGP 2 is reported in Figures A.55 and A.56.

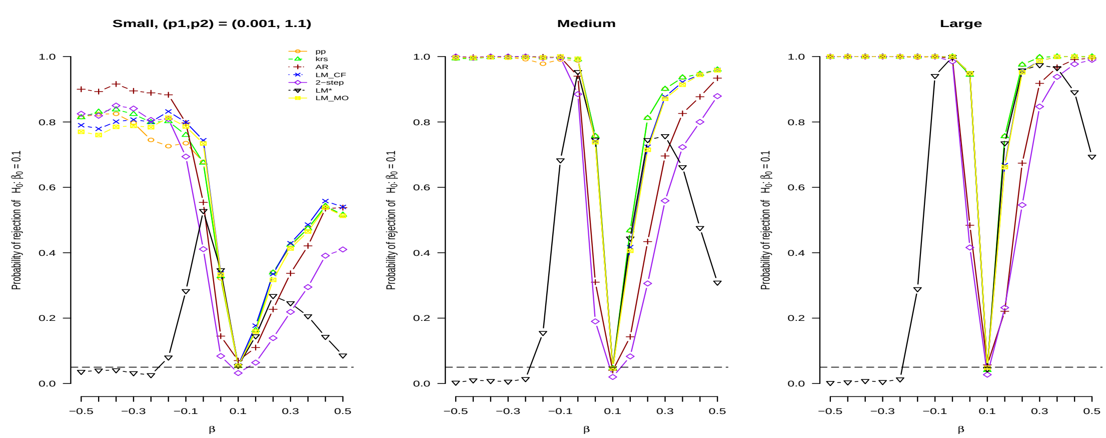

Figure A.37: Power Curve for DGP 1 given in (1.5.1) with $(p_1, p_2) = (0.001, 1.1)$ and Parameter Space $= \mathcal{B}$. The nominal size of 5% is represented by the horizontal dotted line
Note: The orange line with circle represents *pp* test; the green line with upward-pointing triangle represents *krs* test; the brown dash line with additive sign represents AR test given in (1.2.5); the blue dotted line with cross represents LM test with cross-fit variance; the purple dash line with diamond represents the 2-step test proposed by Mikusheva and Sun (2022) with overall 5% significance level; dark line with downward-pointing triangle represents LM^*; the yellow dash line with rectangle represents the LM test proposed by Matsushita and Otsu (2021).

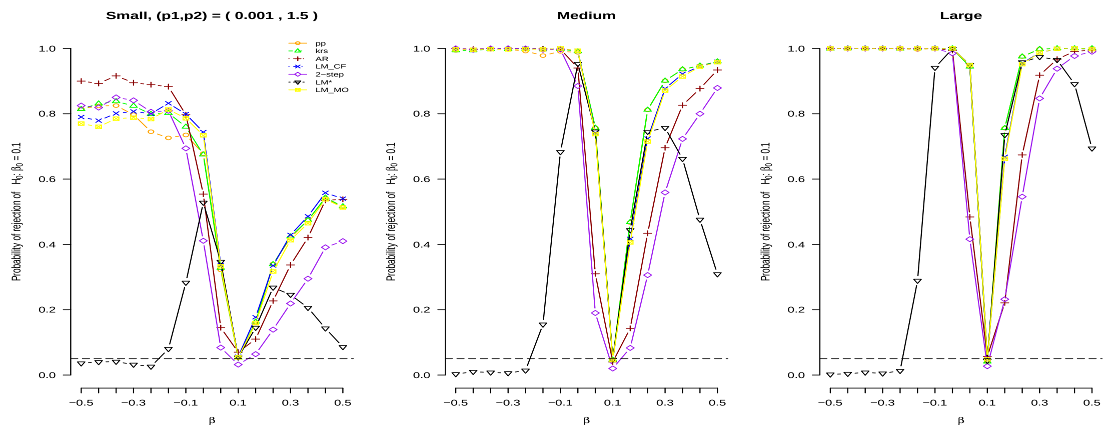

Figure A.38: Power Curve for DGP 1 with $(p_1, p_2) = (0.001, 1.5)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

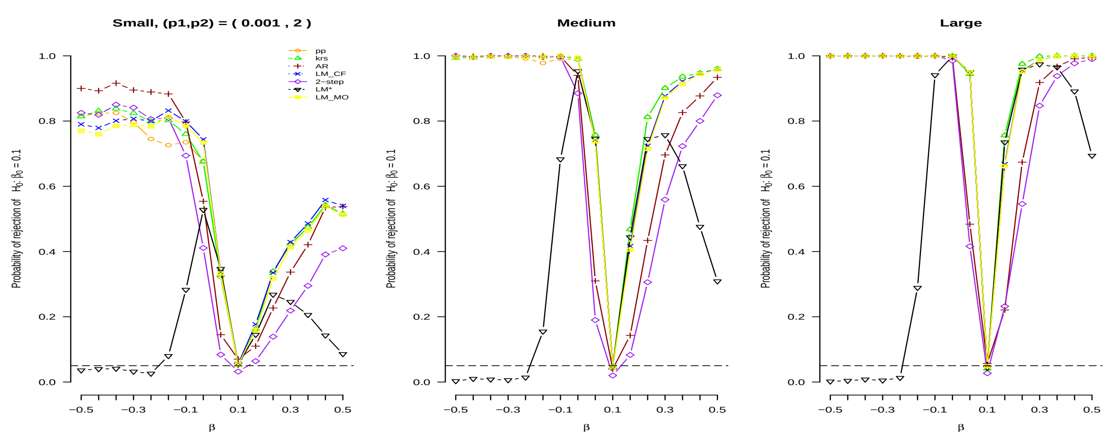

Figure A.39: Power Curve for DGP 1 with $(p_1, p_2) = (0.001, 2)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

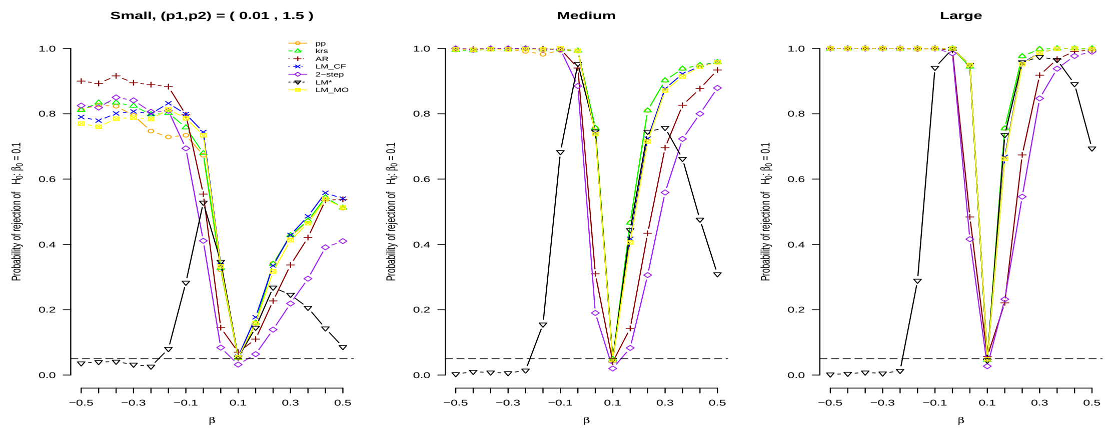

Figure A.40: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 1.5)$ and Parameter Space $= \mathcal{B}$
Note: The lines are explained under Figure A.37.

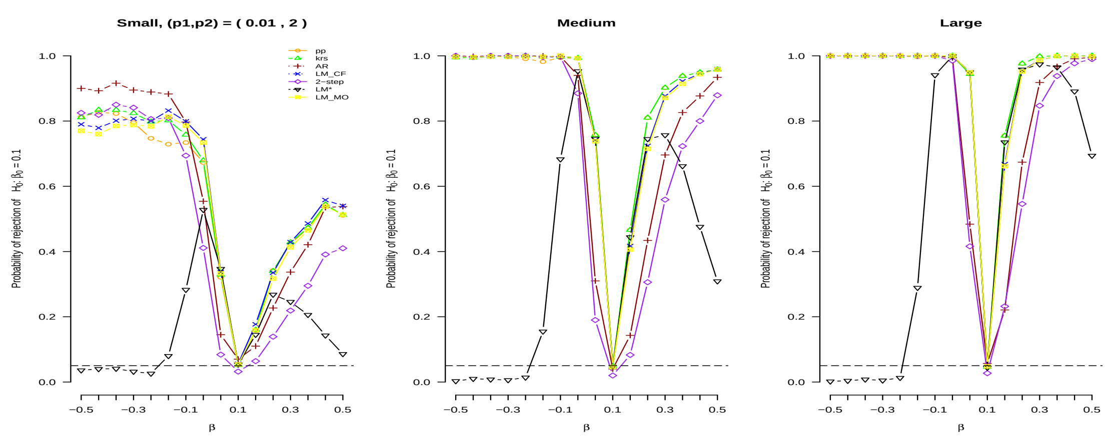

Figure A.41: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 2)$ and Parameter Space $= \mathcal{B}$
Note: The lines are explained under Figure A.37.

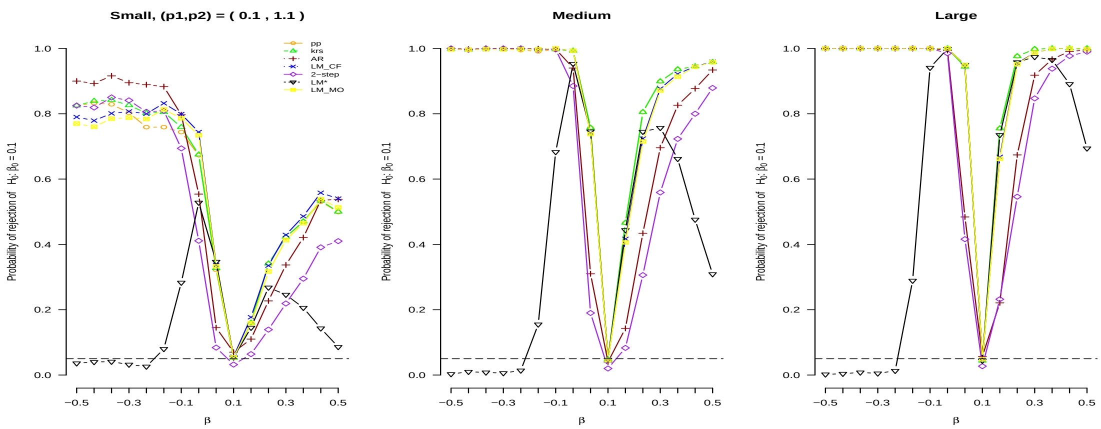

Figure A.42: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 1.1)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

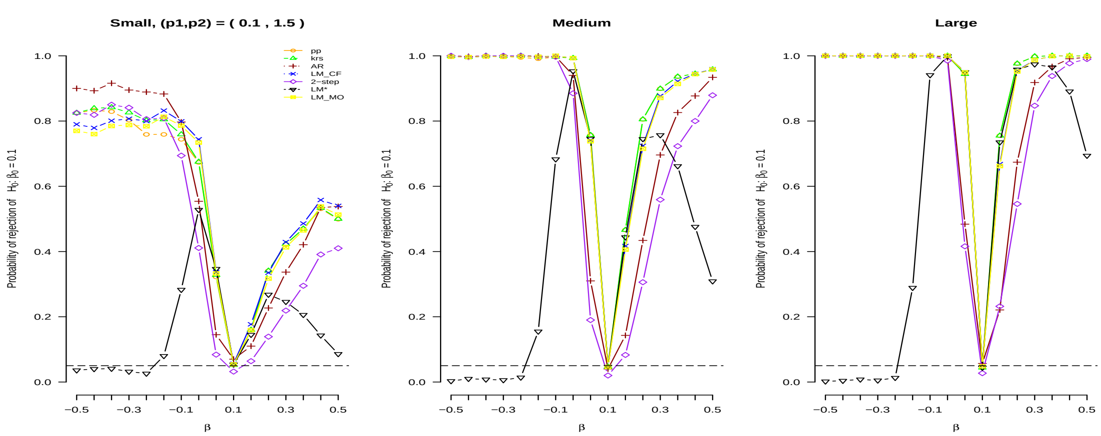

Figure A.43: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 1.5)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

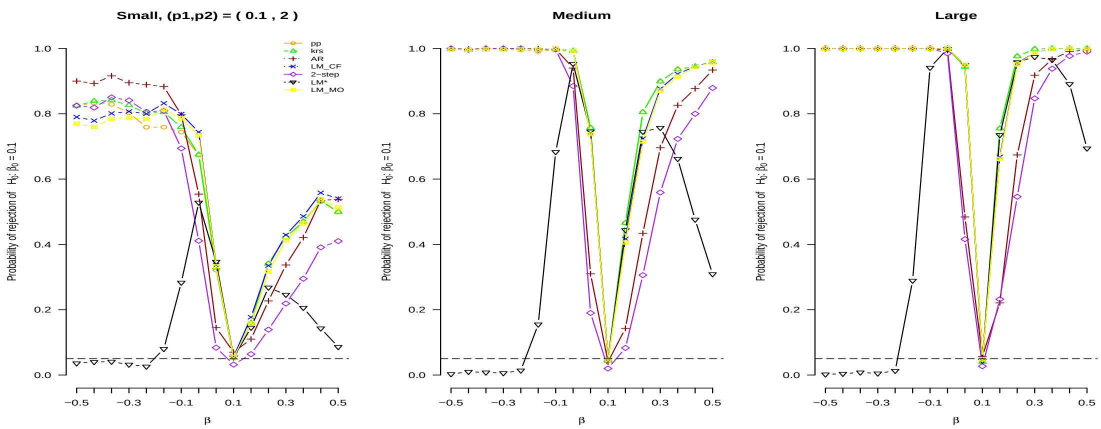

Figure A.44: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 2)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

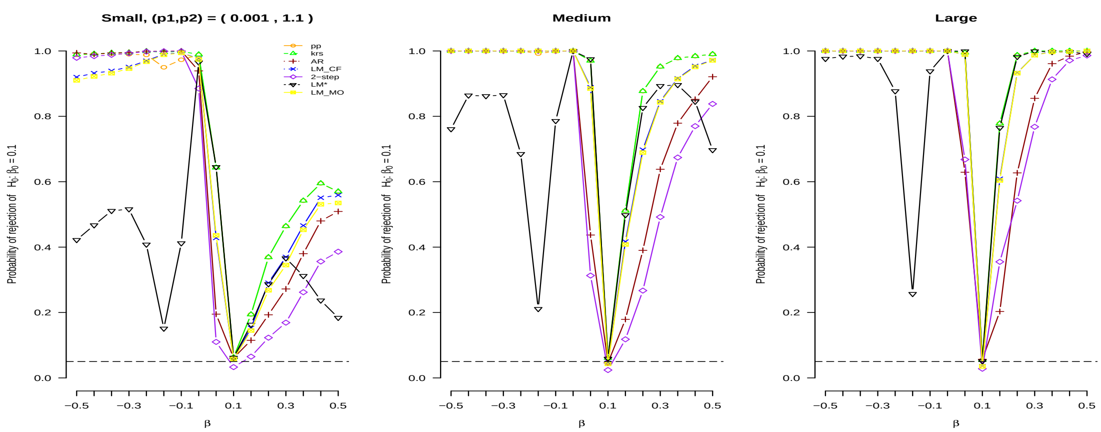

Figure A.45: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 1.1)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

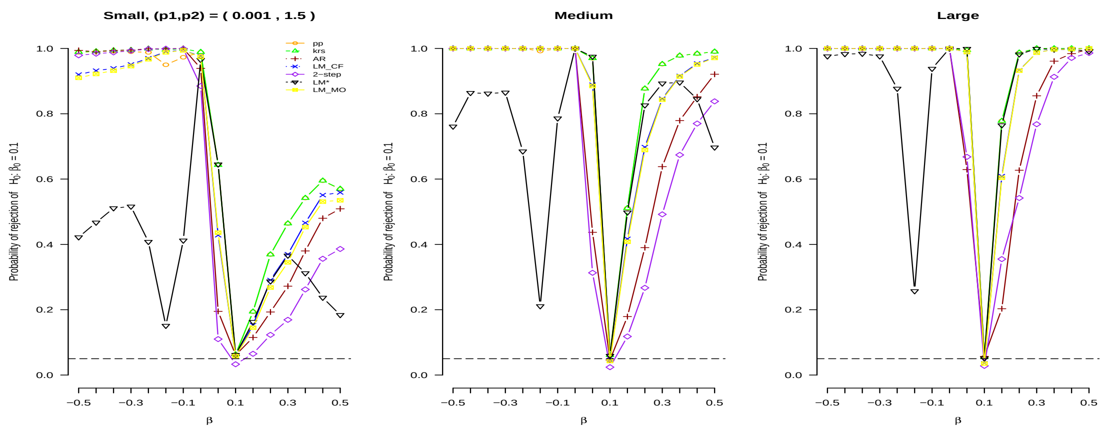

Figure A.46: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 1.5)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

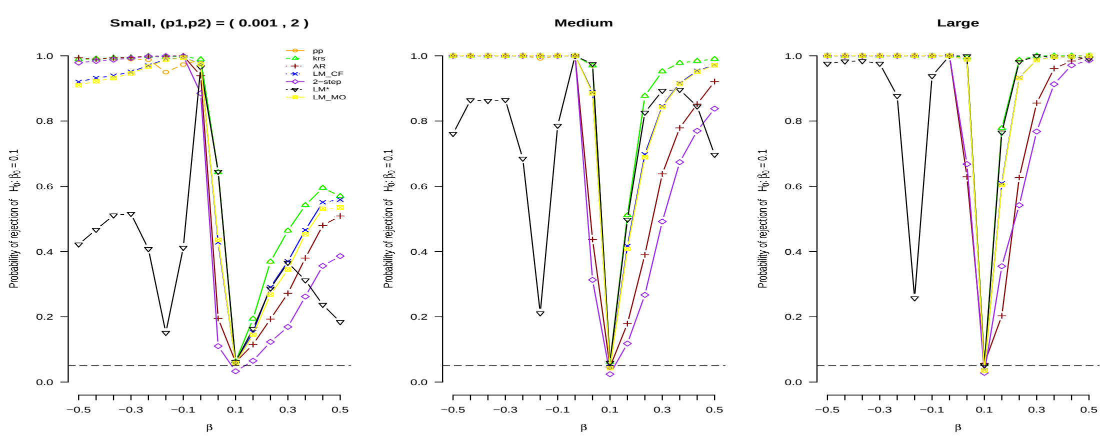

Figure A.47: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 2)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

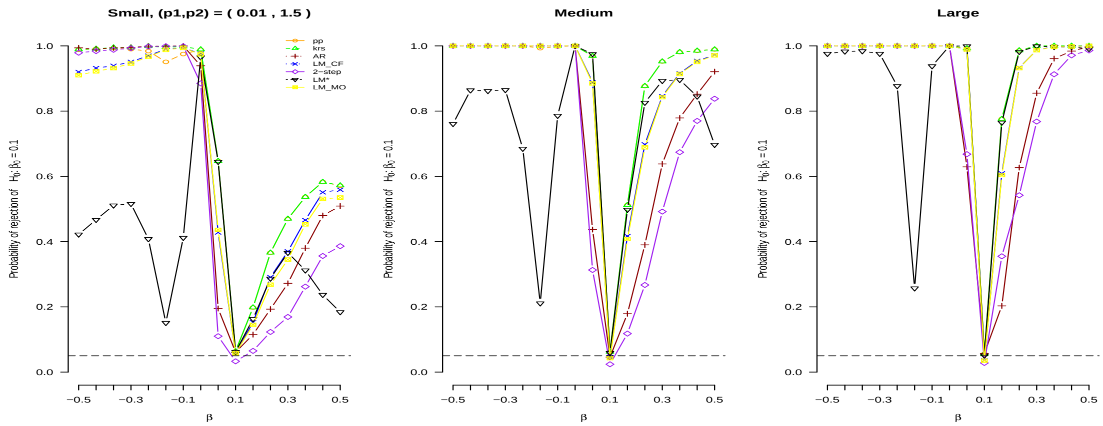

Figure A.48: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 1.5)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

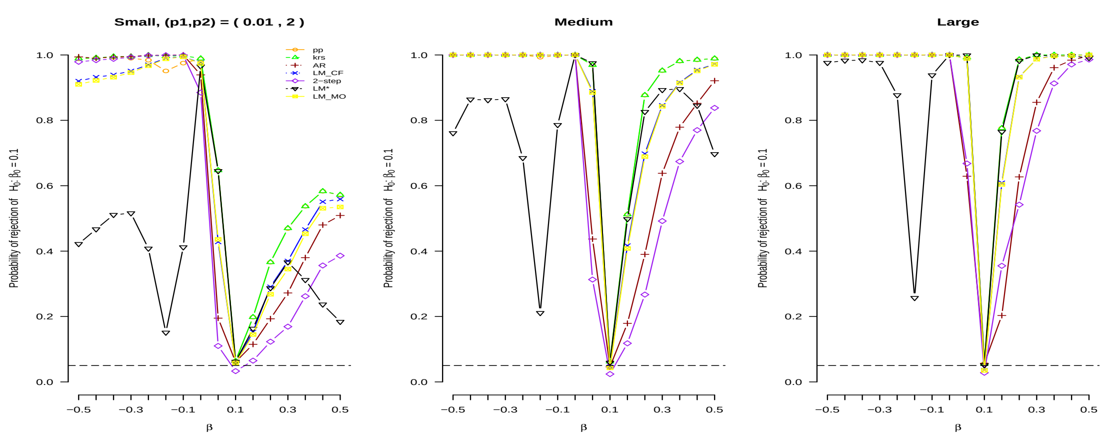

Figure A.49: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 2)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

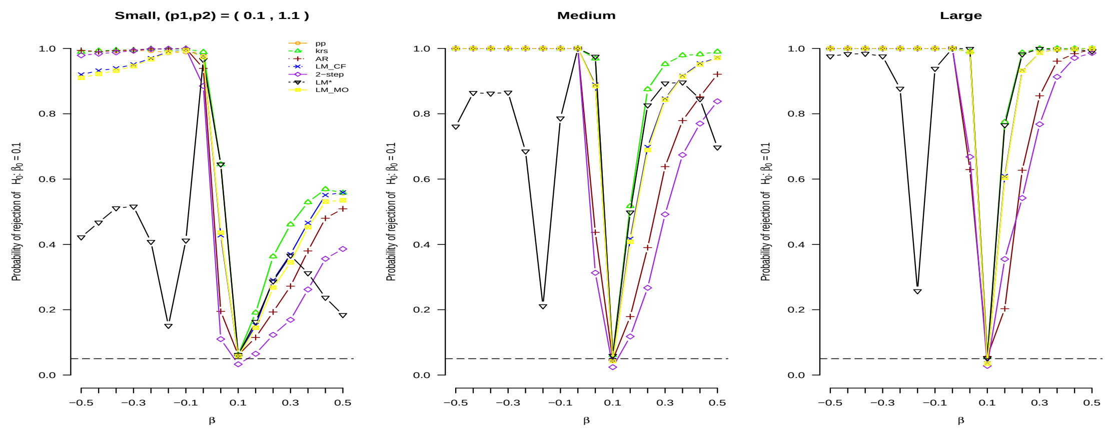

Figure A.50: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 1.1)$ and Parameter Space = $\mathcal{B}$

Note: The lines are explained under Figure A.37.

