1-2024

A conditional linear combination test with many weak instruments

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

Wenjie WANG
Nanyang Technological University

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Citation

LIM, Dennis; WANG, Wenjie; and ZHANG, Yichong. A conditional linear combination test with many weak instruments. (2024). *Journal of Econometrics*. 238, (2), 1-20.
Available at: https://ink.library.smu.edu.sg/soe_research/2618

A Conditional Linear Combination Test with Many Weak Instruments*

Dennis Lim† Wenjie Wang‡ Yichong Zhang§

Abstract

We consider 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), we choose the weights in the linear combination based on a decision-theoretic rule that is adaptive to the identification strength. Under both weak and strong identifications, the proposed test controls asymptotic size and is admissible among certain class of tests. Under strong identification, our linear combination test has optimal power against local alternatives among the class of invariant or unbiased tests which are constructed based on jackknife AR and LM tests. Simulations and an empirical application to Angrist and Krueger's (1991) dataset confirm the good power properties of our test.

Keywords: Many instruments, power, size, weak identification

JEL codes: C12, C36, C55

*We thank the editor, the associate editor, and two referees for extensive comments which have led to many improvements. We are grateful to Andrii Babii, Harold D. Chiang, Keisuke Hirano, Ming Li, Adam Rosen, and all participants at the SH3 conference on econometrics 2023 and the University of Macau Econometrics Seminar for their valuable comments. Wenjie Wang acknowledges the financial support from Singapore Ministry of Education Tier 1 grants RG53/20 and RG104/21. Yichong Zhang acknowledges the financial support from a Lee Kong Chian fellowship and the NSFC under the grant No. 72133002. Any and all errors are our own.

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

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

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

1 Introduction

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 some 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. 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: Power Comparision 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,

[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.

Hausman, Newey, and Woutersen (2012), Hausman, Newey, Woutersen, Chao, and Swanson (2012), Hansen and Kozbur (2014), Carrasco and Tchuente (2015), Wang and Kaffo (2016), Kolesár (2018), Matsushita and Otsu (2022), Sølvsten (2020), Crudu, Mellace, and 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 ratio of the concentration parameter over the square root of the number of instruments 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), 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

[2] Specifically, the rate of growth of the concentration parameter, which measure the overal 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).

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 (2022) 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, $U \overset{d}{=} V$ 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 $U_n \overset{d}{=} V_n + o_P(1)$ if there exist $\tilde{U}_n \overset{d}{=} U_n$ and $\tilde{V}_n \overset{d}{=} V_n$ such that $\tilde{U}_n - \tilde{V}_n = o_P(1)$.

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

$$Y_i = X_i\beta + e_i, \quad X_i = \Pi_i + V_i, \quad \forall i \in [n], \tag{2.1}$$

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 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 (2022)) 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:*

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix} \right), \tag{2.2}$$

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

Although there are no controls in the model (2.1), we further verify Assumption 1 in Section A of the Online Supplement 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,

$$\begin{pmatrix} Q_{e(\beta_0),e(\beta_0)} - \Delta^2 \mathcal{C} \\ Q_{X,e(\beta_0)} - \Delta \mathcal{C} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \overset{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) & \Phi_{13}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) & \tau(\beta_0) \\ \Phi_{13}(\beta_0) & \tau(\beta_0) & \Upsilon \end{pmatrix} \right) + o_p(1), \tag{2.3}$$

[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) ($\mu_n^2/\sqrt{K} \to \infty$).

[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 (2023) 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 $\begin{pmatrix} Q_{e(\beta_0),e(\beta_0)} \\ Q_{X,e(\beta