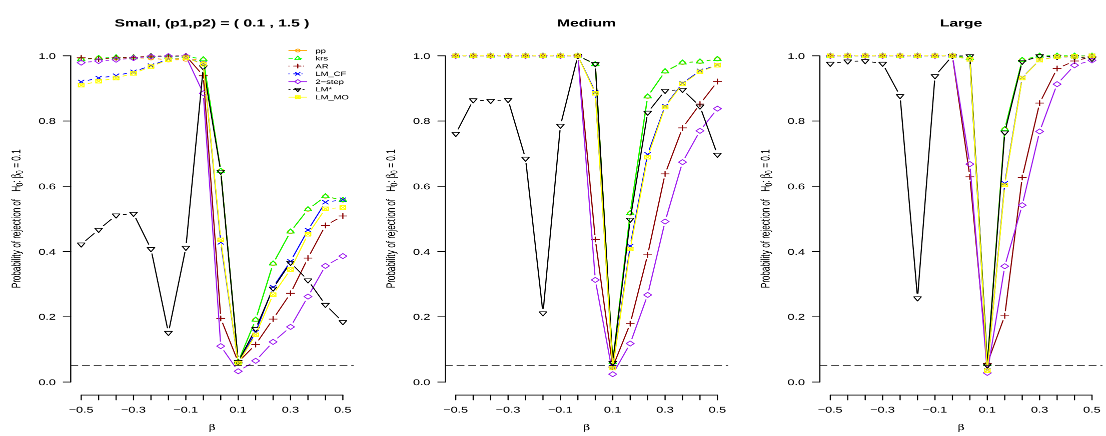

Figure A.51: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 1.5)$ and Parameter Space = $\mathcal{B}$

Note: The lines are explained under Figure A.37.

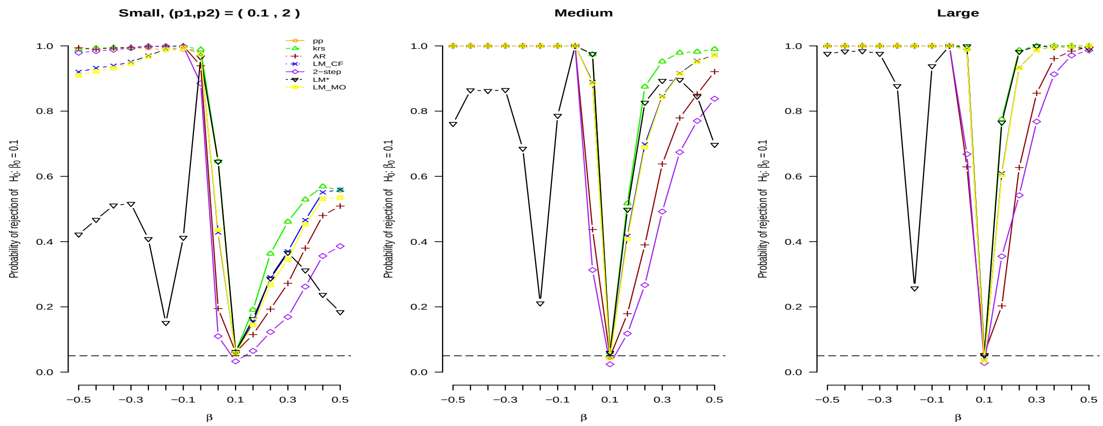

Figure A.52: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 2)$ and Parameter Space = $\mathcal{B}$
Note: The lines are explained under Figure A.37.

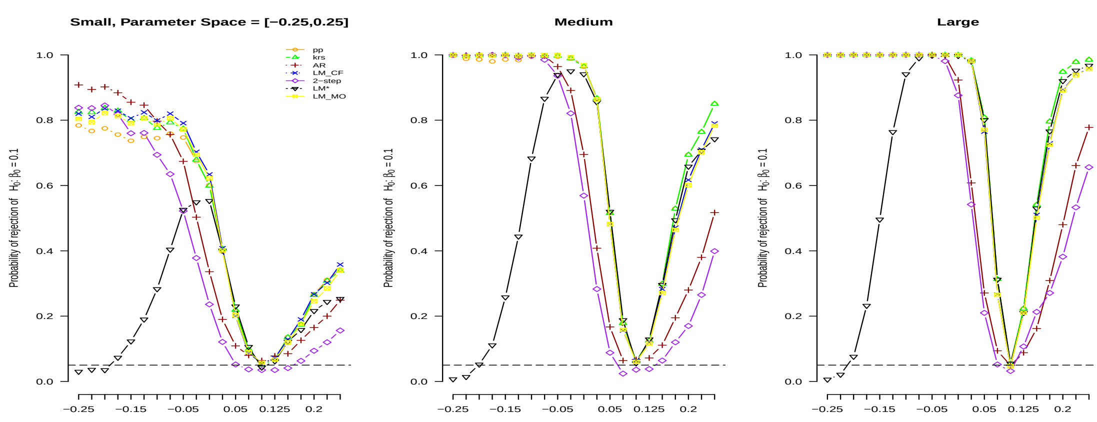

Figure A.53: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 1.1)$ and Parameter Space = $\mathcal{B}_2$
Note: The lines are explained under Figure A.37.

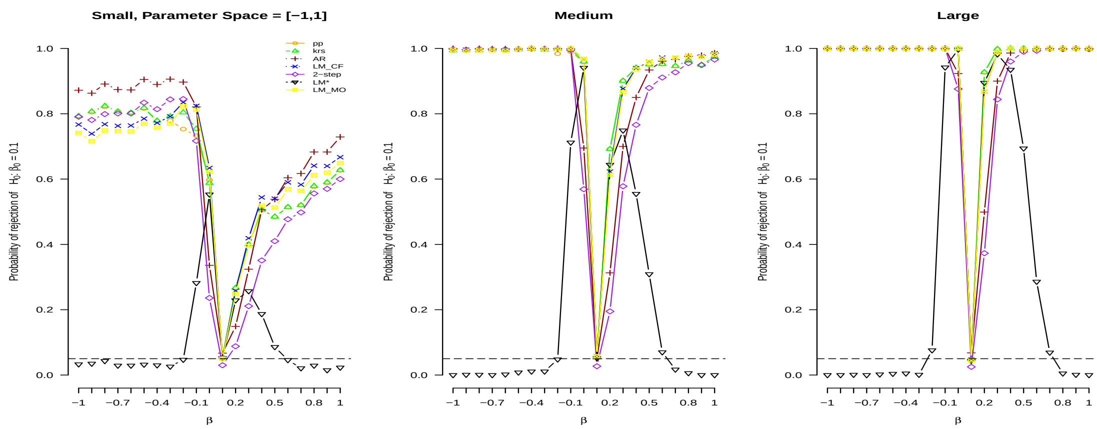

Figure A.54: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 1.1)$ and Parameter Space $= \mathcal{B}_3$
Note: The lines are explained under Figure A.37.

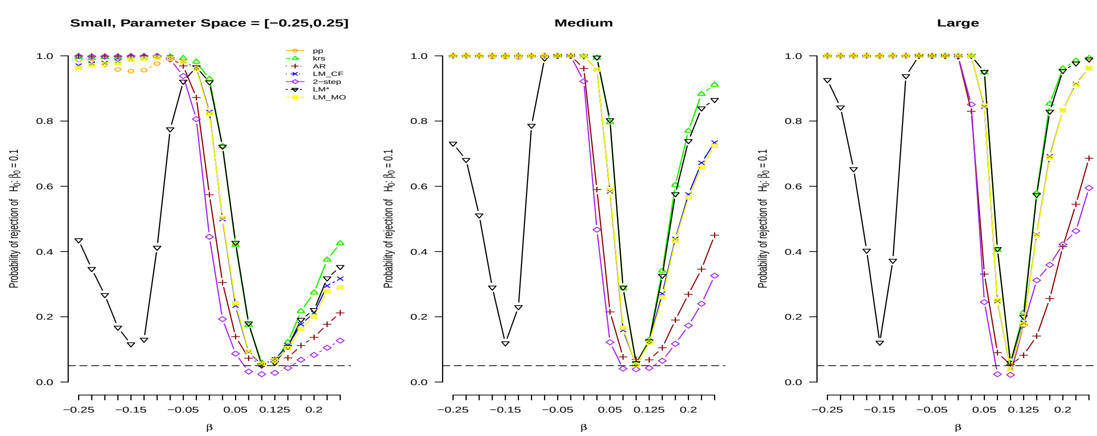

Figure A.55: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 1.1)$ and Parameter Space $= \mathcal{B}_2$
Note: The lines are explained under Figure A.37.

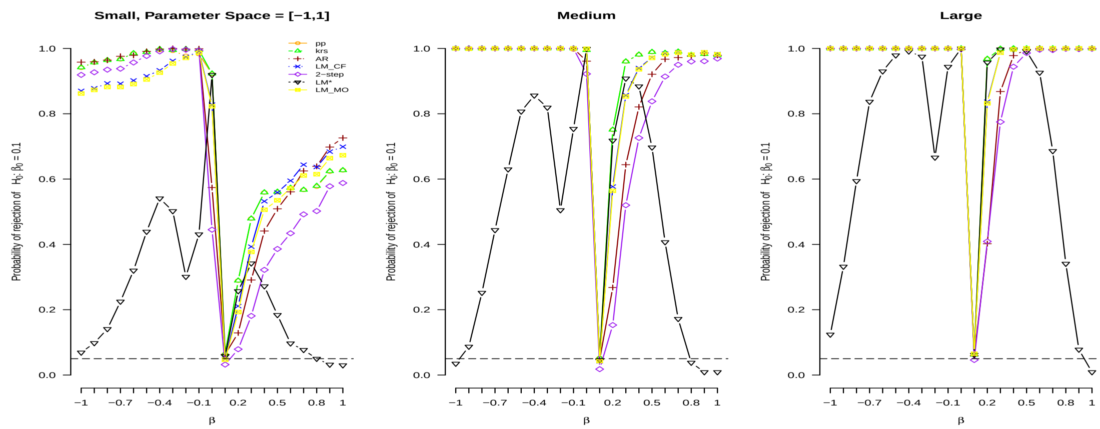

Figure A.56: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 1.1)$ and Parameter Space $= \mathcal{B}_3$
Note: The lines are explained under Figure A.37.

A.22 Additional Results for the Empirical Application

For the first set of robustness check, we ran 1001 equal-spaced grid-points from parameter space $\mathcal{B} = [-0.5, 0.5]$ (step size $= 0.001$) over the 9 different variations of (p_1, p_2), which we furnish in Table A.1. The first row is the specification used in the main text, $(p_1, p_2) = (0.01, 1.1)$. We do not include 'jackknife AR', 'jackknife LM', 'JIVE-t' and 'Two-step' since variations of (p_1, p_2) will not affect the result of those methods. We find that our results are similar to the main text.

(p_1, p_2)-values	pp with 180 IVs (5%)	krs with 180 IVs (5%)	pp with 1530 IVs (5%)	krs with 1530 IVs (5%)
(0.01,1.1)	[0.067,0.128]	[0.067,0.128]	[0.037,0.133]	[0.037,0.133]
(0.001,1.1)	[0.072,0.127]	[0.072,0.127]	[0.041,0.132]	[0.041,0.132]
(0.001,1.5)	[0.067,0.127]	[0.067,0.127]	[0.038,0.132]	[0.038,0.132]
(0.001,2)	[0.066,0.128]	[0.066,0.128]	[0.039,0.133]	[0.039,0.133]
(0.01,1.5)	[0.067,0.127]	[0.067,0.127]	[0.04,0.134]	[0.04,0.134]
(0.01,2)	[0.071,0.125]	[0.071,0.125]	[0.041,0.133]	[0.041,0.133]
(0.1,1.1)	[0.069,0.126]	[0.069,0.126]	[0.037,0.132]	[0.037,0.132]
(0.1,1.5)	[0.072,0.126]	[0.072,0.126]	[0.044,0.132]	[0.044,0.132]
(0.1,2)	[0.069,0.127]	[0.069,0.127]	[0.035,0.132]	[0.035,0.132]

Table A.1: Confidence Intervals under different values of (p_1, p_2) with Parameter Space $\mathcal{B}$

For the second set of robustness checks, we consider two different parameter spaces, namely $\mathcal{B}_2 = [-1, 1]$ and $\mathcal{B}_3 = [-0.25, 0.25]$. Both parameter spaces have 1001 equal-spaced grid-points, and we have retained the values $(p_1, p_2) = (0.01, 1.1)$ as in our main text. Table A.2 reports the results. Overall, these additional robustness checks show that the results reported in our main text are reliable and hold for different parameter spaces.

Parameter Space	pp with 180 IVs (5%)	krs with 180 IVs (5%)	pp with 1530 IVs (5%)	krs with 1530 IVs (5%)
$\mathcal{B}$	[0.067,0.128]	[0.067,0.128]	[0.037,0.133]	[0.037,0.133]
$\mathcal{B}_2$	[0.068,0.124]	[0.068,0.124]	[0.042,0.134]	[0.042,0.134]
$\mathcal{B}_3$	[0.07,0.1275]	[0.07,0.1275]	[0.037,0.1335]	[0.037,0.1335]

Table A.2: Confidence Intervals under $(p_1, p_2) = (0.01, 1.1)$ with varying Parameter Space $\mathcal{B}_2$ and $\mathcal{B}_3$

Appendix B

Technical Results for Chapter 2

B.1 Proofs For Main Text

B.1.1 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

$$\begin{aligned} Q_{e,e} &= Q_{\widetilde{e},\widetilde{e}} + 2Q_{\widetilde{e},P^W\widetilde{e}} + Q_{P^W\widetilde{e},P^W\widetilde{e}} \\ Q_{\mathcal{E},\mathcal{E}} &= Q_{\varepsilon,\varepsilon} + 2Q_{\varepsilon,P^W\varepsilon} + Q_{P^W\varepsilon,P^W\varepsilon} \end{aligned} \tag{B.1.1}$$

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

(i) $\mathbb{E}\vartheta_i = 0$

(ii) $\mathbb{E}[\vartheta_i^2] = \mathbb{E}[\widetilde{e}_i^2] = \mathbb{E}[\varepsilon_i^2]$

(iii) $\{(\vartheta_i\}_{i=1}^n$ is independent of $\{\widetilde{e}_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$

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

$$Q_{e,e} - Q_{\mathcal{E},\mathcal{E}} = Remainder_n + O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{B.1.2}$$

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 (B.1.2). In **step 2** we deal with the difference between the middle-term on the right-side of (B.1.1), 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 (B.1.2), 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 (B.1.2) (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

$$
\begin{aligned}
Q_{P^W\widetilde{\varepsilon},P^W\widetilde{\varepsilon}} - Q_{P^W\vartheta,P^W\vartheta} &= O_p(\frac{p_n d_W^2}{K^{1/2}}) \\
Q_{P^W\varepsilon,P^W\varepsilon} - Q_{P^W\vartheta,P^W\vartheta} &= O_p(\frac{p_n d_W^2}{K^{1/2}})
\end{aligned} \tag{B.1.3}
$$

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

$$
Q_{P^WU,P^WU} - Q_{P^W\widetilde{U},P^W\widetilde{U}} = O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{B.1.4}
$$

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

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

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

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

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

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

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 (B.1.4) follows. An application of (B.1.4) with $(U, \widetilde{U})$ replaced by $(\widetilde{e}, \vartheta)$ and (ε, ϑ) yields the first and second equation of (B.1.3) respectively.

Step 2: We show that

$$\begin{aligned}
2Q_{\widetilde{e},P^W\widetilde{e}} - 2Q_{\vartheta,P^W\vartheta} &= \mathcal{H}_n^{(1)} - \frac{2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}P_{ii}^W(\widetilde{e}_i\widetilde{e}_j - \vartheta_i\vartheta_j) = \mathcal{H}_n^{(1)} + O_p(\frac{p_n d_W^2}{K^{1/2}})\\
2Q_{\varepsilon,P^W\varepsilon} - 2Q_{\vartheta,P^W\vartheta} &= \mathcal{H}_n^{(2)} - \frac{2}{\sqrt{K}}\sum_{i\in[n]}P_{ii}P_{ii}^W(\varepsilon_i\varepsilon_j - \vartheta_i\vartheta_j) = \mathcal{H}_n^{(2)} + O_p(\frac{p_n d_W^2}{K^{1/2}}) \quad \text{(B.1.5)}
\end{aligned}$$

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 $P^W P = 0$,

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

which implies that

$$
Q_{P^W U,T} - Q_{P^W \widetilde{U},\widetilde{T}} = -\frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W U_j T_i + \frac{1}{\sqrt{K}} \sum_{i\in[n]} \sum_{j\neq i} P_{ii} P_{ij}^W \widetilde{U}_j \widetilde{T}_i + O_p(\frac{p_n d_W^2}{K^{1/2}}), \tag{B.1.6}
$$

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 (B.1.6) would yield the first equation of (B.1.5). The second equation of (B.1.5) 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

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

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

$$
F_n = \mathcal{F}_n' + O_p\left(\left[\frac{p_n^{1/2} + p_n^{3/2} (p_n^W)^{1/2} d_W}{K^{1/2}}\right]^{1/3}\right) \tag{B.1.7}
$$

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,

$$\left| f_n(x+y) - f_n(x) - y\partial f_n(x) - \frac{1}{2}y^2\partial^2 f_n(x) \right| \le \frac{|y|^3}{9\delta_n h_n^2} \tag{B.1.8}$$

and

$$1 - B(C_h)\mathbb{1}\{x \in A\} \le f_n(x) \le B(C_h) + (1 - B(C_h))\mathbb{1}\{x \in A^{3\delta_n}\}, \tag{B.1.9}$$

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,

$$\begin{aligned} &|\mathbb{E}f_n(F_n) - \mathbb{E}f_n(\mathcal{F}_n)| \\ &\le \sum_{i\in[n]} |\mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_i, \varepsilon_{i+1}, ..., \varepsilon_n)) - \mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{i-1}, \varepsilon_i, ..., \varepsilon_n))|, \end{aligned} \tag{B.1.10}$$

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

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

so that together with (B.1.8), letting $x = \lambda_{n-1}$, $y = \Delta_n$ and $\widetilde{\Delta}_n$, we have

$$\begin{aligned}
&|\mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_n)) - \mathbb{E}f_n(\mathcal{G}_n(\widetilde{e}_1, ..., \widetilde{e}_{n-1}, \varepsilon_n))| \\
&\leq |\mathbb{E}\partial f_n(\lambda_{n-1})(\widetilde{\Delta}_n - \Delta_n)| + \frac{1}{2}|\mathbb{E}\partial^2 f_n(\lambda_{n-1})(\widetilde{\Delta}_n^2 - \Delta_n^2)| + \frac{\mathbb{E}|\widetilde{\Delta}_n|^3 + \mathbb{E}|\Delta_n|^3}{9\delta_n h_n^2} \\
&= \frac{\mathbb{E}|\Delta_n|^3 + \mathbb{E}|\widetilde{\Delta}_n|^3}{9\delta_n h_n^2}.
\end{aligned} \tag{B.1.11}$$

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 $|a+b|^3 \leq 2(a^2+b^2)\cdot|a+b| \leq 8(|a|^3+|b|^3)$, we have by independence of the errors across i that

$$\mathbb{E}|\Delta_n|^3 \leq \frac{C}{K^{3/2}}\mathbb{E}\left|\sum_{i\in[n]}(P_{in} + P_{ii}P_{in}^W + P_{nn}P_{in}^W)\widetilde{e}_i\right|^3 \tag{B.1.12}$$

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

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i\in[n-1]}\theta_i\right|^3 \overset{(i)}{\leq} 8\mathbb{E}\left|\sum_{i\in[n-1]}\theta_i\xi_i\right|^3 \overset{(ii)}{\leq} 8\int_0^\infty t^2\mathbb{P}\left(\left|\sum_{i\in[n-1]}\theta_i\xi_i\right| > t\right)dt \\
&= 8\mathbb{E}\int_0^\infty t^2\mathbb{P}\left(\left|\sum_{i\in[n-1]}\theta_i\xi_i\right| > t\middle|\mathcal{I}_{n-1}\right)dt \overset{(iii)}{\leq} 16\mathbb{E}\int_0^\infty t^2 exp(-\frac{1}{2}\frac{t^2}{\sum_{i\in[n-1]}\theta_i^2})dt \\
&\overset{(iv)}{\leq} C\mathbb{E}\left(\sum_{i\in[n-1]}\theta_i^2\right)^{3/2} \overset{(v)}{\leq} C\left(\mathbb{E}(\sum_{i\in[n-1]}\theta_i^2)^2\right)^{3/4}
\end{aligned} \tag{B.1.13}$$

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 $\theta_i = P_{in}\widetilde{e}_i$,

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

so that

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

Similarly we can obtain

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

Hence, by (B.1.12) and (B.1.13), we have

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

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 (B.1.10), 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

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

Therefore, by (B.1.9) we have

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

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

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

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 $\frac{B(C_h)+\frac{C}{C_h^2}}{1-B(C_h)} \leq \tau$, 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 (B.1.7) 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

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

and note that by (B.1.3) and (B.1.5),

$$R_n - \widetilde{\mathcal{R}}_n = O_p(\frac{p_n d_W^2}{K^{1/2}}) \tag{B.1.14}$$

and

$$\mathcal{R}_n - \widetilde{\mathcal{R}}_n = O_p(\frac{p_n d_W^2}{K^{1/2}}). \tag{B.1.15}$$

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 (B.1.7) and (B.1.14); the last line follows from (B.1.15). 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

$$\begin{aligned}\frac{\sum_{i\in[n]} P_{ii}\widetilde{e}_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} &= O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right)\\ \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{K} - \frac{\sum_{i\in[n]} P_{ii}\vartheta_i^2}{K} &= O_p\left(\frac{p_n^{1/2}}{K^{1/2}}\right)\end{aligned} \tag{B.1.16}$$

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

$$J_n = \mathcal{J}_n' + O_p(\frac{p_n^{1/2}}{K}). \tag{B.1.17}$$

Putting everything together, we have

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

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

where (i) follows from (B.1.16) and (B.1.17). This completes the proof of the second part of Theorem 1.

B.1.2 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,

$$
\lim_{n_k\to\infty} \mathbb{P}_{\lambda_{n_k}}\left(\widehat{Q}(\beta_0) > C_{\alpha,df}(\widehat{\Phi}_1(\beta_0))\right) = \alpha. \tag{B.1.18}
$$

$$
\lim_{n_k\to\infty} \lim_{B\to\infty} \mathbb{P}_{\lambda_{n_k}}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) = \alpha \tag{B.1.19}
$$

Then (B.1.18) and (B.1.19) satisfy **Assumption B*** of Andrews, Cheng, and Guggenberger (2020b). 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 (B.1.18) and (B.1.19). We break the proof into two parts, part I and II, which deals with (B.1.18) and (B.1.19) respectively. For each part, we deal with fixed and diverging instruments separately. We drop the dependence on β_0 for notational simplicity.

Part I (analytical tests):

Fixed K case: Consider first the case when K is fixed. We can write the rejection criteria (2.2.8) as

$$
\widehat{Q}(\beta_0) > q_{1-\alpha}(F_{\widetilde{w}_n}) + (q_{1-\alpha}(F_{\widetilde{w}_n}) - 1)\left(\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2 + 1/df}} - 1\right) \tag{B.1.20}
$$

We denote $Q(\beta_0)$ as $Q_n(\beta_0)$ to reflect its relationship to the sample size n. Under the null, by Theorem B.4.1.1 and Lemma B.2.3, we know that for any sub-sequence n_j, there exists a further sub-sequence n_{j_k} such that

$$\widehat{Q}_{n_{j_k}}(\beta_0) \rightsquigarrow \sum_{i\in[K]} w_i^* \chi_{1,i}^2 =: \overline{\chi}_{w^*}^2 \tag{B.1.21}$$

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.2.3. 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 $w_n \to w^*$. This is because if for any sub-sequence we can show size-control for a further sub-sequence, then size-control holds for the entire sequence. Note that

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

$$(b) \quad \sum_{i\in[n]} P_{ii}^2\sigma_i^4 \leq \overline{C}^2 p_n K = o(1)$$

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

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

where (i) follows from our assumption of consistent estimator; (ii) from the second part of Theorem B.3.0.1; (iii) follows from (b); (iv) follows from Lemma B.2.1. 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,

$$\frac{\sqrt{\widehat{\Phi}_1}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2} = \frac{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}\cdot\frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}e_i^2}\cdot\frac{\sqrt{\widehat{\Phi}_1}}{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}$$

$$\overset{(e),(f)}{=} \frac{\sqrt{\frac{1}{K}\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]}P_{ii}\sigma_i^2}(1+o_p(1))(\sqrt{2}+o_p(1)) = \sqrt{2}\frac{\sqrt{\sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\sigma_i^2\sigma_j^2}}{\sum_{i\in[n]}P_{ii}\sigma_i^2}+o_p(1)$$

$$\overset{(a)}{=} \sqrt{2}||w_n||+o_p(1) = \sqrt{2}||w^*||+o_p(1), \tag{B.1.22}$$

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

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

as $\frac{1}{df} = o(1)$. Therefore,

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

so we can write (B.1.20) as

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

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

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

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

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

applying Theorem 1.3.4(vi) of the same reference yields

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

We have therefore shown that for fixed K, (B.1.18) is satisfied.

Diverging K: assume now that $K\to\infty$. By Theorem B.4.2.1 we have

$$\frac{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2}{\sqrt{\widehat{\Phi}_1}}\left(\widehat{Q}(\beta_0) - 1\right) = Q_{e,e} \rightsquigarrow \mathcal{N}(0,1) \tag{B.1.23}$$

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

$$Var(F_{\widetilde{w}_n} - 1 \mid \mathcal{I}) = \mathbb{E}\left(\sum_{i\in[K]}\widetilde{w}_{i,n}(\chi^2_{1,i} - 1)\right) = s_n^2. \tag{B.1.24}$$

Additionally, we have

$$\lim_{K\to\infty}\frac{C\max_i \widetilde{w}_{i,n}^2}{\sum_{i\in[n]}\widetilde{w}_{i,n}^2} = 0. \tag{B.1.25}$$

To see (B.1.25), note that $\max_i \widetilde{w}_{i,n} = o_p(1)$ by Lemma B.2.3. 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.2.3, 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 (B.1.25) we can obtain

$$\lim_{K \to \infty} \frac{1}{s_n^4} \sum_{i \in [K]} \mathbb{E}(\widetilde{w}_{i,n}(\chi_{1,i}^2 - 1))^4 \leq \lim_{K \to \infty} \frac{C \sum_{i \in [n]} \widetilde{w}_{i,n}^4}{s_n^4} \leq \lim_{K \to \infty} \frac{C \max_i \widetilde{w}_{i,n}^2 \sum_{i \in [n]} \widetilde{w}_{i,n}^2}{(\sum_{i \in [K]} \widetilde{w}_{i,n}^2)^2}$$
$$= \lim_{K \to \infty} \frac{C \max_i \widetilde{w}_{i,n}^2}{\sum_{i \in [K]} \widetilde{w}_{i,n}^2} = 0. \tag{B.1.26}$$

Since the Lyapunov condition (B.1.24) and (B.1.26) is satisfied, by the Lyapunov Central Limit Theorem, conditional on $\mathcal{I}$ we have

$$\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2 + 1/df}} \overset{(i)}{=} \frac{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2}}{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2} + 1/df} \frac{F_{\widetilde{w}_n} - 1}{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2}}$$
$$= (1 + o_p(1)) \frac{F_{\widetilde{w}_n} - 1}{\sqrt{2 \sum_{i \in [K]} \widetilde{w}_{i,n}^2}} \rightsquigarrow \mathcal{N}(0, 1). \tag{B.1.27}$$

where (i) follows from observing that $1 = \sum_{i \in [K]} \widetilde{w}_{i,n} \leq ||\widetilde{w}_n||_F \sqrt{K}$ by cauchy-schwartz inequality, so that $\frac{1}{||\widetilde{w}_n||_F df} \leq \frac{\sqrt{K}}{df} = o(1)$ by assumption. Since the distributional convergence in (B.1.27) holds for any sequence $\widetilde{w}_{i,n}$, then it must hold unconditionally by Lemma B.2.4.

Hence, asymptotically, by (B.1.23) we have exact α-level size control whenever

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

We can rearrange this rejection criteria as

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

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 (B.1.18). The proof of part I is complete.

Part II (bootstrap tests):

We can first establish that for any fixed sample size n, conditioning on data, for any $z\in\mathbb{R}$,

$$\frac{\sum_{\ell\in[B]} 1\left\{\widehat{J}^{BS,\ell}\le z\right\}}{B} \xrightarrow{\widehat{p}} \widehat{P}_{\mathcal{L}}\left(\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_i\eta_j}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}}\le z\middle|\widehat{P}\right) \tag{B.1.28}$$

as $B\to\infty$, where we drop the dependence of $\widehat{J}^{BS,\ell}$ on $(e(\beta_0),\mathcal{L},\widehat{\Phi}_1(\beta_0))$ for notational simplicity; $\xrightarrow{\widehat{p}}$ and $\mathbb{P}_{\mathcal{L}}(\cdot|\widehat{P})$ means convergence in probability and probability measure under the law $\mathcal{L}$ conditioning on the data, respectively; $\Phi_1^{BS,n}(\beta_0) := \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2(\beta_0)e_j^2(\beta_0)$; random variables $\{\eta_i\}_{i\in[n]} \overset{d}{\sim} \mathcal{L}$. First observe that $\widehat{\Phi}_1^{BS,\ell}(\beta_0)\xrightarrow{\hat{p}}\Phi_1^{BS,n}(\beta_0)$ by $\mathbb{E}(\eta_i|e_i)=0$, $Var(\eta_i|e_i)=e_i^2$, and the assumption that $\widehat{\Phi}_1(\beta_0)$ satisfies (2.2.12). Second, observe that $\left\{\widehat{J}^{BS,\ell}\right\}_{\ell\in[B]}$ are i.i.d., so that (B.1.28) follows from the law of large numbers.

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. Since $\widetilde{w}_n\xrightarrow{p} w^*$ implies some sub-sequence converges almost-surely, we can assume $\widetilde{w}_n\xrightarrow{a.s.} w^*$ over the full

Note that

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) = \frac{\sum_{i\in[n]} P_{ii}e_i^2(\widehat{Q}_s(\beta_0)-1)}{\sqrt{K\widehat{\Phi}_1}} = \frac{\widehat{Q}(\beta_0)-1}{\sqrt{2}||w^*||} + o_p(1) \rightsquigarrow \sum_{i\in[K]} \frac{w_i^*}{\sqrt{2}||w^*||}(\chi^2_{1,i}-1) \tag{B.1.29}$$

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 (B.1.21). Next, we will show that $\mathbb{P}$-almost surely, for any $z \in \mathbb{R}$,

$$\widehat{P}_{\mathcal{L}}\left(\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_i\eta_j}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}} \leq z \middle| \widehat{P}\right) \to \mathbb{P}\left(\sum_{i\in[K]} \frac{w_i^*}{\sqrt{2}||w^*||}(\chi^2_{1,i}-1) \leq z\right) \tag{B.1.30}$$

as $n \to \infty$. Conditional on data, $\mathbb{P}_{\lambda_n}$-almost surely we have

$$\begin{aligned}
\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_i\eta_j}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}} &= \frac{\sum_{i\in[n]} P_{ii}\eta_i^2}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}}\left(\frac{\eta' P\eta}{\sum_{i\in[n]} P_{ii}\eta_i^2} - 1\right) \\
&\overset{(i)}{=} \frac{\sum_{i\in[n]} P_{ii}\eta_i^2}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}}\left(\sum_{i\in[K]} \widetilde{w}_{i,n}^{BS}\chi^2_{1,i} - 1\right) + o_{\widehat{p}}(1) \\
&\overset{(ii)}{=} \sum_{i\in[K]} \frac{\widetilde{w}_{i,n}^{BS}}{\sqrt{2}||w^*||}(\chi^2_{1,i}-1) + o_{\widehat{p}}(1) \\
&\overset{(iii)}{=} \sum_{i\in[K]} \frac{\widetilde{w}_{i,n}}{\sqrt{2}||w^*||}(\chi^2_{1,i}-1) + o_{\widehat{p}}(1) \\
&= \sum_{i\in[K]} \frac{w_{i,n}^*}{\sqrt{2}||w^*||}(\chi^2_{1,i}-1) + o_{\widehat{p}}(1)
\end{aligned}$$

where (i) follows from Theorem 1 adapted to conditioning on data[1], $\widetilde{w}_n^{BS} := (\widetilde{w}_{1,n}^{BS}, ..., \widetilde{w}_{K,n}^{BS})'$

[1] Although Theorem 1 requires the fourth moment to be bounded from above, we

are the eigenvalues of $\frac{(Z'\Lambda_\eta Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda_\eta Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\eta_i^2}$ and $\Lambda_\eta := diag(\eta_1^2, ..., \eta_n^2)$; (ii) follows from

$$\frac{\sum_{i\in[n]} P_{ii}\eta_i^2}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}} = \sqrt{2}||\widetilde{w}_n|| + o_{\widehat{p}}(1) = \sqrt{2}||w^*|| + o_{\widehat{p}}(1),$$

which is analogous to (B.1.22); (iii) follows from Lemma B.2.3 adapted to the conditioned data, where there exists for every sub-sequence n_j a further sub-sequence n_{j_k} such that under the null

$$\max_{i\in[K]}(\widetilde{w}_{i,n_{j_k}}^{BS} - \widetilde{w}_{i,n_{j_k}})^2 = o_{\widehat{p}}(1),$$

and we can assume without loss of generality that this holds under the full sequence. This proves (B.1.30). Finally, by Vaart (1998)[Lemma 21.2], (B.1.30) implies

$$q_{1-\alpha}\left(\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_i\eta_j}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}}\right) \xrightarrow{\widehat{p}} q_{1-\alpha}\left(\sum_{i\in[K]}\frac{w_{i,n}^*}{\sqrt{2}||w^*||}(\chi_{1,i}^2 - 1)\right),$$

so that conditioning on data and combining with (B.1.28) yields, WPA1 (with respect to law $\mathcal{L}$)

$$\lim_{n\to\infty}\lim_{B\to\infty} C_{\alpha,df_{BS}}^B(\widehat{\Phi}_1(\beta_0), \mathcal{L}) = q_{1-\alpha}\left(\sum_{i\in[K]}\frac{w_{i,n}^*}{\sqrt{2}||w^*||}(\chi_{1,i}^2 - 1)\right),$$

noting that $df_{BS} = o(1)$. The preceding equation holds $\mathbb{P}_{\lambda_n}$-almost surely, so that by bounded convergence theorem,

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}_{\lambda_n}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_{\alpha,df_{BS}}^B(\widehat{\Phi}_1(\beta_0), \mathcal{L})\right) = \alpha$$

---

note that $sup_{i\in\mathbb{N}}\ e_i^4 < \infty$ with probability greater than $1-\varepsilon$ for any $\varepsilon > 0$. Therefore, following the arguments later on, we can prove a version of (B.1.19), that is $\alpha(1-\varepsilon) \leq \liminf_{n_k\to\infty}\lim_{B\to\infty}\mathbb{P}_{\lambda_{n_k}}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_{\alpha,df_{BS}}^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) \leq \limsup_{n_k\to\infty}\lim_{B\to\infty}\mathbb{P}_{\lambda_{n_k}}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C_{\alpha,df_{BS}}^B(\widehat{\Phi}_1^{BS}(\beta_0), \mathcal{L})\right) \leq \alpha(1-\varepsilon)+\varepsilon$. since $\varepsilon > 0$ was arbitrary, we have (B.1.19) itself. Hence we can assume without loss of generality that $sup_{i\in\mathbb{N}}\ e_i^4 < \infty$ with probability one.

This completes the proof of the fixed K case.

Diverging K: assume now that $K \to \infty$. Then by Chao et al. (2012)[Lemma A2],

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) \rightsquigarrow \mathcal{N}(0,1) \tag{B.1.31}$$

Furthermore, by applying Chao et al. (2012)[Lemma A2] conditioned on data, we have[2]

$$\widehat{P}_{\mathcal{L}}\left(\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}\eta_i\eta_j}{\sqrt{K\Phi_1^{BS,n}(\beta_0)}} \leq z \,\middle|\, \widehat{P}\right) \xrightarrow{\widehat{p}} \mathbb{P}\left(\mathcal{N}(0,1) \leq z\right), \tag{B.1.32}$$

so that combining with (B.1.31), (B.1.28), using bounded convergence theorem and $df_{BS} = o(1)$ yields

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}_{\lambda_n}\left(\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) > C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L})\right) = \alpha$$

This completes the proof for the diverging K case.

B.1.3 Proof of Theorem 3

We first prove the first part of the statment. Note that (B.1.27) holds for any sequence of $\Delta_n \to \Delta^\dagger$ not necessarily zero, i.e.

$$\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2 + 1/df}} \rightsquigarrow \mathcal{N}(0,1) \tag{B.1.33}$$

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

$$\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) > \sqrt{\widehat{\Phi}_1(\beta_0)} \cdot q_{1-\alpha}\left(\frac{F_{\widetilde{w}_n} - 1}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2 + 1/df}}\right) \tag{B.1.34}$$

[2]Note that the following equation holds true for any sequence of $\Delta_n \to \Delta^\dagger$ not necessarily zero, as long as $\widehat{\Phi}_1(\Delta_n) \xrightarrow{p} \Phi_1(\Delta^\dagger)$, where we have rewritten the dependence of $\widehat{\Phi}_1(\cdot)$ on Δ_n instead of β_0, so that β_0 is seen as "moving" in this case.

By (2.2.12), noting that

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

the estimator $\widehat{\Phi}_1(\beta_0)=O_p(1)$. Therefore the right-hand-side of (B.1.34) is an $O_p(1)$ term. The left-hand-side of (B.1.34) diverges to infinity for $\mathcal{C}\to\infty$ and fixed $\Delta\neq 0$ by Theorem B.4.2.2. The result of the first statement thus follow. For the second part of the statement, note that (B.1.32) holds even under the alternative. Therefore, by (B.1.28), (B.1.32) and $df_{BS}=o(1)$, we have that $\mathbb{P}$-almost surely,

$$\lim_{n\to\infty}\lim_{B\to\infty}C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L})\xrightarrow{\widehat{p}}q_{1-\alpha}(\mathcal{N}(0,1)).$$

Combining with the fact that

$$\widehat{J}(\beta_0,\widehat{\Phi}_1(\beta_0))=\frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}}\sum_{i\in[n]}P_{ii}e_i^2(\beta_0)\left(\widehat{Q}(\beta_0)-1\right)\xrightarrow{p}\infty$$

by Theorem B.4.2.2 yields the second statement.

B.1.4 Proof of Theorem 4

By Theorem B.4.2.2,

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

Therefore, by (B.1.33), for fixed Δ and any estimator $\widehat{\Phi}_1(\beta_0)\xrightarrow{p}\Phi_1(\beta_0)$.

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

$$= 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\Delta^2\mathcal{C}}{\sqrt{\widehat{\Phi}_1(\beta_0)}}\right)$$

$$= 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\Delta^2\mathcal{C}}{\sqrt{\Phi_1(\beta_0)}}\right)$$

Noting that $\Delta = \widetilde{\Delta}$ and $\mathcal{C} = \widetilde{\mathcal{C}}$ completes the first part of the proof. For the second part of the proof, it only remains to show that, $\mathbb{P}$-almost surely,

$$\lim_{n\to\infty}\lim_{B\to\infty} C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L}) \xrightarrow{\hat{p}} q_{1-\alpha}\left(\mathcal{N}\left(\frac{\Delta^2\mathcal{C}}{\sqrt{\Phi_1(\beta_0)}}, 1\right)\right).$$

But this follows directly from (B.1.28), (B.1.32) and $df_{BS} = o(1)$. Finally, we show that

$$\widehat{\Phi}_1^{standard}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0), \tag{B.1.35}$$

$$\widehat{\Phi}_1^{cf}(\beta_0) \xrightarrow{p} \Phi_1(\beta_0). \tag{B.1.36}$$

in order to complete the last part of the proof. Recall from section 2.2.5 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 (2.2.12) we have (B.1.35). Furthermore, by $\Pi' M \Pi \leq \frac{\Pi'\Pi}{K} \to 0$, (B.1.36) follows from Mikusheva and Sun (2022)[Theorem 3].

B.1.5 Proof of Theorem 5

Note that $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$ by (2.2.12) and $\Delta \to 0$. Furthermore, $\frac{\Delta^2\mathcal{C}}{\sqrt{\widehat{\Phi}_1(\beta_0)}} = \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}} + o(1) = \frac{\widetilde{\Delta}^2\widetilde{\mathcal{C}}}{\sqrt{\Phi_1(\beta_0)}}$, so that by Theorem B.4.2.2 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 (B.1.33) we have

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

This proves the first part of the statement. For the second part of the statement, it only remains to show that, $\mathbb{P}$-almost surely,

$$
\lim_{n\to\infty}\lim_{B\to\infty} C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L}) \xrightarrow{\widehat{p}} q_{1-\alpha}\left(\mathcal{N}\left(\frac{\Delta^2\mathcal{C}}{\sqrt{\Phi_1(\beta_0)}},1\right)\right),
$$

which follows directly from (B.1.28), (B.1.32) and $df_{BS} = o(1)$.

B.1.6 Proof of Lemma 2.4.1

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

$$
\begin{aligned}
&(a)\quad ||w_n||_F^2 \cdot (\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2 = \sum_{i,j\in[n]} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) \\
&(b)\quad \sum_{i\in[n]} P_{ii}^2\sigma_i^4(\beta_0) \le Cp_nK = o(1) \\
&(c)\quad \widehat{\Phi}_1(\beta_0) = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + \mathcal{D}(\Delta)\ \text{ by assumption of (2.2.12)}
\end{aligned}
$$

Hence

$$
\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} \overset{(i)}{=} \frac{\sqrt{\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + O_p(1)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + O_p(1)} + o_p(1)
$$

$$\overset{(a),(b)}{=} \sqrt{2}||w_n||_F + O_p(1) \leq \sqrt{2}||D_{w_n} + \Lambda_H||_F + \sqrt{2}||\Lambda_H||_F + O_p(1)$$
$$\overset{(ii)}{=} \sqrt{2}||D_{w_n} + \Lambda_H||_F + O_p(1)$$

where (i) follows from (c) and Lemma B.2.1; Λ_H is defined in Lemma B.2.3 and $D_{w_n} := diag(w_{1,n}, ..., w_{K,n})$; (ii) follows from $||\Lambda_H||_F^2 = ||\Omega_H(\beta_0)||_F^2 = \frac{\Delta^4 \sum_{i,j\in[n]} P_{ij}^2 \Pi_i^2 \Pi_j^2}{\sum_{i\in[K]} P_{ii}\sigma_i^2(\beta_0)} \leq \frac{\Delta^4 CK}{\underline{C}K} \leq C$. Furthermore, we have by Lemma B.2.3

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

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

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2 + 1/df}} = \frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}}{\sqrt{2||\widetilde{w}_n||_F^2 + 1/df}} \leq \frac{\sqrt{2}||D_n + \Lambda_H||_F + O_p(1)}{\sqrt{2||\widetilde{w}_n||_F^2 + 1/df}}$$
$$= \frac{\sqrt{2}||D_n + \Lambda_H||_F + O_p(1)}{\sqrt{2}||\Lambda_H + D_n||_F + o_p(1)} \overset{p}{\to} 1 + O_p(1) = O_p(1)$$

which completes the proof.

B.1.7 Proof of Lemma 2.4.2

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

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

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

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

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

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

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

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

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

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

B.1.8 Proof of Theorem 6

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

$$n^{-1/2}((Z'\widetilde{e})', (Z'\Delta_n\widetilde{v})')' \rightsquigarrow (I_K, I_K)\mathcal{N}\left(0, \Sigma(\Delta^\dagger)\right) \tag{B.1.37}$$

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

$$\begin{aligned}
&e(\beta_0)'Pe(\beta_0) \\
&= (n^{-1/2}Z'\widetilde{e} + \Delta_n n^{-1/2}Z'\widetilde{v} + \Delta_n n^{-1/2}Z'\Pi)'\left(\frac{Z'Z}{n}\right)^{-1}(n^{-1/2}Z'\widetilde{e} + \Delta_n n^{-1/2}Z'\widetilde{v} + \Delta_n n^{-1/2}Z'\Pi) \\
&\rightsquigarrow ((I_K, I_K)\mathcal{N}(0, \Sigma(\Delta^\dagger)) + \Delta^\dagger \mu_K)' Q_{ZZ}^{-1}((I_K, I_K)\mathcal{N}(0, \Sigma(\Delta^\dagger)) + \Delta^\dagger \mu_K) \qquad \text{(B.1.38)}
\end{aligned}$$

To show (B.1.38), note that by assumption 8 we have

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

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

$$\left\{\mathbb{E}\left\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F^2 1\{||(Z_i\widetilde{e}_i, \Delta_n Z_i\widetilde{v}_i)||_F \geq \eta\sqrt{n}\}\right\}\right\}^2$$

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

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 6 and 8; (iii) follows from Markov-inequality. We can then apply the Lindeberg-Feller Central-Limit-Theorem to obtain (B.1.38). Furthermore, note that

$$\left(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\right)^{-1} \geq C(1+\Delta^{\dagger}+\Delta^{\dagger 2})^{-1} + o_p(1) \tag{B.1.39}$$

for some $C > 0$. To see (B.1.39), 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.2.1; (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 (B.1.39) is shown. 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 (B.1.38) diverges to ∞, as $d_n^{-1} \to \infty$. By (B.1.38) and (B.1.39) we have

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

Furthermore, by lemma 2.4.2 we know that $q_{1-\alpha}(F_{\widetilde{w}_n}) = O_p(1)$; by lemma 2.4.1 and (B.1.20), we have

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

This completes the proof for the first part for the statement of Theorem 6. For the second part, WPA1,

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) = \frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \sum_{i\in[n]} P_{ii} e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) \to \infty \qquad \text{(B.1.40)}$$

by $\widehat{Q}(\beta_0) \to \infty$ and WPA1,

$$\frac{\sum_{i\in[n]} P_{ii} e_i^2(\beta_0)}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \overset{(i)}{\geq} \frac{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \overset{(ii)}{\geq} \frac{\underline{C}\sum_{i\in[n]} P_{ii}}{\sqrt{KC_1}} \geq \frac{\underline{C}\sqrt{K}}{\sqrt{C_1}} > 0$$

where (i) follows from Lemma B.2.1; (ii) follows from assumption 6 and $\widehat{\Phi}_1(\beta_0) \leq C_1$ for some $C_1 > 0$ WPA1. Furthermore, by (B.1.28) and (B.1.32), $\mathbb{P}$-almost surely,

$$\lim_{n\to\infty}\lim_{B\to\infty} C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0),\mathcal{L}) \xrightarrow{\widehat{p}} q_{1-\alpha}\left(\mathcal{N}\left(\frac{\Delta^2\mathcal{C}}{\sqrt{\Phi_1(\beta_0)}},1\right)\right),$$

so that combining with (B.1.40) yields the second statement of Theorem 6.

B.1.9 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 (B.1.37), $\Delta_n n^{-1/2}Z'\widetilde{v} = o_p(1)$, whence

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

Furthermore, by Lemma B.2.1, $p_n\frac{\Pi'\Pi}{K} = O(1)$ and $\Delta \to 0$, we have

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

where β is the true parameter. Therefore we have

$$\begin{aligned} &\widehat{Q}(\beta_0) = \frac{(n^{-1/2}Z'\widetilde{e} + \widetilde{\Delta}\widetilde{\mu})'\left(\frac{Z'Z}{n}\right)^{-1}(n^{-1/2}Z'\widetilde{e} + \widetilde{\Delta}\widetilde{\mu})}{\sum_{i\in[n]} P_{ii}\widetilde{\sigma}_i^2} + o_p(1) \\ &= \left((Z'\Lambda_0 Z)^{-1/2}Z'\widetilde{e} + (n^{-1}Z'\Lambda_0 Z)^{-1/2}\widetilde{\Delta}\widetilde{\mu}\right)'\Omega(\beta)\left((Z'\Lambda_0 Z)^{-1/2}Z'\widetilde{e} + (n^{-1}Z'\Lambda_0 Z)^{-1/2}\widetilde{\Delta}\widetilde{\mu}\right) + o_p(1) \\ &\rightsquigarrow \left(\mathcal{N}(0,I_K) + \Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)'\Omega^*(\beta)\left(\mathcal{N}(0,I_K) + \Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) = \mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)'\Omega^*(\beta)\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) \end{aligned} \tag{B.1.41}$$

where $\Omega(\beta)$ is defined in (2.2.6), $\Lambda_0 := diag(\Lambda_{0,1}, ..., \Lambda_{0,n})$ and the convergence follows from (B.1.37) and $\Omega^*(\beta) := \lim_{n\to\infty} \Omega(\beta)$. Next, we deal with the critical value. If we show that

$$\widetilde{w}_n \xrightarrow{p} w^* \quad and \quad \frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]} \widetilde{w}_{i,n}^2} + 1/df} \xrightarrow{p} 1, \tag{B.1.42}$$

then by (B.1.41) and (B.1.20) we can obtain

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

which completes the first part of the proof. Note that by Lemma B.2.1, 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 2.4.1 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.2.3 we have that

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

Finally,

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

where the last equality follows by recalling from (B.1.27) that

$$\frac{||\widetilde{w}_n||}{||\widetilde{w}_n|| + 1/df} = 1 + o_p(1).$$

Therefore, together with the assumption that $w_n \to w^*$ (which holds as $\lim_{n\to\infty} \Omega(\beta_0) \to \Omega^*(\beta_0)$), (B.1.42) is shown. This proves the first statement of the theorem. To prove the second part of the theorem, note that $\widehat{\Phi}_1(\beta_0) \xrightarrow{p} \Phi_1(\beta_0)$ by (2.2.12). Furthermore, observe that by (B.1.41) and Lemma B.2.1,

$$\widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) = \frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}} \sum_{i\in[n]} P_{ii} e_i^2(\beta_0) \left(\widehat{Q}(\beta_0) - 1\right) = \frac{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}{\sqrt{K\Phi_1(\beta_0)}} \left(\widehat{Q}(\beta_0) - 1\right) + o_p(1)$$

$$= \frac{1}{\sqrt{2}||w_n||} \left(\widehat{Q}(\beta_0) - 1\right) + o_p(1) \rightsquigarrow \frac{\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)' \Omega^*(\beta)\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) - 1}{\sqrt{2}||w^*||} \tag{B.1.43}$$

where the last equality follows from the proof of Lemma 2.4.1. Finally, by (B.1.28) and (B.1.30) we have $\mathbb{P}$-almost surely,

$$\lim_{n\to\infty} \lim_{B\to\infty} C^B_{\alpha, df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L}) \xrightarrow{\widehat{p}} q_{1-\alpha}\left(\sum_{i\in[K]} \frac{w_i^*}{\sqrt{2}||w^*||}(\chi^2_{1,i} - 1)\right),$$

so that combing with (B.1.43) yields the second statement of Theorem 7.

B.1.10 Proof of Corollary 2.4.1

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

B.1.11 Proof of Theorem 8:

We prove the first statement of Theorem 8 first. 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

$$\widehat{Q}(\beta_0) \overset{(i)}{=} \frac{(\mathbb{A}_n + \widetilde{\mu})'(\frac{Z'Z}{n})^{-1}(\mathbb{A}_n + \widetilde{\mu})}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \widetilde{\Delta}^2 \sum_{i\in[n]} P_{ii}\Pi_i^2 + o_p(1)}$$

$$\overset{(ii)}{=} (\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})' \frac{Z'\Lambda(\beta_0)P\Lambda(\beta_0)Z}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0) + \widetilde{\Delta}^2 \sum_{i\in[n]} P_{ii}\Pi_i^2} (\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1)$$
$$= (1+\mathcal{Y}_n)^{-1}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})' \frac{Z'\Lambda(\beta_0)P\Lambda(\beta_0)Z}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} (\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1)$$
$$\overset{(iii)}{=} (1+\mathcal{Y}_n)^{-1}(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu})'\Omega(\beta_0)(\mathbb{V}_n^{-1/2}\mathbb{A}_n + \mathbb{V}_n^{-1/2}\widetilde{\mu}) + o_p(1)$$
$$\overset{(iv)}{\rightsquigarrow} (1+\mathcal{Y}_n)^{-1} \left(\mathcal{N}(0, I_K) + \Sigma(\widetilde{\Delta})\widetilde{\mu}\right)' \Omega^*(\beta_0) \left(\mathcal{N}(0, I_K) + \Sigma(\widetilde{\Delta})\widetilde{\mu}\right) \tag{B.1.44}$$

where (i) follows from Lemma B.2.1; (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 (2.2.6); (iv) follows from (B.1.37). To deal with the critical-value, note that by Lemma B.2.3 we have that

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

so that

$$||\widetilde{w}_n||_F^2 = ||w_n + \Lambda^H||_F^2 + o_p(1) = ||w_n||_F^2 + \frac{\widetilde{\Delta}^2 \sum_{i\in[n]} P_{ii}\Pi_i^2}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} + 2w_n'\Lambda^H + o_p(1)$$
$$= ||w_n||_F^2 + \mathcal{Y}_n + 2w_n'\Lambda^H + o_p(1) \tag{B.1.45}$$

where $\Lambda^H = (\lambda_{1,n}^H, ..., \lambda_{K,n}^H)$ is defined in Lemma B.2.3. Furthermore,

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

where (i) follows from Lemma B.2.1 and (c) in the proof of Lemma 2.4.1; (ii) follows from (b) in the proof of Lemma 2.4.1; (iii) follows from (a) in the proof of Lemma 2.4.1. Therefore we have

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2+1/df}} \overset{(i)}{=} \frac{||w_n||_F}{(1+\mathcal{Y}_n)\left(\sqrt{||w_n||_F^2+\mathcal{Y}_n+2w_n'\Lambda^H}+1/df\right)}+o_p(1)$$
$$\overset{(ii)}{=} \frac{||w^*||_F}{\sqrt{||w^*||_F^2+2w^{*\prime}\Lambda_H}}+o_p(1). \tag{B.1.46}$$

where (i) follows from (B.1.45); (ii) follows from $||w_n-w^*||_F=o(1)$, $1/df=o(1)$, and

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

(iii) follows from $\sigma_i^2(\beta_0)\geq \underline{C}>0$ by assumption 6, (iv) follows from $\Pi'\Pi=O(1)$ and $\frac{p_n}{K}=o(1)$ by assumption 6. Furthermore, we can show that

$$\Lambda_H=(n^{-1}Z'Z)^{-1/2}\frac{Z'H_nZ}{n}(n^{-1}Z'Z)^{-1/2}\to 0, \tag{B.1.47}$$

which follows from

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

where (i) follows from $\sup_i||Z_i||_F<\infty$ by assumption 8. Therefore, combining (B.1.46) and (B.1.47) yields

$$\frac{\frac{\sqrt{\widehat{\Phi}_1(\beta_0)}}{\frac{1}{\sqrt{K}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}}{\sqrt{2\sum_{i\in[K]}\widetilde{w}_{i,n}^2+1/df}} \overset{p}{\to} 1 \tag{B.1.48}$$

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, (B.1.44), (B.1.48) and (2.2.9), we have the first statement of Theorem 8. For the second statement, note that we just showed

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

Therefore by (B.1.44) and $\mathcal{Y}_n = o(1)$, we have

$$\begin{aligned} \widehat{J}(\beta_0, \widehat{\Phi}_1(\beta_0)) &= \frac{1}{\sqrt{K\widehat{\Phi}_1(\beta_0)}}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\left(\widehat{Q}(\beta_0) - 1\right) = \frac{1}{\sqrt{2}||w^*||}\left(\widehat{Q}(\beta_0) - 1\right) + o_p(1) \\ &\rightsquigarrow \frac{\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right)'\Omega^*(\beta_0)\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right) - 1}{\sqrt{2}||w^*||} \end{aligned} \tag{B.1.49}$$

Next, by (B.1.28) and (B.1.30) we have $\mathbb{P}$-almost surely,

$$\lim_{n\to\infty}\lim_{B\to\infty} C^B_{\alpha,df_{BS}}(\widehat{\Phi}_1(\beta_0), \mathcal{L}) \xrightarrow{\widehat{p}} q_{1-\alpha}\left(\sum_{i\in[K]}\frac{w_i^*}{\sqrt{2}||w^*||}(\chi^2_{1,i} - 1)\right),$$

so that combining with (B.1.49) yields the second statement of Theorem 8. Finally, the last part of the theorem is shown in exactly the same way as the last part of the proof of Theorem 4.

B.1.12 Proof of Corollary 2.4.2

Repeat the proof of corollary 2.4.1 and replace $\mathbb{M}_i$ by $\overline{\mathbb{M}}_i$ for each i

B.2 Proofs for Technical Lemmas

Lemma B.2.1. *Under Assumption 5 and 6, 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.2.1:

To begin, recall

$$\sigma_i^2(\beta_0) = \widetilde{\sigma}_i^{\,2} + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i \tag{B.2.1}$$

Furthermore,

$$\begin{aligned} e_i^2(\beta_0) &= (e_i + \Delta X_i)^2 = ((M_i^W)'\widetilde{e} + \Delta\Pi_i + \Delta v_i)^2 \\ &= ((M_i^W)'\widetilde{e})^2 + 2\Delta\Pi_i(M_i^W)'\widetilde{e} + 2\Delta v_i(M_i^W)'\widetilde{e} + \Delta^2\Pi_i^2 + 2\Delta^2\Pi_i v_i + \Delta^2 v_i^2 \\ &= A_{i,1} + 2\Delta A_{i,2} + 2\Delta A_{i,3} + \Delta^2 A_{i,4} + 2\Delta^2 A_{i,5} + \Delta^2 A_{i,6} \end{aligned} \tag{B.2.2}$$

We will show that

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,1} - \widetilde{\sigma}_i^2) = O_p\left(\sqrt{\frac{p_n}{K}} + \sqrt{p_n^W}\right) \tag{B.2.3}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,2} = O_p(\sqrt{\frac{p_n}{K}}), \tag{B.2.4}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,3} - \widetilde{\gamma}_i) = O_p(\sqrt{\frac{p_n}{K}} + \sqrt{p_n^W}), \tag{B.2.5}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,4} = O_p(\Delta^2 p_n \frac{\Pi'\Pi}{K}) \tag{B.2.6}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}A_{i,5} = O_p(\sqrt{\frac{p_n}{K} + p_n^W}). \quad \text{and} \tag{B.2.7}$$

$$\frac{1}{K}\sum_{i\in[n]} P_{ii}(A_{i,6}-\widetilde{\varsigma}_i^2) = O_p(\sqrt{\frac{p_n}{K}}+\sqrt{p_n^W}) \tag{B.2.8}$$

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

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

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

$$\begin{aligned}
\mathbb{E}(B_2)^2 &\le \frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\in[n]} P_{ii}P_{i'i'}\sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{i'j'}^W \mathbb{E}(\widetilde{e}_i\widetilde{e}_j\widetilde{e}_{i'}\widetilde{e}_{j'}) \\
&= \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2 \sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{ij'}^W \mathbb{E}(\widetilde{e}_i^2\widetilde{e}_j\widetilde{e}_{j'}) + \frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\neq i} P_{ii}P_{i'i'}\sum_{j\in[n]}\sum_{j'\in[n]} P_{ij}^W P_{i'j'}^W \mathbb{E}(\widetilde{e}_i\widetilde{e}_j\widetilde{e}_{i'}\widetilde{e}_{j'}) \\
&\le \frac{C}{K^2}\sum_{i\in[n]} P_{ii}^2\sum_{j\in[n]}(P_{ij}^W)^2 + \frac{C}{K^2}\sum_{i\in[n]}\sum_{i'\neq i} P_{ii}P_{i'i'}(P_{ii}^W P_{i'i'}^W + (P_{ii'}^W)^2) \\
&\le C p_n^W
\end{aligned} \tag{B.2.9}$$

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

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

so that putting it all together yields (B.2.3). 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,

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

and

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

we obtain (B.2.4). For (B.2.5), 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

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

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

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

and

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

as in (B.2.9); the argument for $B_7 = O_p(\sqrt{p_n^W})$ is analogous to B_6. Furthermore, by

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

we have (B.2.5). Next, (B.2.6) is obvious. For (B.2.7), noting that $v_i v_{i'} = \widetilde{v}_i\widetilde{v}_{i'} + \sum_{\ell\in[n]} P_{i\ell}^W\widetilde{v}_\ell\sum_{\ell\in[n]} P_{i'\ell}^W\widetilde{v}_\ell -$

$\sum_{\ell\in[n]} P^W_{i'\ell}\widetilde{v}_\ell\widetilde{v}_i - \sum_{\ell\in[n]} P^W_{i\ell}\widetilde{v}_\ell\widetilde{v}_{i'}$, we have

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

Finally we deal with (B.2.8). 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

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

Furthermore, similar to (B.2.9) we have

$$\mathbb{E}B_{10}^2 \leq Cp_n^W = O(p_n^W)$$

and

$$\mathbb{E}B_{11} \leq \frac{C}{K}\sum_{i\in[n]} P_{ii}\sum_{j\in[n]}(P^W_{ij})^2 \leq Cp_n^W = O(p_n^W)$$

This completes the proof of (B.2.8). By the assumption of $\frac{p_n}{K} = o(1)$ and $p_n^W = o(1)$, each term from (B.2.3)-(B.2.8) except (B.2.6) is $o_p(1)$. Hence Lemma B.2.1 is shown.

□

Lemma B.2.2. *Suppose Assumption 5 and 6 holds. Then for fixed Δ not necessarily zero,*

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

Proof of Lemma B.2.2:

Step 1: We first show that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2\sigma_j^2(\beta_0)=\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2\sigma_j^2(\beta_0)+o_p(1) \tag{B.2.10}$$

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,

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

where the inequality is from

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

Second,

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

Third, note that

$$C_{i,3} = \sum_{j\neq i}(P_{ij}^W)^2\widetilde{e}_j^2 + \sum_{j\neq i}\sum_{k\neq i,j} P_{ij}^W P_{kj}^W \widetilde{e}_j \widetilde{e}_k \tag{B.2.11}$$

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 $|ab| \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$

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

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

Sixth,

$$\frac{1}{K^2} B_{2,3} \overset{(B.2.11)}{=} \frac{1}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0) \sigma_{j'}^2(\beta_0) \mathbb{E} \left((\sum_{k \neq i} P_{ik}^W \widetilde{e}_k \widetilde{e}_i)(\sum_{k \neq i'} (P_{i'k}^W)^2 \widetilde{e}_k^2 \right)$$
$$+ \frac{1}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0) \sigma_{j'}^2(\beta_0) \mathbb{E} \left((\sum_{\ell \neq i} P_{i\ell}^W \widetilde{e}_\ell \widetilde{e}_i)(\sum_{k \neq i'} \sum_{k' \neq i', k} P_{i'k}^W P_{k'k}^W \widetilde{e}_k \widetilde{e}_{k'}) \right)$$
$$\leq \frac{C}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0) \sigma_{j'}^2(\beta_0) P_{ii}^W$$
$$+ \frac{C}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i} P_{ij}^2 P_{i'j'}^2 \sigma_j^2(\beta_0) \sigma_{j'}^2(\beta_0) \sum_{\ell \neq i} (|P_{i\ell}^W P_{i'\ell}^W P_{i\ell}^W| + (P_{i\ell}^W)^2 |P_{ii'}^W|)$$
$$\leq \frac{C p_n^W}{K^2} \sum_{i,i' \in [n]} \sum_{j \neq i} \sum_{j' \neq i} P_{ij}^2 P_{i'j'}^2 \leq C p_n^W = o(1).$$

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 $2|ab| \leq a^2 + b^2$,

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

The proof of (B.2.10) 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

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 v_i^2 \sigma_j^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \widetilde{\varsigma}_i^2 \sigma_j^2(\beta_0) + o_p(1) \tag{B.2.12}$$

and

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 v_i e_i \sigma_j^2(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \widetilde{\gamma}_i \sigma_j^2(\beta_0) + o_p(1) \tag{B.2.13}$$

Finally, we will show that

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_j^2(\beta_0)\Pi_i e_i = o_p(1) \tag{B.2.14}$$

and

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \sigma_j^2(\beta_0)\Pi_i v_i = o_p(1) \tag{B.2.15}$$

We will only show (B.2.14) since (B.2.15) 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

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

where (i) follows from

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

and

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

where (ii) follows from Cauchy-Schwartz inequality. Therefore, by Markov inequality we have (B.2.14). Combining (B.2.10)-(B.2.15) yields Lemma B.2.2

□

Lemma B.2.3. *Suppose Assumption 5, 6 and 7 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 $\Omega_H(\beta_0) := \frac{U' H_n U}{\sum_{i \in [n]} P_{ii} \sigma_i^2(\beta_0)}$, $H_n := diag(T_{1,n}, ..., T_{n,n})$ 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.2.3:

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

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(e_i^2(\beta_0)-\sigma_i^2(\beta_0)-T_i)(e_j^2(\beta_0)-\sigma_j^2(\beta_0)-T_j)=o_p(1) \tag{B.2.16}$$

where the last equality follows from

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

where (i) follows from noting that by repeating the proof of Theorem B.3.0.1, we can 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.2.2, 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

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

where the last equality follows from (B.2.16) and

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

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

$$||U'\hat{\Lambda}U - U'\Lambda U - U'H_nU||_F = o_p(1) \tag{B.2.17}$$

Then we have

$$\begin{aligned}
&||\widehat{\Omega}_{s,n}(\beta_0)-\Omega_{s,n}(\beta_0)-\Omega_H(\beta_0)||_F^2 = \left|\left|\frac{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)U'\Lambda U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\cdot\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_F^2\\
&= \frac{1/K^2}{\left(\frac{1}{K}\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)\cdot\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\right)^2}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0))U'\Lambda U\right|\right|_F^2\\
&\overset{(i)}{=} \frac{1/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4+o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-H_n)U - \sum_{i\in[n]} P_{ii}e_i^2(\beta_0)U'\Lambda U\right|\right|_F^2\\
&\overset{(ii)}{\le} \frac{2/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4+o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\cdot U'(\hat{\Lambda}-\Lambda-H_n)U\right|\right|_F^2\\
&+ \frac{2/K^2}{(\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^4+o_p(1)}\left|\left|\sum_{i\in[n]} P_{ii}(e_i^2(\beta_0)-\sigma_i^2(\beta_0))\cdot U'\Lambda U\right|\right|_F^2
\end{aligned}$$

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

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

where (i) follows from Lemma B.2.1; (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) $\left|\left|\frac{1}{K}\sum_{i\in[n]} P_{ii}\{e_i^2(\beta_0) - \sigma_i^2(\beta_0)\}\right|\right|_F^2 = ||o_p(1)||_F^2 = o_p(1)$ by Lemma $B.2.1$

(c) $\left|\left|U'(\hat{\Lambda} - \Lambda - H_n)U\right|\right|_F^2 = o_p(1)$ by $(B.2.17)$

(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

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

We will show that[3]

$$\max_i \widetilde{w}_{i,n} = o_p(K^{-1/2}) = o_p(1) \tag{B.2.18}$$

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

$$\begin{aligned}\max_i w_{i,n} = ||\Omega_s(\beta_0)||_S &= \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||U'\Lambda U||_S \leq \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||U||_S^2||\Lambda||_S \\ &\overset{(i)}{=} \frac{1}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}||\Lambda||_S = \frac{\max_i \sigma_i^2(\beta_0)}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \overset{(ii)}{\leq} \frac{C/K}{\frac{1}{K}\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} = o(K^{-1/2})\end{aligned} \tag{B.2.19}$$

where (i) follows by $U'U = I_K$; (ii) follows from expression (B.2.1). Furthermore, we have

$$\max_i \lambda^H_{i,n} = ||\Omega_H(\beta_0)||_S = \frac{||U'H_nU||_S}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)} \leq \frac{||H_n||_S}{K\underline{C}} = \frac{\max_i \Delta^2\Pi_i^2}{K\underline{C}} \leq \frac{C}{K} = o(K^{-1/2}) \tag{B.2.20}$$

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 $\Omega_H(\beta_0) = Q_3'\Lambda_HQ_3$, 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

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

[3]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

$$\le \left(||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S + ||\Omega_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A} + \Omega_H(\beta_0) - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S\right)^2$$

$$\overset{(ii)}{\le} 4||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 + 4||\Omega_s(\beta_0) - \mathcal{A}'\Omega_s(\beta_0)\mathcal{A}||_S^2 + 4||\Omega_H(\beta_0) - \mathcal{B}'\Omega_H(\beta_0)\mathcal{B}||_S^2$$

$$\overset{(iii)}{\le} 4||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 + o(K^{-1}) \tag{B.2.21}$$

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

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

with (iv) following from $\mathcal{A}'\mathcal{A} = I_K$ and (v) following in the same manner as (B.2.19). The second part of (iii) follows from

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

Next, we can express

$$||\widehat{\Omega}_s(\beta_0) - \Omega_s(\beta_0) - \Omega_H(\beta_0)||_S^2 = \left|\left|\frac{U'\hat{\Lambda}U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} - \frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_S^2$$

$$\le 2\left|\left|\frac{U'(\hat{\Lambda} - \Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}\right|\right|_S^2 + 2\left|\left|\frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)} - \frac{U'(\Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}\right|\right|_S^2$$

$$\le 2\left|\left|\frac{U'(\hat{\Lambda} - \Lambda - H_n)U}{\sum_{i\in[n]} P_{ii}e_i^2(\beta_0)}\right|\right|_S^2 + \frac{2(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0) - \sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0))^2 \cdot ||U'(\Lambda - H_n)U||_S^2}{\left(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0) \cdot \sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)\right)^2}$$

$$\overset{(i)}{=} \frac{2||U'(\hat{\Lambda} - \Lambda - H_n)U||_S^2}{(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0))^2} + o(K^{-2}) \tag{B.2.22}$$

where (i) follows from Lemma B.2.1 and $||U'(\Lambda - H_n)U||_S^2 \le ||\Lambda - H_n||_S^2 = \max_i(\sigma_i^2(\beta_0) - \Delta^2\Pi_i^2)^2 \le C$, in the same manner as in (B.2.19). We now separate the problem into two cases now to consider: **(A)** $\frac{K}{n} = o(1)$ and **(B)** $\frac{K}{n} \to c^* > 0$[4]. Suppose for the moment that

[4]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

we are under case **(A)**. Then

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

where (ii) follows from (B.2.16) 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 7. Therefore, by Lemma B.2.1 we have

$$
\frac{2||U'(\hat{\Lambda}-\Lambda-H_n)U||_S^2}{(\sum_{i\in[n]} P_{ii}e_i^2(\beta_0))^2} = o(K^{-1}) \tag{B.2.23}
$$

so that combining (B.2.19), (B.2.20), (B.2.21),(B.2.22) and (B.2.23) 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.2.18).

Next, suppose we are now under case **(B)**. Denote $\widehat{\Lambda} := diag(e_1^2+\Delta^2 v_1^2+2\Delta e_1 v_1, ..., e_n^2+\Delta^2 v_n^2+2\Delta e_n v_n)$ and $\Lambda^\dagger := 2diag(\Delta\Pi_1 e_1+\Delta^2\Pi_1 v_1, ..., \Delta\pi_n e_n+\Delta^2\Pi_n v_n)$. Then

$$
||U'(\hat{\Lambda}-\Lambda-H_n)U||_S^2 = ||U'(\widehat{\Lambda}-\Lambda+\Lambda^\dagger)U||_2^s \leq 2||U'(\widehat{\Lambda}-\Lambda)U||_S^2 + 2||U'\Lambda^\dagger U||_S^2 \tag{B.2.24}
$$

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

$$
||U'\Lambda^\dagger U||_S^2 \leq ||U'\Lambda^\dagger U||_F^2 = 4\sum_{i,j\in[n]} P_{ij}^2(\Delta\Pi_i e_i+\Delta^2\Pi_i v_i)(\Delta\Pi_j e_j+\Delta^2\Pi_j v_j)
$$

B.2.3 is provided for some sub-sequence.

$$= 4 \sum_{i,j \in [n]} P_{ij}^2 (\Delta^2 \Pi_i \Pi_j e_i e_j + 2\Delta^3 \Pi_i \Pi_j e_i v_j + \Delta^4 \Pi_i \Pi_j v_i v_j) \tag{B.2.25}$$

Furthermore,

$$\sum_{i,j \in [n]} P_{ij}^2 \Pi_i \Pi_j e_i e_j = \sum_{i,j \in [n]} P_{ij}^2 \Pi_i \Pi_j \left(\widetilde{e}_i \widetilde{e}_j - 2\widetilde{e}_j (P_i^W)' \widetilde{e} + (P_i^W)' \widetilde{e} (P_j^W)' \widetilde{e} \right) = o(K) \tag{B.2.26}$$

where the last equality follows from

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

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

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

where (i) follows from $2|ab| \leq a^2 + b^2$. In the same way as we have shown (B.2.26), 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.2.25) we can conclude

$$||U' \Lambda^\dagger U||_S^2 = o(K). \tag{B.2.27}$$

Next, we will show that

$$||U'(\widehat{\Lambda} - \Lambda)U||_S^2 = o(K) \tag{B.2.28}$$

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.2.28) it suffices to show

$$||U'(\widehat{\Lambda}_1 - \Lambda_1)U||_S^2 = o(K), \tag{B.2.29}$$

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

$$\begin{aligned}
&||U'(\widehat{\Lambda}_1 - \Lambda_1)U||_S^2 \leq 2||\widehat{\Lambda}_{1,1} - \Lambda_1||_S^2 + 2||U'(\widehat{\Lambda}_1 - \widehat{\Lambda}_{1,1})U||_S^2 \\
&\leq 2||\widehat{\Lambda}_{1,1} - \Lambda_1||_S^2 + 2||U'(\widehat{\Lambda}_1 - \widehat{\Lambda}_{1,1})U||_F^2 = \max_i (e_i^2 - \widetilde{\sigma}_i^2)^2 + \sum_{i,j \in [n]} P_{ij}^2 ((P_i^W)'\widetilde{e})^2 ((P_j^W)'\widetilde{e})^2 \\
&+ 4 \sum_{i,j \in [n]} P_{ij}^2 (\widetilde{e}_i (P_i^W)'\widetilde{e})(\widetilde{e}_j (P_j^W)'\widetilde{e}) - 4 \sum_{i,j \in [n]} P_{ij}^2 \widetilde{e}_i (P_i^W)'\widetilde{e}((P_j^W)'\widetilde{e})^2
\end{aligned} \tag{B.2.30}$$

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

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

$$\leq C\frac{n^{1/2}}{K} = C\frac{n^{1/2}}{K^{1/2}}\frac{1}{K^{1/2}} \leq C\frac{1}{K^{1/2}} = o(1).$$

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

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

$$\leq \sum_{i,j\in[n]} P_{ij}^2(\sum_{\ell\in[n]}(P_{i\ell}^W)^4 + \sum_{\ell\in[n]}\sum_{\ell'\in[n]}(P_{i\ell}^W)^2(P_{i\ell'}^W)^2) \leq (p_n^W)^2K = o(K)$$

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

$$\leq C\sum_{i,j\in[n]} P_{ij}^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2 \leq p_n^W\sum_{i,j\in[n]} P_{ij}^2 = o(K)$$

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

Putting everything together into (B.2.30) yields (B.2.29), which in turn yields (B.2.28). Combining (B.2.24), (B.2.27) and (B.2.28) yields

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

Combining the preceding equation with Lemma B.2.1, (B.2.19), (B.2.20), (B.2.21) and (B.2.22) 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.2.18) for **Case (B)**. The proof for diverging K case is complete.

□

Lemma B.2.4. *(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.2.4:

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

$$\mathbb{E}f(X_i) \to \mathbb{E}f(X) \tag{B.2.31}$$

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

$$\mathbb{E}\left(f(X_i)|\mathcal{A}\right)(\omega) = \int_{\mathbb{R}} f(x)\mathbb{P}_{X_i|\mathcal{A}}(dx,\omega) \quad \forall \omega \in N_i^c \tag{B.2.32}$$

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 ($B$.2.32) 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.2.32)

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

which proves (B.2.31)

□

Lemma B.2.5. *Assume that we do not have controls W in the data-generating process of (2.2.1). 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 8 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.2.5:

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 $Z'e(\beta_0) = Z'\xi + \Delta\sqrt{n}\widetilde{\theta}_K$.

$$n^{-1/2}Z'e(\beta_0) \rightsquigarrow \mathcal{N}\left(\Delta\Sigma_{1,K}^{1/2}(\Delta)\widetilde{\theta}_K, \Sigma_1(\Delta)\right) \tag{B.2.33}$$

where the convergence follows from the Lindeberg-Feller Central-Limit-Theorem, assumption 8, $\frac{\Pi'\Pi}{n^2} = o(1)$ 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

$$\begin{aligned}\frac{1}{n}\sum_{i\in[n]}\mathbb{E}\{||Z_i\xi||_F^2 1(||Z_i\xi||_F > \eta\sqrt{n})\} &\overset{(i)}{\leq} \frac{1}{n}\sum_{i\in[n]}\sqrt{\mathbb{E}||Z_i\xi||_F^4\mathbb{P}(||Z_i\xi||_F > \eta\sqrt{n})} \\ &\overset{(iii)}{\leq} \frac{C}{n}\sum_{i\in[n]}\frac{\mathbb{E}||Z_i\xi_i||_F^2}{\eta n} \leq \frac{C}{n}\sum_{i\in[n]}\frac{1}{\eta n} = \frac{C}{\eta n} \to 0\end{aligned}$$

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

$$\frac{Z'\widehat{\Lambda}(\beta_0)Z}{n} = \Sigma_{1,K}(\Delta) + \Delta^2\Theta_K'\Theta_K + o_p(1) \tag{B.2.34}$$

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

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

Therefore, by (B.2.33) and (B.2.34), we have

$$
\begin{aligned}
J_{n,K} &\rightsquigarrow \mathcal{Z}(\Delta\widetilde{\theta}_K)'(I_K + \Delta^2\Sigma_1(\Delta)^{-1/2}\Theta_K'\Theta\Sigma_{1,K}(\Delta)^{-1/2})^{-1}\mathcal{Z}(\Delta\widetilde{\theta}_K) \\
&\leq \frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{\lambda_{min}(I_K + \Delta^2\Sigma_{1,K}(\Delta)^{-1/2}\Theta_K'\Theta_K\Sigma_{1,K}(\Delta)^{-1/2})} \\
&= \frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2\lambda_{min}(\Sigma_{1,K}(\Delta)^{-1/2}\Theta_K'\Theta_K\Sigma_{1,K}(\Delta)^{-1/2})} \\
&\leq \frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2\lambda_{min}(\Sigma_{1,K}(\Delta)^{-1})\lambda_{min}(\Theta_K'\Theta_K)} \\
&= \frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2\frac{\lambda_{min}(\Theta_K'\Theta_K)}{\lambda_{max}(\Sigma_{1,K}(\Delta))}} \leq \frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{1 + \Delta^2 C_3},
\end{aligned}
\tag{B.2.35}
$$

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

$$
\frac{\frac{\chi^2_K(\Delta^2||\widetilde{\theta}_K||_F^2)}{K}}{1 + \Delta^2 C_3} = \frac{1 + \frac{\Delta^2||\widetilde{\theta}_K||_F^2}{K}}{1 + \Delta^2 C_3} < 1 \tag{B.2.36}
$$

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 $q_{1-\alpha}(\frac{\chi^2_K}{K}) \to 1$, combining with (B.2.35) and (B.2.36) yields

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

for some $\eta_1 > 0$.

B.3 Two Estimators Satisfying Criteria (2.2.12)

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 B.3.0.1 (Standard estimator)**.** *Suppose Assumption 5 and 6 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 B.3.0.2 (Cross-fit estimator)**.** *Suppose Assumption 5 and 6 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

$$\mathcal{D}^{cf}(\Delta) = \mathbb{E}\Bigg(\frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi V_j(\Delta)M_j'\Pi$$

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

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

B.3.1 Proof of Theorem B.3.0.1

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

$$T_0 = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1 + \sum_{i\in[4]}\Delta^i) \tag{B.3.1}$$

$$T_1 = \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_j^2(\widetilde{\sigma}_i^2 + \Delta^2\widetilde{\varsigma}_i^2 + 2\Delta\widetilde{\gamma}_i) + o_p(1 + \Delta^3 + \Delta^4) \tag{B.3.2}$$

$$T_2 = o_p(1 + \Delta^2 + \Delta^3 \tag{B.3.3}$$

$$T_3 = \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\Pi_j^2 \tag{B.3.4}$$

$$T_4 = o_p(1 + \Delta^3 + \Delta^4) \tag{B.3.5}$$

$$T_5 = o_p(1 + \Delta^2 + \Delta^3 + \Delta^4) \tag{B.3.6}$$

Combining (B.3.1)–(B.3.6) yields the second equation of Theorem B.3.0.1. By recalling that $\sigma_i^2(\beta_0) = \widetilde{\sigma}_i^2 + \Delta^2\widetilde{\zeta}_i^2 + 2\Delta\widetilde{\gamma}_i$. 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 B.3.0.1.

Step 1: We show

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 e_i^2 e_j^2 = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2\sigma_j^2 + o_p(1) \tag{B.3.7}$$

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

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

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

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

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,

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

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

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

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 $|ab| \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$,

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

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

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

implying that

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

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

$$A_1 = \Phi_1 + o_p(1) \tag{B.3.8}$$

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 (B.3.8) as

$$A_1 = \frac{2}{K} \sum_{i \in n} \sum_{j \neq i} P_{ij}^2 \widetilde{\sigma}_i^2 \widetilde{\sigma}_j^2 + o_p(1) \tag{B.3.9}$$

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

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

and (ii) follows from

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

The proof of $B_2 = o_p(1)$ is analogous to (ii). Hence (B.3.9) is shown, which proves (B.3.7).

Step 2: We show (B.3.1) In a similar way to showing (B.3.7) we have

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

Therefore by expression (B.2.1),

$$\begin{aligned}\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2V_i^2(\Delta)V_j^2(\Delta) &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2e_j^2 + \frac{2\Delta^4}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2v_j^2 + \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_ie_iv_je_j \\ &+ \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2v_j^2 + \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2v_je_j + \frac{4\Delta^3}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2v_je_j \\ &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1+\sum_{i\in[4]}\Delta^i) \qquad \text{(B.3.10)}\end{aligned}$$

Therefore (B.3.1) is shown

Step 3: We show (B.3.2). Note that we have

$$\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_i^2\Pi_j^2 = \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\Pi_j^2 + o_p(1+\Delta^2)$$
$$\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2v_i^2\Pi_j^2 = \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\varsigma}_i^2\Pi_j^2 + o_p(1+\Delta^2)$$
$$\frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2e_iv_i\Pi_j^2 = \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\gamma}_i\Pi_j^2 + o_p(1+\Delta^2) \qquad \text{(B.3.11)}$$

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

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

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 (B.3.11) 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 (B.3.2).

Step 4: We show (B.3.3). We can express

$$\Pi_j V_j(\Delta) V_i^2(\Delta) = \Pi_j e_j V_i^2(\Delta) + \Delta \Pi_j v_j V_i^2(\Delta)$$

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

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_j e_j e_i^2 = o_p(1), \tag{B.3.12}$$

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

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

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

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

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

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

Furthermore,

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

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

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

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 (B.3.12), which proves (B.3.3)

Step 5: We prove (B.3.5). Since $V_i(\Delta) = e_i + \Delta v_i$, it suffices to prove

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

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

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

and

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

Step 6: We prove (B.3.6). 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,

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

and

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

We have shown (B.3.6), and the proof is complete.

B.3.2 Proof of Theorem B.3.0.2

Observe that we can express

$$
\begin{aligned}
\widehat{\Phi}_1^{cf}(\beta_0) &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(V_i(\Delta)+\Delta\Pi_i)M_i'(V(\Delta)+\Delta\Pi)(V_j(\Delta)+\Delta\Pi_j)M_j'(V(\Delta)+\Delta\Pi)\\
&= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)V_j(\Delta)M_j'V(\Delta) + \frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi V_j(\Delta)M_j'\Pi\\
&+ \frac{2\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\Pi_i M_i'e(\beta_0)\Pi_j M_j'e(\beta_0) + \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)V_j(\Delta)M_j'\Pi\\
&+ \frac{4\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'V(\Delta)\Pi_j M_j'e(\beta_0) + \frac{4\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 V_i(\Delta)M_i'\Pi\Pi_j M_j'e(\beta_0)\\
&\equiv \sum_{\ell=0}^{5} T_\ell
\end{aligned}
$$

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

$$
T_0 = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2(\beta_0)\sigma_j^2(\beta_0) + o_p(1+\sum_{i\in[4]}\Delta^i) \tag{B.3.13}
$$

as well as

$$
\begin{aligned}
&T_\ell = \mathbb{E}T_\ell + o_p(1+\sum_{i\in[4]}\Delta^i) \quad \text{for} \quad \ell\in\{1,...,5\} \quad \text{and}\\
&\sum_{\ell\in[n]}\mathbb{E}T_\ell = \mathcal{D}^{cf}(\Delta)
\end{aligned} \tag{B.3.14}
$$

When $\Delta = 0$, it is clear that $T_1 = T_2 = ... = T_5 = 0$, so that the case of Theorem B.3.0.2 for $\Delta = 0$ is shown immediately upon proving (B.3.13); this is shown in **Step 1** below. We can therefore focus on the case of $\Delta \neq 0$.

Step 1: We prove (B.3.13):

Sub-step 1: We show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2[e_iM_i'e][e_jM_j'e]=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2+o_p(1) \tag{B.3.15}$$

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

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,1}A_{j,1}=\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2+o_p(1) \tag{B.3.16}$$

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

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

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

$$\begin{aligned}
&|\mathbb{E}\xi_{ij}\xi_{\ell k}| \\
&\leq C|M_{ii}M_{jk} + M_{ij}M_{ik})(M_{\ell\ell}M_{jk} + M_{\ell j}M_{\ell k})| + C|(M_{jj}M_{i\ell} + M_{ij}M_{\ell j})(M_{kk}M_{i\ell} + M_{k\ell}M_{i\ell})| \\
&+ C(M_{i\ell}M_{jk} + M_{ik}M_{\ell j})^2 + C(P_{ij}P_{k\ell} + P_{i\ell}P_{jk})^2
\end{aligned}$$

The second inequality of (c) follows from Mikusheva and Sun (2022)[Lemma S1.2]. Specifically, we have

$$\begin{aligned}
&\frac{1}{K^2}\sum_{i,j,k,\ell}P_{ij}^2P_{k\ell}^2|M_{ii}M_{jk}M_{\ell\ell}M_{jk}| \leq \frac{1}{K^2}\sum_{i,j,k,\ell}P_{ij}^2P_{k\ell}^2M_{jk}^2 = \frac{1}{K^2}\sum_{j,k,\ell}P_{ii}P_{k\ell}^2M_{jk}^2 \leq \frac{p_n}{K^2}\sum_{k,\ell}P_{k\ell}^2M_{kk} \\
&\leq \frac{p_n}{K^2}\sum_{k,\ell}P_{k\ell}^2 = \frac{p_n}{K},
\end{aligned}$$

with the rest of the terms in $|\mathbb{E}\xi_{ij}\xi_{\ell k}|$ dealt in a similar manner. Therefore (B.3.16) 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

$$\begin{aligned}
\mathbb{E}\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2A_{i,2}^2 &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2(P_i^W)'\mathbb{E}(\widetilde{e}_i^2\widetilde{e}\widetilde{e}')P_i^W = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\sum_{k\in[n]}(P_{ik}^W)^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_j^2 \\
&\leq \frac{Cp_n^W}{K}\sum_{i,j\in[n]}P_{ij}^2 = Cp_n^W = o(1)
\end{aligned}$$

so that by Markov inequality,

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,2}^2 = o_p(1) \tag{B.3.17}$$

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

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,3}^2 = o_p(1) \tag{B.3.18}$$

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

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,4}^2 = o_p(1) \tag{B.3.19}$$

By the simple inequality $|ab| \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$,

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}A_{j,\ell'} \leq \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}^2 + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{j,\ell'}^2 \tag{B.3.20}$$

Restricting $(\ell,\ell') \in \{2,3,4\}\times\{2,3,4\}$, by (B.3.17)-(B.3.19), using (B.3.20) we have

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,\ell}A_{j,\ell'} = o_p(1) \tag{B.3.21}$$

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 (B.3.16) to show that

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,2} = \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}(A_{i,1}A_{j,2}) + o_p(1) = o_p(1) \tag{B.3.22}$$

where the last equality follows from Markov inequality and

$$\begin{aligned}\left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}(A_{i,1}A_{j,2})\right| &= \left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\sum_{\ell\in[n]}M_{i\ell}P_{i\ell}^W\mathbb{E}(\widetilde{e}_i^2\widetilde{e}_\ell^2)\right| \leq \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sum_{\ell\in[n]}|M_{i\ell}P_{i\ell}^W| \\ &\overset{(i)}{\leq} \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\sum_{\ell\in[n]}M_{i\ell}^2\sum_{\ell\in[n]}(P_{i\ell}^W)^2 = \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 M_{ii}P_{ii}^W \\ &\leq \frac{Cp_n^W}{K}\sum_{i,j\in[n]}P_{ij}^2 = Cp_n^W = o(1)\end{aligned}$$

where (i) follows from Cauchy-Schwartz inequality. Next, we will show

$$\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,3} = o_p(1) \tag{B.3.23}$$

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

$$\begin{aligned}
(a)\quad &\sum_{\mathcal{J}_1}|M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}| = \sum_k M_{ik}^2M_{jk}^2(P_{ik}^W)^2 \le M_{ii}(p_n^W)^2\le p_n^W\\
(b)\quad &\sum_{\mathcal{J}_2}|M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}| \le C\sum_{k,\ell'}|M_{ik}P_{ik}^W M_{jk}||M_{i\ell'}P_{i\ell'}^W M_{j\ell'}|\\
&\le C(p_n^W)^2\sum_{k,\ell'}|M_{ik}M_{jk}||M_{i\ell'}M_{j\ell'}| = Cp_n^W(\sum_k|M_{ik}M_{jk}|)^2\\
&\overset{(i)}{\le} Cp_n^W\sum_k M_{ik}^2\sum_k M_{jk}^2 = Cp_n^W M_{jj}M_{jj}\le Cp_n^W\\
(c)\quad &\sum_{\mathcal{J}_3}|M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}| \le C\sum_{k,\ell,m}|M_{ik}P_{ik}^W M_{j\ell}M_{i\ell}P_{im}^W M_{jm}|\\
&\overset{(ii)}{\le} CM_{ii}P_{ii}^W M_{jj}M_{ii}P_{ii}^W M_{jj}\le Cp_n^W\\
(d)\quad &\sum_{\mathcal{J}_4}|M_{ik}P_{ik'}^W M_{j\ell}M_{i\ell'}P_{im}^W M_{jm'}| \le C\sum_{k,\ell'}|M_{ik}P_{ik}^W M_{jk}M_{i\ell'}P_{i\ell'}^W M_{ji}|\\
&\le C\sum_{k,\ell'}|M_{ik}P_{ik}^W M_{jk}M_{i\ell'}P_{i\ell'}^W| \le Cp_n^W\sum_k|M_{ik}M_{jk}|\sum_{\ell'}|M_{i\ell'}P_{i\ell'}^W|\\
&\overset{(iii)}{\le} Cp_n^W M_{ii}M_{jj}M_{ii}P_{ii}^W\le Cp_n^W
\end{aligned}$$

where (i),(ii) and (iii) follows by Cauchy-Schwartz inequality. Putting (a)-(d) together we have

$$\mathbb{E}\widetilde{e}_i^2((M_i'\widetilde{e})((P_i^W)'\widetilde{e})(M_j'\widetilde{e}))^2 \leq Cp_n^W. \tag{B.3.24}$$

Hence

$$\mathbb{E}\left(\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,3}\right)^2 = \frac{4}{K^2}\sum_{i,i'}\sum_{j\neq i}\sum_{j'\neq i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2\mathbb{E}[\widetilde{e}_i M_i'\widetilde{e}((P_j^W)'\widetilde{e}M_j'\widetilde{e})][\widetilde{e}_{i'}M_{i'}'\widetilde{e}((P_{j'}^W)'\widetilde{e}M_{j'}'\widetilde{e}]$$

$$\overset{(i)}{\leq}\frac{2}{K^2}\sum_{i,i'}\sum_{j\neq i}\sum_{j'\neq i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2\mathbb{E}[\widetilde{e}_i M_i'\widetilde{e}((P_j^W)'\widetilde{e}M_j'\widetilde{e})]^2 + \frac{2}{K^2}\sum_{i,i'}\sum_{j\neq i}\sum_{j'\neq i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2\mathbb{E}[\widetilde{e}_{i'}M_{i'}'\widetilde{e}((P_{j'}^W)'\widetilde{e}M_{j'}'\widetilde{e}]^2$$

$$\overset{(ii)}{\leq}\frac{Cp_n^W}{K^2}\sum_{i,i'}\sum_{j\neq i}\sum_{j'\neq i'}\widetilde{P}_{ij}^2\widetilde{P}_{i'j'}^2 \leq \frac{Cp_n^W}{K^2}\sum_{i,i',j,j'}P_{ij}^2P_{i'j'}^2 = Cp_n^W = o(1)$$

where (i) follows from $2|ab| \leq a^2+b^2$ and (ii) follows from (B.3.24). By Markov inequality, (B.3.23) is shown. Finally,

$$\mathbb{E}\left|\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 A_{i,1}A_{j,4}\right| \overset{(i)}{\leq} \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\left(\mathbb{E}(\widetilde{e}_i(P_j^W)'\widetilde{e})^2 + \mathbb{E}(M_i'\widetilde{e}(P_j^W)'\widetilde{e})^2\right)$$

$$= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\left(\sum_{\ell\in[n]}(P_{j\ell}^W)^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_\ell^2 + \mathbb{E}(M_i'\widetilde{e}(P_j^W)'\widetilde{e})^2\right) \overset{(ii)}{=} o(1)$$

where (i) follows from $2|ab| \leq a^2+b^2$ and (ii) follows from

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\sum_{\ell\in[n]}(P_{j\ell}^W)^2\mathbb{E}\widetilde{e}_i^2\widetilde{e}_\ell^2 \leq \frac{C}{K}\sum_{i,j\in[n]}P_{ij}^2P_{jj}^W \leq Cp_n^W = o(1)$$

and

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}(M_i'\widetilde{e}(P_j^W)'\widetilde{e})^2 \leq \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\left(\sum_{k,\ell}(M_{ik})^2(P_{j\ell}^W)^2 + \sum_k (M_{ik})^2(P_{jk}^W)^2\right)$$

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

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

Therefore

$$\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 A_{i,1} A_{j,4} = o_p(1). \tag{B.3.25}$$

Putting (B.3.16)-(B.3.25) yields (B.3.15).

Sub-step 2: In a similar way to **sub-step 1**, we can show that

$$\begin{aligned}
\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 e_i M_i' e e_j M_j' v &= \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \widetilde{\sigma}_i^2 \widetilde{\gamma}_j + o_p(1) \\
\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 v_i M_i' v v_j M_j' v &= \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \widetilde{\varsigma}_i^2 \widetilde{\varsigma}_j^2 + o_p(1) \\
\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} \widetilde{P}_{ij}^2 v_i M_i' e v_j M_j' e &= \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \widetilde{\gamma}_i \widetilde{\gamma}_j + o_p(1)
\end{aligned} \tag{B.3.26}$$

By expression (B.2.1) 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 (B.3.15) and (B.3.26) yields (B.3.13).

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 (2.2.12) for $\ell \in \{1, ..., 5\}$, in order to complete the proof of (B.3.14). Note first that

$$\mathbb{E}e_i^2 = \mathbb{E}(\widetilde{e}_i - (P_i^W)'\widetilde{e})^2 = \widetilde{\sigma}_i^2 + \sum_{\ell\in[n]} (P_{i\ell}^W)^2 \widetilde{\sigma}_i^2 - 2P_{ii}^W \widetilde{\sigma}_i^2 \leq C$$

since $\sum_{\ell\in[n]}(P_{i\ell}^W)^2 = P_{ii}^W \le 1$, by property of a projection matrix. Similarly,

$$\mathbb{E}v_i^2 \le C \quad \text{and} \quad \mathbb{E}v_i e_i \le C,$$

so that

$$\mathbb{E}V_i^2(\Delta) = \mathbb{E}e_i^2 + \Delta^2\mathbb{E}v_i^2 + 2\Delta\mathbb{E}v_i e_i \le C(1+\Delta+\Delta^2) \tag{B.3.27}$$

By the inequality $(a+b)^2 \le 2a^2 + 2b^2$ and noting that $\widetilde{P}_{ij}^2 \le CP_{ij}^2$, we have

$$\begin{aligned}
\mathbb{E}|T_1| &\le \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2\mathbb{E}V_i^2(\Delta)(M_i'\Pi)^2 \le \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}V_i^2(\Delta)(M_i'\Pi)^2 \\
&\le \frac{C\Delta^2(1+\Delta+\Delta^2)}{K}\sum_{i\in[n]}P_{ii}(M_i'\Pi)^2 \le \frac{C\Delta^2(1+\Delta+\Delta^2)p_n}{K}\sum_{i\in[n]}(M_i'\Pi)^2 \\
&= \frac{C\Delta^2(1+\Delta+\Delta^2)p_n}{K}\Pi' M\Pi = O\left(\Delta^2+\Delta^3+\Delta^4\right)
\end{aligned}$$

For T_2, note that

$$\mathbb{E}(M_i'V(\Delta))^2 \le C(1+\Delta+\Delta^2) \tag{B.3.28}$$

To see this, it suffices to show $\mathbb{E}(M_i'e)^2 \le C$, 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

$$\begin{aligned}
\mathbb{E}(M_i'e)^2 &= \mathbb{E}(M_i'M^W\widetilde{e})^2 = \mathbb{E}((M_i^W)'\widetilde{e} - P_i'\widetilde{e})^2 \le 2\mathbb{E}((M_i^W)'\widetilde{e})^2 + 2\mathbb{E}(P_i'\widetilde{e})^2) \\
&= 2\sum_{\ell\in[n]}(M_{i\ell}^W)^2\widetilde{\sigma}_\ell^2 + 2\sum_{\ell\in[n]}P_{i\ell}^2\widetilde{\sigma}_\ell^2 \le CM_{ii}^W + CP_{ii} \le C
\end{aligned}$$

since $M_{ii}^W, P_{ii} \le 1$. This implies (B.3.28). Expressing $M_i'e(\beta_0) = M_i'V(\Delta) + \Delta M_i'\Pi$, we have

$$\mathbb{E}|T_2| \le \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\mathbb{E}(M_i'e(\beta_0))^2 \le \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\Pi_i^2\mathbb{E}((M_i'V(\Delta))^2 + \Delta^2(M_i'\Pi)^2)$$

$$\leq \frac{C\Delta^2(1+\Delta+\Delta^2)}{K}\sum_{i,j\in[n]}P_{ij}^2\Pi_i^2 + \frac{C\Delta^4}{K}\sum_{i,j\in[n]}P_{ij}^2(M_i'\Pi)^2$$
$$\leq \frac{C\Delta^2(1+\Delta+\Delta^2)p_n\Pi'\Pi}{K} + \frac{C\Delta^4}{K}\sum_{i\in[n]}P_{ii}(M_i'\Pi)^2$$
$$\leq \frac{C\Delta^2(1+\Delta+\Delta^2)p_n\Pi'\Pi}{K} + C\Delta^4\frac{p_n\Pi'M\Pi}{K} = O\left(\Delta^2+\Delta^3+\Delta^4\right)$$

Next, to deal with T_3 we first show that

$$\mathbb{E}V_i^2(\Delta)\cdot(M_i'V(\Delta))^2 \leq C(1+\sum_{i\in[4]}\Delta^i) \tag{B.3.29}$$

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

$$\mathbb{E}e_i^2((M_i^W)'\widetilde{e})^2 = \mathbb{E}\widetilde{e}_i^2((M_i^W)'\widetilde{e})^2 + \mathbb{E}((P_i^W)'\widetilde{e})^2((M_i^W)'\widetilde{e})^2 - 2\mathbb{E}\widetilde{e}_i(P_i^W)'\widetilde{e}((M_i^W)'\widetilde{e})^2$$
$$\leq C\left(\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + \sum_{\ell\in[n]}(P_{i\ell}^W)^2\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + (\sum_{\ell\in[n]}|P_{i\ell}^W M_{i\ell}^W|)^2 + CP_{ii}^W\sum_{\ell\in[n]}(M_{i\ell}^W)^2 + M_{ii}^W\sum_{\ell\in[n]}|P_{i\ell}^W M_{i\ell}^W|\right.$$
$$\leq C\left(M_{ii}^W + P_{ii}^W M_{ii}^W + (M_{ii}^W)^2 P_{ii}^W\right) \leq C.$$

Hence (B.3.29) is shown. Then

$$\mathbb{E}|T_3| \leq \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\mathbb{E}(V_i^2(\Delta)\cdot(M_i'V(\Delta))^2 + V_j^2(\Delta)\cdot(M_j'\Pi)^2)$$
$$\overset{(B.3.27),(B.3.29)}{\leq} \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2 + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2(M_j'\Pi)^2$$
$$\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i)\frac{p_n\Pi'M\Pi}{K} = O\left(\sum_{i\in[5]}(1+\frac{p_n\Pi'M\Pi}{K})\Delta^i\right) = O\left(\sum_{i\in[5]}\Delta^i\right)$$

Next,

$$
\begin{aligned}
\mathbb{E}|T_4| &\leq \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \mathbb{E}\left(V_i^2(\Delta)(M_i'V(\Delta))^2 + \Pi_j^2(M_j'e(\beta_0))^2\right) \\
&\overset{(B.3.29)}{\leq} \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(M_j'e(\beta_0))^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(M_j'V(\Delta)+\Delta M_j'\Pi)^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(M_j'V(\Delta))^2 + \frac{C\Delta}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(\Delta M_j'\Pi)^2 \\
&\overset{(B.3.28)}{\leq} C\Delta(1+\sum_{i\in[4]}\Delta^i) + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 + \frac{C\Delta(1+\sum_{i\in[4]}\Delta^i)}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(M_j'\Pi)^2 \\
&\leq C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i) + C\Delta(1+\sum_{i\in[4]}\Delta^i)\frac{p_n\Pi'M\Pi}{K} = O\left(\sum_{i\in[5]}\Delta^i\right)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathbb{E}|T_5| &\leq \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \mathbb{E}\left(V_i^2(\Delta)(M_i'\Pi)^2 + \Pi_j^2(M_j'e(\beta_0))^2\right) \\
&\overset{(B.3.27)}{\leq} \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 + \frac{C\Delta^2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\mathbb{E}(M_j'e(\beta_0))^2 \\
&\overset{(i)}{\leq} C\Delta^2 + C\Delta^2\frac{p_n\Pi'M\Pi}{K} = O(\Delta^2)
\end{aligned}
$$

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 (B.3.14) is shown, and the proof is complete.

B.4 Limit Problem For Fixed And Diverging Numbers of Instruments

B.4.1 Limit Problem for Fixed K

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 B.4.1.1 (Fixed K asymptotics)**.** *Suppose Assumption 5 and 6 holds. Then for fixed K, under the null*

$$\widehat{Q}(\beta_0) \stackrel{d}{=} \sum_{i\in[K]} w_{i,n}\chi^2_{1,i} + o_p(1)$$

where the $\chi^2_{1,i}$ are independent chi-squares with one degree-of-freedom and $D_n := diag(w_{1,n}, ..., w_{K,n})$ are the eigenvalues of $\frac{(Z'\Lambda Z)^{1/2}(Z'Z)^{-1}(Z'\Lambda Z)^{1/2}}{\sum_{i\in[n]} P_{ii}\sigma_i^2(\beta_0)}$.

B.4.2 Limit Problem for Diverging K

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 $\overline{\mathcal{C}} := Q_{\widetilde{\Pi},\widetilde{\Pi}} \to \infty$ and weak identification otherwise. Under the arguments of Chao et al. (2012) and Mikusheva and Sun (2022), by assumption 5 and 6, one can obtain the following asymptotics for diverging K: Under both Weak and Strong

Identification, for $K \to \infty$,

$$\begin{pmatrix} Q_{\widetilde{e},\widetilde{e}} \\ Q_{\widetilde{X},\widetilde{e}} \\ Q_{\widetilde{X},\widetilde{X}} - \overline{\mathcal{C}} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \overline{\Phi}_1 & \overline{\Phi}_{12} & \overline{\Phi}_{13} \\ \overline{\Phi}_{12} & \overline{\Psi} & \overline{\tau} \\ \overline{\Phi}_{13} & \overline{\tau} & \overline{\Upsilon} \end{pmatrix} \right) \tag{B.4.1}$$

for $\overline{\mathcal{C}} := Q_{\widetilde{\Pi},\widetilde{\Pi}}$, for some $(\overline{\Phi}_1, \overline{\Phi}_{12}, \overline{\Phi}_{13}, \overline{\Psi}, \overline{\tau}, \overline{\Upsilon})$. We can therefore take (B.4.1) as given whenever assumption 5 and 6 holds. Under a fixed number of controls, one can usually obtain an analogous result to (B.4.1) with the replacement of $(\widetilde{e}, \widetilde{X})$ with (e, X). However, even when the number of controls increase with sample size, as long as these controls grow slower than $K^{(1-\eta)/4}$, we will have the following result:

Theorem B.4.2.1. *Suppose Assumptions 5 and 6 hold. Then for $K \to \infty$, under the null,*

$$Q_{e,e} \rightsquigarrow \mathcal{N}(0, \Phi_1)$$

where $\Phi_1 = \frac{2}{K} \sum_{i\in[n]} \sum_{j\in[n]} P_{ij}^2 \widetilde{\sigma}_i^2 \widetilde{\sigma}_j^2$ Furthermore, under the alternative, if we further assume that $\frac{\Pi'\Pi}{K} = O(1)$, then

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix} \right) \tag{B.4.2}$$

for some $(\Phi_{12}, \Phi_{13}, \Psi, \tau, \Upsilon)$. Therefore we have that

$$Q_{e(\beta_0),e(\beta_0)} \rightsquigarrow \mathcal{N}(\Delta^2 \mathcal{C}, \Phi_1(\beta_0))$$

where $\mathcal{C} := Q_{\Pi,\Pi}$, $\Phi_1(\beta_0) = \Delta^4 \Upsilon + 4\Delta^3 \tau + \Delta^2(4\Psi + 2\Phi_{13}) + 4\Delta\Phi_{12} + \Phi_1$

Note that Theorem B.4.2.1 can be seen as a minor extension of Theorem A.1 in Lim, Wang, and Zhang (2024) in that the dimensions of controls were taken as fixed in that paper.

Theorem B.4.2.2 (Diverging K asymptotics)**.** *Suppose Assumption 5 and 6 holds. Then* *for $K \to \infty$, for $\beta = \beta_0$ we have*

$$\frac{1}{\sqrt{K}} \sum_{i \in [n]} P_{ii} e_i^2(\beta_0) \left(\widehat{Q}(\beta_0) - 1 \right) \rightsquigarrow \mathcal{N}(0, \Phi_1).$$

If we further assume that $\frac{\Pi'\Pi}{K} = O(1)$, under fixed alternative Δ we have

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

B.4.3 Proofs for Section B.4

Proof of Theorem B.4.1.1

By Lemma B.2.1 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 $\mathcal{Z} \sim \mathcal{N}(0, I_K)$.

Proof of Theorem B.4.2.1

We will show that

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix} \right) \tag{B.4.3}$$

so that by writing $Q_{e(\beta_0),e(\beta_0)} = Q_{e+\Delta X,e+\Delta X} = Q_{e,e} + \Delta^2 Q_{X,X} + 2\Delta Q_{X,e}$, then

$$Q_{e(\beta_0),e(\beta_0)} - \Delta^2 \mathcal{C} = \begin{pmatrix} 1 & 2\Delta & \Delta^2 \end{pmatrix} \begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}(0, \Phi_1(\beta_0))$$

which completes the proof.

We will show the following:

$$\begin{aligned}
&(A) \quad Q_{e,e} = Q_{\widetilde{e},\widetilde{e}} + o_p(1) \rightsquigarrow \mathcal{N}(0, \Phi_1) \\
&(B) \quad Q_{X,e} = Q_{\widetilde{v},\widetilde{e}} + \frac{\sum_{i\in[n]}(G_i + \theta_i)\widetilde{e}_i}{\sqrt{K}} + o_p(1) \\
&(C) \quad Q_{X,X} = Q_{\Pi,\Pi} + Q_{\widetilde{v},\widetilde{v}} + 2\frac{\sum_{i\in[n]}(G_i + \theta_i)\widetilde{v}_i}{\sqrt{K}} + o_p(1)
\end{aligned}$$

where $\theta_i := \sum_{j\neq i} P_{ij}\Pi_j$ and $G_i := \sum_{j\in[n]} \Pi_j P_{jj} P^W_{ij}$. To proof the second part of the theorem, given that $\{\widetilde{e}_i, \widetilde{v}_i\}_{i\in[n]}$ are independent, we can follow the proof of Chao et al. (2012)[Lemma A2] to show the joint asymptotic normality of

$$\left(Q_{\widetilde{e},\widetilde{e}}, Q_{\widetilde{v},\widetilde{e}}, Q_{\widetilde{v},\widetilde{v}}, \frac{\sum_{i\in[n]}(G_i + \theta_i)\widetilde{e}_i}{\sqrt{K}}, \frac{\sum_{i\in[n]}(G_i + \theta_i)\widetilde{v}_i}{\sqrt{K}} \right)$$

Then (B.4.3) follows from $(A), (B)$ and (C). In particular, if $\frac{\Pi'\Pi}{K} = O(1)$, then denoting $\pi_j := \Pi_j P_{jj}$ and noting $G_i = (P^W_i)'\pi$,

$$\begin{aligned}
Var\left(\frac{\sum_{i\in[n]}(G_i + \theta_i)\widetilde{e}_i}{\sqrt{K}}\right) &= \frac{\sum_{i\in[n]}(G_i + \theta_i)^2\widetilde{\sigma}_i^2}{K} \leq \frac{C\sum_{i\in[n]} G_i^2}{K} + \frac{C\sum_{i\in[n]} \theta_i^2}{K} \\
&\overset{(i)}{\leq} \frac{C\sum_{i\in[n]} G_i^2}{K} + \frac{C\Pi'\Pi}{K} = \frac{C\pi'\sum_{i\in[n]} P^W_i (P^W_i)'\pi}{K} + O(1) \\
&= \frac{C\pi'(P^W)^2\pi}{K} + O(1) \leq \frac{C\pi'\pi}{K} + O(1) = \frac{C\sum_{i\in[n]} P_{ii}^2\Pi_i^2}{K} + O(1) \\
&= Cp_n^2\frac{\Pi'\Pi}{K} + O(1) = O(1)
\end{aligned}$$

where (i) follows from Mikusheva and Sun (2022)[Lemma S1.4(a)]. In a similar manner we can show that $Var\left(\frac{\sum_{i\in[n]}(G_i+\theta_i)\widetilde{v}_i}{\sqrt{K}}\right) = O(1)$. This implies the joint asymptotic normality of

$$(Q_{e,e}, Q_{X,e}, Q_{X,X} - Q_{\Pi,\Pi}),$$

completing the proof of (B.4.3).

To this end, we begin by showing (A), which proves the first part of Theorem B.4.2.1. Suppose only that assumption 5 and 6 holds. Then WPA1, where the equalities are in terms of distribution,

$$Q_{e,e} = \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij}e_ie_j}{\sqrt{K}} \overset{(i)}{=} \frac{1}{\sqrt{K}}\varepsilon' P\varepsilon - \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{\sqrt{K}} \overset{(ii)}{=} \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{\sqrt{K}}\left(\sum_{i\in[K]} w_{i,n}\chi^2_{1,i} - 1\right)$$

where (i) follows from Theorem 1 for fixed K and $M^W P = P$; (ii) follows in the same way as the proof of Theorem B.4.1.1. Therefore, defining $T_n := \frac{\sum_{i\in[n]} P_{ii}\widetilde{\sigma}_i^2}{\sqrt{K}}$ and noting that T_n is away from zero, we have WPA1

$$\begin{aligned} Q_{e,e} &\overset{d}{=} \frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2}{\sqrt{K\Phi_1}}\left(\sum_{i\in[K]} w_{i,n}\chi^2_{1,i} - 1\right) = \frac{T_n}{\sqrt{\Phi_1}}\frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2/\sqrt{K}}{T_n}\left(\sum_{i\in[K]} w_{i,n}\chi^2_{1,i} - 1\right) \\ &\overset{(i)}{=} \frac{\sum_{i\in[n]} P_{ii}\widetilde{\sigma}_i^2}{\sqrt{K\Phi_1}}\sum_{i\in[K]} w_{i,n}(\chi^2_{1,i} - 1) \overset{(ii)}{=} \sum_{i\in[K]}\frac{w_{i,n}}{\sqrt{2}||w_n||_F}(\chi^2_{1,i} - 1) \rightsquigarrow \mathcal{N}(0,1) \end{aligned}$$

where (i) follows from $\frac{\sum_{i\in[n]} P_{ii}\varepsilon_i^2/\sqrt{K}}{T_n} \overset{p}{\to} 1$ as a consequence of Lemma B.2.1, as well as the fact that $\sum_{i\in[K]} w_{i,n} = 1$; (ii) follows from $\Phi_1 = \frac{2}{K}\sum_{i,j\in[n]} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2$ and $||w_n||_F = \frac{\sqrt{\sum_{i,j\in[n]} P_{ij}^2\widetilde{\sigma}_i^2\widetilde{\sigma}_j^2}}{\sum_{i\in[n]} P_{ii}\widetilde{\sigma}_i^2}$: this follows from (a) in the proof of Lemma 2.4.1. It remains to show that $Q_{e,e} = Q_{\widetilde{e},\widetilde{e}} + o_p(1)$, which follows from

$$\begin{aligned} Q_{e,e} - Q_{\widetilde{e},\widetilde{e}} &= \frac{\widetilde{e}' P\widetilde{e}}{\sqrt{K}} - \frac{\sum_{i\in[n]} P_{ii}e_i^2}{\sqrt{K}} - Q_{\widetilde{e},\widetilde{e}} = \frac{\sum_{i\in[n]} P_{ii}(\widetilde{e}_i^2 - e_i^2)}{\sqrt{K}} \\ &= \frac{\sum_{i\in[n]} P_{ii}(2\widetilde{e}_i P_i^W\widetilde{e} - (P_i^W\widetilde{e})^2))}{\sqrt{K}} = o_p(1), \end{aligned} \quad \text{(B.4.4)}$$

where the last equality follows fom an application or Markov inequality and

$$\mathbb{E}\left(\frac{\sum_{i\in[n]} P_{ii}\widetilde{e}_i P_i^W \widetilde{e}}{\sqrt{K}}\right)^2 = \frac{\sum_{i\in[n]}\sum_{j\in[n]} P_{ii}P_{jj}\mathbb{E}(\widetilde{e}_i\widetilde{e}_j P_i^W\widetilde{e}\cdot P_j^W\widetilde{e})}{K}$$
$$\leq \frac{C}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ii}P_{jj}((P_{ij}^W)^2 + P_{ii}^W P_{jj}^W) \leq \frac{Cp_n^W p_n}{K}\sum_{i\in[n]} P_{ii} + \frac{Cp_n^2 d_W^2}{K}$$
$$\leq Cp_n^W p_n + \frac{Cp_n^w d_W^2}{K} \overset{(i)}{=} o(1)$$

and

$$\mathbb{E}\left(\frac{\sum_{i\in[n]} P_{ii}(P_i^W\widetilde{e})^2}{\sqrt{K}}\right) = \frac{\sum_{i\in[n]} P_{ii}\sum_{j\in[n]}(P_{ij}^W)^2\widetilde{\sigma}_j^2}{\sqrt{K}} \leq C\frac{\sum_{i\in[n]} P_{ii}P_{ii}^W}{\sqrt{K}} \leq Cp_n\frac{d_W}{\sqrt{K}} = o(1),$$

where (i) follows from $p_n^W = o(1)$ and $d_W^2 = O(K^{(1-\eta)/2}) = o(K)$. The proof of (A) is complete.

It remains to prove (B) and (C) in order to complete the proof for the second part of the theorem. We first prove (B). By a similar proof to (B.4.4) we can show that

$$Q_{v,e} = Q_{\widetilde{v},\widetilde{e}} + o_p(1)$$

so that

$$Q_{X,e} = Q_{\Pi,e} + Q_{v,e} = Q_{\Pi,\widetilde{e}} - Q_{\Pi,P^W\widetilde{e}} + Q_{\widetilde{v},\widetilde{e}} + o_p(1) = Q_{\Pi+\widetilde{v},\widetilde{e}} + \frac{\sum_{i\in[n]} P_{ii}\Pi_i(P_i^W)'\widetilde{e}}{\sqrt{K}} + o_p(1)$$
$$= Q_{\widetilde{v},\widetilde{e}} + \frac{\sum_{i\in[n]}(G_i+\theta_i)\widetilde{e}_i}{\sqrt{K}} + o_p(1)$$

To prove (C), note that by a similar proof to (B.4.4) we can show that

$$Q_{v,v} = Q_{\widetilde{v},\widetilde{v}} + o_p(1).$$

Furthermore, as in the proof of (B), by some rearrangement we can show that

$$Q_{\Pi,v} = Q_{\Pi,\widetilde{v}} + Q_{\Pi,P^W\widetilde{v}} = \frac{\sum_{i\in[n]}(G_i+\theta_i)\widetilde{v}_i}{\sqrt{K}},$$

so that putting it together,

$$Q_{X,X} = Q_{\Pi,\Pi} + 2Q_{\Pi,v} + Q_{v,v} = Q_{\Pi,\Pi} + 2\frac{\sum_{i\in[n]}(G_i+\theta_i)\widetilde{v}_i}{\sqrt{K}} + Q_{\widetilde{v},\widetilde{v}} + o_p(1),$$

which completes the proof of $(A), (B)$ and (C), thereby completing the proof of the second part of Theorem B.4.2.1.

Proof of Theorem B.4.2.2

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 B.4.2.1,

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

B.5 Details regarding Testing Under Rank Deficiency

In this section we provide details of the our testing procedure as well as its asymptotic properties.

B.5.1 Analytical Test under Rank Deficiency

The analogous statistic $\widehat{Q}(\beta_0)$ given in (2.2.4) under the ridge-projection matrix is

$$\widehat{Q}^{\gamma_n}(\beta_0) := \frac{e(\beta_0)'P_{\gamma_n}e(\beta_0)}{\sum_{i\in[n]} P_{ii,\gamma_n}e_i^2(\beta_0)}, \tag{B.5.1}$$

with the corresponding critical value as

$$C_{\alpha,df,\gamma_n}(\widehat{\Phi}_1^{\gamma_n}(\beta_0)) := 1 + \frac{\sqrt{\widehat{\Phi}_1^{\gamma_n}(\beta_0)}}{\frac{1}{\sqrt{r}}\sum_{i\in[n]} P_{ii,\gamma_n} e_i^2(\beta_0)} \left(\frac{q_{1-\alpha}(F_{\widetilde{w}_n}) - 1}{\sqrt{2\sum_{i\in[r]}(\widetilde{w}_{i,n}^{\gamma_n})^2 + 1/df}} \right), \quad \text{(B.5.2)}$$

where $\widetilde{w}_n^{\gamma_n} = (\widetilde{w}_{1,n}^{\gamma_n}, \cdots, \widetilde{w}_{r,n}^{\gamma_n})'$ are the eigenvalues of

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

$\widehat{\Lambda}(\beta_0)$ is defined as in section 2.2.3, P_{ij,γ_n} are the (i,j) entries of P_{γ_n} and

$$df^{-1} = o(r^{-1/2}). \quad \text{(B.5.3)}$$

Note that the rank of $\widehat{\Omega}^{\gamma_n}(\beta_0)$ equals r, so that it has only r non-zero eigenvalues. The variance estimator $\widehat{\Phi}_1^{\gamma_n}(\beta_0)$ satisfies

$$\widehat{\Phi}_1^{\gamma_n}(\beta_0) = \Phi_1^{\gamma_n}(\beta_0) + \mathcal{D}^{\gamma_n}(\Delta) + o_p(1 + \sum_{i\in[4]} \Delta^i) \quad \text{(B.5.4)}$$

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

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

We have two estimators satisfying (B.5.4) that are analogous to the standard and cross-fit estimator of section 2.2.5; namely,

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

and

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

where $M_{\gamma_n} := I_n - P_{\gamma_n}$. The proof that $\widehat{\Phi}_1^{\gamma_n,standard}(\beta_0)$ and $\widehat{\Phi}_1^{\gamma_n,cf}(\beta_0)$ satisfies (B.5.4) follows in exactly the same way as the proof of Theorems B.3.0.1 and B.3.0.2 respectively, with an additional usage of Lemma B.5.1; hence we omit them to avoid repetition. Our analytical test rejects $H_0 : \beta = \beta_0$ at α significance-level if

$$\widehat{Q}^{\gamma_n^*}(\beta_0) > C_{\alpha,df}^{\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)).$$

The intuition for size-control is exactly the same as what was described in section 2.2.3.

B.5.2 Bootstrap-based Test under Rank Deficiency

The Bootstrap-based statistic is defined as

$$\widehat{J}^{\gamma_n}(\beta_0, \widehat{\Phi}_1^{\gamma_n}(\beta_0)) := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n} e_i(\beta_0)e_j(\beta_0)}{\sqrt{r\widehat{\Phi}_1^{\gamma_n}(\beta_0)}} \tag{B.5.5}$$

with $\widehat{\Phi}_1^{\gamma_n}(\beta_0)$ satisfying (B.5.4) with the additional requirement that it can be constructed from $e(\beta_0)$ and P_{γ_n}. We reject $H_0 : \beta = \beta_0$ at α significance-level if

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

where $C_{\alpha,df_{BS}}^{\gamma_n}(\widehat{\Phi}_1^{\gamma_n}(\beta_0), \mathcal{L})$ is the critical value that depends (1) on some large positive integer B, (2) significance-level α, (3) i.i.d. random variables $\{\kappa_i\}_{i\in[n]}$ following the probability law $\mathcal{L}$ with the property that its mean is zero, variance is one, fourth moment is bounded, (4) the structure of the variance estimator $\widehat{\Phi}_1^{\gamma_n}(\beta_0)$ and (5) sequence of γ_n. 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 i.i.d. 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 $\widehat{\Phi}_1^{\gamma_n}(\beta_0)$ is assumed to be constructed by using only $e(\beta_0)$ and P_{γ_n}, we construct $\widehat{\Phi}_1^{\gamma_n,\ell}(\beta_0)$ in exactly the same way that $\widehat{\Phi}_1^{\gamma_n}(\beta_0)$ was constructed, but replacing $(e(\beta_0), P_{\gamma_n})$ with $(\eta_\ell, P_{\gamma_n})$, where $\eta_\ell = (\eta_{1,\ell}, ..., \eta_{n,\ell})'$. Once this is done, we can construct the statistic

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

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

$$C^{\gamma_n}_{\alpha,df_{BS}}(\widehat{\Phi}_1^{\gamma_n}(\beta_0), \mathcal{L}) := inf\left\{z \in \mathbb{R} : 1-\alpha \leq \frac{\sum_{\ell\in[B]} 1\left\{\widehat{J}^{\gamma_n,\ell} \leq z\right\}}{B}\right\} + 1/df_{BS} \quad \text{(B.5.6)}$$

where $df_{BS}^{-1} = o(1)$ is a deterministic sequence.

B.5.3 Asymptotic Size Control under Rank Deficiency

Define $p_n^{\gamma_n} := \max_{i\in[n]} P_{ii,\gamma_n}$. We make the following assumption:

Assumption 13. *Suppose* $p_n^{\gamma_n^*} \leq \overline{C}\frac{r}{n}$ *for some* $\overline{C} < \infty$

Let $\overline{\lambda}_n \in \overline{\Lambda}_n$ 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 ($\{\overline{\Lambda}_n\}_{n\geq 1}$):

$$\left(\begin{array}{c}\{\widetilde{e}_i, \widetilde{v}_i\}_{i\in[n]} \text{ are independent, } \mathbb{E}\widetilde{e}_i = \mathbb{E}\widetilde{v}_i = 0,\\ \frac{p_n^{\gamma_n^*}}{r} = o(1), p_n^W = o(1), d_W = O(K^{(1-\eta)/4}) \text{ for any } \eta > 0,\\ \max_i \Pi_i^2 + \max_i \mathbb{E}\widetilde{e}_i^8 + \max_i \mathbb{E}\widetilde{v}_i^8 \leq \overline{C} < \infty,\\ \Pi'\Pi, \sigma_i^2(\beta_0), \zeta_i^2(\beta_0) \geq \underline{C} \text{ under the null,}\\ \underline{C} \leq \lambda_{min}(\frac{W'W}{n}) \leq \lambda_{max}(\frac{W'W}{n}) \leq \overline{C},\\ \exists\gamma_n \in [\overline{\gamma}, \infty), h \geq 1 \text{ s.t. } \sum_{i\in[n]}\sum_{j\neq i} P^2_{ij,\gamma_n} \geq \underline{C}r^h, \overline{\gamma} = 0 \text{ if } r = K, \overline{\gamma} = \gamma_- \text{ if } r < K\\ \widehat{\Phi}_1^{\gamma_n^*}(\beta_0) \text{ satisfies (B.5.4) under the null,}\\ \text{where } 0 < \underline{C}, \overline{C}, \gamma_- < \infty \text{ are some fixed constants}\end{array}\right) \quad \text{(B.5.7)}$$

Then our test has size-control uniformly over the set of DGPs that satisfy (B.5.7). We formalize the statement as follows:

Theorem B.5.3.1. *Suppose* $\{\overline{\Lambda}_n\}_{n\geq 1}$ *satisfies* (B.5.3), (B.5.7) *and assumption 13. Then under the null, for both fixed and diverging instruments, with possibly more instruments than sample-size, we have exact size-control for the proposed tests, i.e.*

$$\liminf_{n\to\infty} \inf_{\overline{\lambda}_n\in\overline{\Lambda}_n} \mathbb{P}_{\overline{\lambda}_n}\left(\widehat{Q}^{\gamma_n^*}(\beta_0) > C_{\alpha,df,\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0))\right)$$
$$= \limsup_{n\to\infty} \sup_{\overline{\lambda}_n\in\overline{\Lambda}_n} \mathbb{P}_{\overline{\lambda}_n}\left(\widehat{Q}^{\gamma_n^*}(\beta_0) > C_{\alpha,df,\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0))\right) = \alpha$$

and

$$\liminf_{n\to\infty} \inf_{\overline{\lambda}_n\in\overline{\Lambda}_n} \lim_{B\to\infty} \mathbb{P}_{\overline{\lambda}_n}\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n}(\beta_0)) > C^{\gamma_n^*}_{\alpha,df_{BS}}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right)$$
$$= \limsup_{n\to\infty} \sup_{\overline{\lambda}_n\in\overline{\Lambda}_n} \lim_{B\to\infty} \mathbb{P}_{\overline{\lambda}_n}\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n^*}(\beta_0)) > C^{\gamma_n^*}_{\alpha,df_{BS}}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right) = \alpha$$

B.5.4 Asymptotic Power Properties under Rank Deficiency

The power-properties of our ridge-projection-based-tests are similar to Theorems 3–8. We first expound on the notion of identification parameter under rank-deficiency of instruments. Recall in section 2.4.2 we began by introducing the notion of identification parameter $\mathcal{G} := Q_{\Pi,\Pi}$. Under rank-deficiency of instruments, we have an analogous notion of identification parameter, namely $\mathcal{G} := \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n^*}\Pi_i\Pi_j}{\sqrt{r}}$. We say that we have strong identification if $\mathcal{G}\to\infty$ and weak identification otherwise.

Power Properties – Diverging Rank

We first discuss the asymptotic-power under diverging rank,[5] and consider three cases for some sequence $d_n \to 0$: (1) Strong identification and local alternative, where $d_n\mathcal{G} = \widetilde{\mathcal{G}}$ and $\Delta = \widetilde{\Delta} d_n^{1/2}$ for some fixed $\widetilde{\Delta}, \widetilde{\mathcal{G}} \in \mathbb{R}$; (2) Strong identification and fixed alternative, where $d_n\mathcal{G} = \widetilde{\mathcal{G}}$ and $\Delta = \widetilde{\Delta}$; (3) Weak identification and fixed alternative, where $\mathcal{G} = \widetilde{\mathcal{G}}$ and $\Delta = \widetilde{\Delta}$. We make the following assumption:

[5] This implies that the number of instruments diverge. We make no assumptions regarding the number of instruments; in particular we allow $K >> n$.

Assumption 14. *Suppose that* $\frac{p_n^{\gamma_n^*}}{r} = o(1)$ *and* $p_n^W := \max_i P_{ii}^W = o(1)$, *and* $d_W = O(r^{(1-\eta)/4})$ *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.*

Note that assumption 14 is very similar to assumption 6, the only difference is that we have replaced K with r, p_n by $p_n^{\gamma_n^*}$, and removed the requirement that $p_n \leq \delta < 1$ for some constant $\delta > 0$ (since this clearly wouldn't hold whenever $K >> n$). Under the usual conditions of $r = K < n$, by noting that for any $0 \leq \gamma_1 \leq \gamma_2$, we have $p_n^{\gamma_2} \leq p_n^{\gamma_1} \leq p_n$,[6] so that a sufficient condition for $\frac{p_n^{\gamma_n^*}}{r} = o(1)$ is given by $\frac{p_n}{K} = o(1)$. We only require $\frac{p_n^{\gamma_n^*}}{r} = o(1)$ instead of $\frac{p_n^{\gamma_n}}{r} = o(1)$ for some sequence of γ_n out of being conservative. Recall that γ_n^* is the maximum of the arguments that maximize $\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n}^2$, so that in essence, $\frac{p_n^{\gamma_n^*}}{r} = o(1)$ is the weakest requirement in the sense that it is possible for $\frac{p_n^{\gamma_1}}{r} \neq o(1)$ for some $\gamma_1 < \gamma_n^*$ with the property that γ_1 maximizes $\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n}^2$, yet we can still have $\frac{p_n^{\gamma_n^*}}{r} = o(1)$.

Similar to (B.4.1), under the arguments of Dovi et al. (2023)[Theorem 1], whenever assumption 5 and 14 holds, under both weak and strong identification, for $r \to \infty$ and any sequence of γ_n satisfying assumption 9, we have

$$\begin{pmatrix} \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n}\widetilde{e}_i\widetilde{e}_j}{\sqrt{r}} \\ \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n}\widetilde{X}_i\widetilde{e}_j}{\sqrt{r}} \\ \frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n}\widetilde{X}_i\widetilde{X}_j}{\sqrt{r}} - \mathcal{G} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1^\gamma(\beta) & \Phi_{12}^\gamma(\beta) & \Phi_{13}^\gamma(\beta) \\ \Phi_{12}^\gamma(\beta) & \Psi^\gamma(\beta) & \tau^\gamma(\beta) \\ \Phi_{13}^\gamma(\beta) & \tau^\gamma(\beta) & \Upsilon^\gamma(\beta) \end{pmatrix}\right) \tag{B.5.8}$$

for some $(\Phi_1^\gamma(\beta), \Phi_{12}^\gamma(\beta), \Phi_{13}^\gamma(\beta), \Psi^\gamma(\beta), \tau^\gamma(\beta), \Upsilon^\gamma(\beta))$ with β being the true parameter of interest.[7] We have the following power-properties, for which we omit the proof in order to avoid repetition; the proofs are exactly the same as Theorem 3–5, with an additional use of Lemma B.5.1.

Theorem B.5.4.1. *Suppose Assumption 5, 9, 14 and* (B.5.3) *holds, with* $r \to \infty$. *For any estimator* $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)$ *that satisfies* (B.5.4), *we have under strong identification and fixed*

[6]See the expression of $\widetilde{D}_{ii}$ at the start of section B.5.5

[7]Note that Dovi et al. (2023)[Theorem 1] proved the first of the three equations in (B.5.8), with $\Phi_1^\gamma(\beta) = \lim_{n\to\infty}\Phi_1^{\gamma_n}(\beta)$ for any sequence of γ_n satisfying assumption 9.

alternative

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n^*}(\beta_0)) > C^{\gamma_n^*}_{\alpha,df_{BS}}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right) = 1$$

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

Theorem B.5.4.2. *Suppose Assumption 5, 9, 14 and (B.5.3) holds, with $r \to \infty$. For any estimator $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0) \xrightarrow{p} \Phi_1^{\gamma}(\beta_0)$, we have under weak identification and fixed alternative that*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n^*}(\beta_0)) > C^{\gamma_n^*}_{\alpha,df_{BS}}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right) = 1 - F\left(q_{1-\alpha}(\mathcal{N}(0,1)) - \frac{\widetilde{\Delta}^2\widetilde{\mathcal{G}}}{\sqrt{\Phi_1(\beta_0)}}\right)$$

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 $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)$ can be taken as $\widehat{\Phi}_1^{\gamma_n^*,\ell}(\beta_0)$ for $\ell = \{standard, cf\}$ given in section B.5.1.

Under strong identification and local alternative, we have the following result:

Theorem B.5.4.3. *Suppose Assumption 5, 9, 14 and (B.5.3) holds, with $r \to \infty$. For any estimator $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)$ satisfying (B.5.4), under strong identification and local alternative we have*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}\left(\widehat{J}^{\gamma_n^*}(\beta_0,\widehat{\Phi}_1^{\gamma_n^*}(\beta_0))>C_{\alpha,df_{BS}}^{\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0),\mathcal{L})\right)=1-F\left(q_{1-\alpha}(\mathcal{N}(0,1))-\frac{\widetilde{\Delta}^2\widetilde{\mathcal{G}}}{\sqrt{\Phi_1(\beta_0)}}\right)$$

Power Properties – Fixed Rank

We discuss in this section the asymptotic-power when rank is fixed. In general, there are two further cases to consider under fixed rank: (i) K is fixed (ii) $K\to\infty$. In either case, for $K>r$, the implication is that there are $K-r>0$ linearly-dependent columns; these linearly-dependent columns provide no additional information, so that when the rank of instruments is taken to be fixed, we can assume without loss of generality that the number of instruments is fixed, specifically, $r=K$. In essence, the power-properties will be (almost) exactly the same as that described in section 2.4.2. The only difference is that we replace assumption 8 by the following assumption:

Assumption 15. *For every sequence of $\Delta_n\to\Delta^\dagger\in\mathbb{R}$, suppose $\frac{1}{n}\sum_{i\in[n]}\Lambda_{0,i}(\Delta_n)\otimes Z_iZ_i'\to\Sigma(\Delta^\dagger)$ and $\frac{Z'Z+\gamma_n^*I_K}{n}\to Q_{ZZ}$, where $\Sigma(\Delta^\dagger)$ is positive-semi-definite and Q_{ZZ} is positive-definite matrix. Furthermore, assume that $\sup_i||Z_i||_F<\infty$.*

By repeating the exact proof as in Theorem 6–8 and using Lemma B.5.1, we can obtain the following results, which we state without proof.

Theorem B.5.4.4. *Suppose Assumption 5, 9 14, 15, (B.5.3) holds and we are under fixed r. For any estimator $\widehat{\Phi}_1(\beta_0)$ that satisfies (B.5.4), our test consistently differentiates the null from alternative, i.e.*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}\left(\widehat{J}^{\gamma_n^*}(\beta_0,\widehat{\Phi}_1^{\gamma_n^*}(\beta_0))>C_{\alpha,df_{BS}}^{\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0),\mathcal{L})\right)=1$$

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, $\mu_K \equiv \widetilde{\mu}$,[8] while under strong identification, $d_n\mu_K \equiv \widetilde{\mu}$, where $\widetilde{\mu} \in \mathbb{R}^K$ is some constant. Denote

$$\Omega^*(\beta_0) := \lim_{n\to\infty} \frac{(Z'\Lambda(\beta_0)Z)^{1/2}(Z'Z + \gamma_n^* I_K)^{-1}(Z'\Lambda(\beta_0)Z)^{1/2}}{\sum_{i\in[n]} P_{ii,\gamma_n^*}\sigma_i^2(\beta_0)}$$

and assume it is well-defined. We have the following result:

Theorem B.5.4.5. *Suppose Assumption 5, 9 14, 15, (B.5.3) holds and we are under fixed r. Furthermore, let $\frac{p_n^{\gamma_n^*}\Pi'\Pi}{r} = O(1)$ and suppose $\Omega^*(\beta_0)$ is well-defined. Then under strong-identification and local alternative, for any estimator $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)$ that satisfies (B.5.4),*

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

and

$$\begin{aligned}\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}&\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n^*}(\beta_0)) > C_{\alpha,df_{BS}}^{\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right)\\ &= \mathbb{P}\left(\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right)'\Omega^*(\beta_0)\mathcal{Z}_K\left(\Sigma(0)\widetilde{\Delta}\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*})\right)\end{aligned}$$

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

Theorem B.5.4.6. *Suppose Assumption 5, 9 14, 15, (B.5.3) holds and we are under fixed r. Assume $\Omega^*(\beta_0)$ is well-defined and consider any estimator $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0) \xrightarrow{p} \Phi_1^{\gamma}(\beta_0)$. Then under weak-identification and fixed alternative, if we further assume that $\Pi'\Pi = O(1)$, we have*

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

and

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\widehat{J}^{\gamma_n^*}(\beta_0, \widehat{\Phi}_1^{\gamma_n^*}(\beta_0)) > C_{\alpha,df_{BS}}^{\gamma_n^*}(\widehat{\Phi}_1^{\gamma_n^*}(\beta_0), \mathcal{L})\right)$$

[8] 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.

$$= \mathbb{P}\left(\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right)' \Omega^*(\beta_0)\mathcal{Z}_K\left(\Sigma(\widetilde{\Delta})\widetilde{\mu}\right) > q_{1-\alpha}(F_{w^*})\right)$$

where w^* are the eigenvalues of $\Omega^*(\beta_0)$.In particular, if we assume $\Pi' M \Pi \leq \frac{\Pi'\Pi}{K} \to 0$, then $\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)$ can be taken as $\widehat{\Phi}_1^{\gamma_n^*,\ell}(\beta_0)$ for $\ell = \{standard, cf\}$ given in section B.5.1.

B.5.5 Proofs for section B.5

The proofs are analogous to what we have shown before in section 2.4. We require a technical lemma needed for the proofs later on, which is provided by Dovi et al. (2023). We begin by introducing some intuition. We can apply the singular-value-decomposition for our $n \times K$ matrix Z as follows:

$$Z = S\Sigma V'$$

where $S \in \mathbb{R}^{n\times n}$ is such that $S'S = SS' = I_n$, $V \in \mathbb{R}^{K\times K}$ is such that $V'V = VV' = I_K$, and $\Sigma \in \mathbb{R}^{n\times K}$ is such that it can be written as

$$\Sigma = \begin{pmatrix} D & 0_{r\times(K-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}$$

and $D \in \mathbb{R}^{r\times r}$ is a diagonal-matrix with elements $\{D_{ii}\}_{i\in[r]}$. we can then rewrite

$$P_{\gamma_n} = S\Sigma V'(V\Sigma'\Sigma V' + \gamma_n I_K)^{-1}V\Sigma' S' = S\Sigma(\Sigma'\Sigma + \gamma_n I_K)\Sigma' S' = S\widetilde{D}S'$$

where $\widetilde{D} = \Sigma(\Sigma'\Sigma + \gamma_n I_K)^{-1}\Sigma' \in \mathbb{R}^{n\times n}$ is a diagonal-matrix given by entries $\widetilde{D}_{ii} = \frac{D_{ii}^2}{D_{ii}^2+\gamma_n}$ for $i \in [r]$ and zero otherwise. Note that these diagonal entries of $\widetilde{D}$ are also the eigenvalues of P_{γ_n}. The only additional technical lemma needed for the proofs later on is given as follow:

Lemma B.5.1 (Dovi et al. (2023) Lemma 1)**.** *Fix $n \geq 3$. For all $i, j, m = 1, ..., n$ and $\gamma_n \geq 0$ if $r = K$ and $\gamma_n > 0$ for $r < K$, one has*

$$(i) \quad 0 \leq (P_{\gamma_n})_{ii}^{\ell} \leq P_{ii,\gamma_n} \text{ for all positive integers } \ell$$

$$(ii) \quad \sum_{i\in[n]}(P_{ij,\gamma_n})^2 = (P_{\gamma_n})_{jj}^2 \leq P_{jj,\gamma_n}$$

$$(iii) \quad \sum_{i\in[n]} P_{ii,\gamma_n} = \sum_{i\in[r]} \frac{D_{ii}^2}{D_{ii}^2+\gamma_n} \leq r$$

$$(iv) \quad |P_{ij,\gamma_n}| \leq 1$$

(v) *for any* $\mathcal{I}_2 \subset \{1,...,n\}^2$ *and* $\mathcal{I}_3 \subset \{1,...,n\}^3$,

$$(a) \quad \sum_{\mathcal{I}_2} (P_{ij,\gamma_n})^4 \leq r,$$

$$(b) \quad \sum_{\mathcal{I}_3} (P_{ij,\gamma_n})^2 (P_{jm,\gamma_n})^2 \leq r$$

Lemma B.5.1 shows that the ridge-projection matrix has similar properties to the usual projection. Therefore many of the proofs can be repeated with appropriate replacement (i.e. replace K and P with r and P_{γ_n} respectively).

Proof of Theorem B.5.3.1: Note that $\beta_0 = \beta$ since we are under the null. We separate our proof into two cases: (i) r is fixed and (ii) $r \to \infty$. The fixed r case follows in exactly the same way as the proof of Theorem 2 - Fixed K case. In particular, we can show that

$$\widehat{Q}_{n_{j_k}}^{\gamma_n^*}(\beta_0) \rightsquigarrow \sum_{i\in[r]} w_i^* \chi_{1,i}^2$$

where $w^* := (w_1^*, ..., w_r^*)'$ is the limit of $w^{\gamma_n^*}$, where $w^{\gamma_n^*}$ is the eigenvalues of

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

Furthermore, we can show that $F_{\widetilde{w}_{n_{j_k}}^{\gamma_{n_{j_k}}^*}} \rightsquigarrow F_{w^*}$. Finally we can show that

$$\frac{\dfrac{\sqrt{\widehat{\Phi}_1^{\gamma_n^*}}}{\frac{1}{\sqrt{r}}\sum_{i\in[n]} P_{ii,\gamma_n^*} e_i^2}}{\sqrt{2\sum_{i\in[r]}(\widetilde{w}_{i,n}^{\gamma_n^*})^2 + 1/df}} \xrightarrow{p} \frac{\sqrt{2}||w^*||}{\sqrt{2}||w^*||} = 1$$

This concludes the proof for the fixed r case. The diverging r case follows in exactly the same way as the proof of Theorem 2 - Diverging K case. In particular, we can show

$$\frac{\frac{1}{\sqrt{r}}\sum_{i\in[n]} P_{ii,\gamma_n^*} e_i^2}{\sqrt{\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)}}\left(\widehat{Q}^{\gamma_n^*}(\beta_0)-1\right)=\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n^*} e_i e_j}{\sqrt{r\widehat{\Phi}_1^{\gamma_n^*}(\beta_0)}}\rightsquigarrow \mathcal{N}(0,1) \tag{B.5.9}$$

and

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

To see (B.5.9), note that (B.5.7) implies assumption 5, 9 and 14, which in turn implies (B.5.8). An analogous proof to Lim et al. (2024)[Theorem A.1.] yields

$$\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n^*} e_i e_j}{\sqrt{r}}=\frac{\sum_{i\in[n]}\sum_{j\neq i} P_{ij,\gamma_n^*} \widetilde{e}_i\widetilde{e}_j}{\sqrt{r}}+o_p(1),$$

so that combining with (B.5.8) completes the proof for the diverging r case.

Bibliography

Anatolyev, S. and N. Gospodinov (2011). Specification testing in models with many instruments. *Econometric Theory 27*(2), 427–441.

Anatolyev, S. and A. Mikusheva (2022). Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics 229*(1), 103–126.

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

Anderson, T., N. Kunitomo, and Y. Matsushita (2010). On the asymptotic optimality of the liml estimator with possibly many instruments. *Journal of Econometrics 157*(2), 191–204.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

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

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

Andrews, D. W. and P. Guggenberger (2019). Identification-and singularity-robust inference for moment condition models. *Quantitative Economics 10*(4), 1703–1746.

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.

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica 84*(6), 2155–2182.

Andrews, I. (2018). Valid two-step identification-robust confidence sets for GMM. *Review of Economics and Statistics 100*(2), 337–348.

Andrews, I. and A. Mikusheva (2016). Conditional inference with a functional nuisance parameter. *Econometrica 84*(4), 1571–1612.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*(1), 727–753.

Angrist, J. and B. Frandsen (2022). Machine labor. *Journal of Labor Economics 40*(S1), S97–S140.

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.

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

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

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica 62*(3), 657–681.

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

Belloni, A., V. Chernozhukov, and C. Hansen (2011). Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*.

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

Boot, T. and J. W. Ligtenberg (2023). Identification- and many instrument-robust inference via invariant moment conditions. *arXiv:2303.07822*.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*(2), 383–398.

Carrasco, M. and G. Tchuente (2015). Regularized liml for many instruments. *Journal of Econometrics 186*(2), 427–442.

Chamberlain, G. and G. Imbens (2004). Random effects estimators with many instrumental variables. *Econometrica 72*(1), 295–306.

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 (2023a). Jackknife estimation of a cluster-sample iv regression model with many weak instruments. *Journal of Econometrics*.

Chao, J. C., N. R. Swanson, and T. Woutersen (2023b). 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.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–40.

Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. *Econometrica 69*(5), 1161–1191.

Dovi, M.-S., A. B. Kock, and S. Mavroeidis (2023). A ridge-regularised jackknifed anderson-rubin test. *arXiv:2209.03259*.

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

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

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.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica 45*(4), 939–953.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Han, C. and P. C. Phillips (2006). Gmm with many moment conditions. *Econometrica 74*(1), 147–192.

Hansen, C., J. Hausman, and W. Newey (2008). Estimation with many instrumental variables. *Journal of Business and Economic Statistics 26*(4), 398–422.

Hansen, C. and D. Kozbur (2014). Instrumental variables estimation with many weak instruments using regularized jive. *Journal of Econometrics 182*(2), 290–308.

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.

Kaffo, M. and W. Wang (2017). On bootstrap validity for specification testing with many weak instruments. *Economics Letters 157*, 107–111.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

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

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

Kolesár, M. (2018). Minimum distance approach to inference with many instruments. *Journal of Econometrics 204*(1), 86–100.

Kolesar, M., U. K. Muller, and S. T. Roelsgaard (2023). The fragility of sparsity. *arXiv:2311.02299*.

Koziol, J. A. and M. D. Perlman (1978). Combining independent chi-squared tests. *Journal of the American Statistical Association 73*(364), 753–763.

Kubokawa, T., C. P. Robert, and A. M. E. Saleh (1993). Estimation of noncentrality parameters. *The Canadian Journal of Statistics 21*(1), 45–57.

Kuersteiner, G. and R. Okui (2010). Constructing optimal instruments by first-stage prediction averaging. *Econometrica 78*(2), 697–718.

Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association 75*(371), 693–700.

Lee, D. S., J. McCrary, M. J. Moreira, and J. R. Porter (2022). Valid t-ratio inference for iv. *American Economic Review forthcoming.*

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

Lim, D., W. Wang, and Y. Zhang (2024). A conditional linear combination test with many weak instruments. *Journal of Econometrics 238*(2), 105602.

Maestas, N., K. J. Mullen, and A. Strand (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review 103*(5), 1797–1829.

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

Matsushita, Y. and T. Otsu (2020). *Jackknife Lagrange multiplier test with many weak instruments*. LSE, STICERD.

Matsushita, Y. and T. Otsu (2021). Jackknife empirical likelihood: small bandwidth, sparse network and high-dimensional asymptotics. *Biometrika 108*(3), 661–674.

Maurice J. G. Bun, H. F. and R. W. Poldermans (2020). Finite sample properties of the gmm anderson-rubin test. *Econometric Reviews 39*(10), 1042–1056.

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

Monti, K. L. and P. K. Sen (1976). The locally optimal combination of independent test statistics. *Journal of the American Statistical Association 71*(356), 903–911.

Moreira, H. and M. J. Moreira (2019). Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors. *Journal of Econometrics 213*(2), 398–433.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Morimune, K. (1983). Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica 51*(3), 821–841.

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

Okui, R. (2011). Instrumental variable estimation in the presence of many moment conditions. *Journal of Econometrics 165*(1), 70–86.

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.

Sampat, B. and H. L. Williams (2019). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

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.

Shanken, J. (1992). On the estimation of beta-pricing models. *The Review of Financial Studies 5*(1), 1–33.

Sølvsten, M. (2020). Robust estimation with many instruments. *Journal of Econometrics 214*(2), 495–512.

Stock, J. and M. Yogo (2005a). Asymptotic distributions of instrumental variables statistics with many instruments. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg 6*, 109–120.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H. and M. Yogo (2005b). Testing for weak instruments in linear IV regression. In D. W. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Chapter 6, pp. 80–108. Cambridge, U.K.: Cambridge University Press.

Vaart, A. W. V. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

Wang, W. and M. Kaffo (2016). Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics 192*(1), 231–268.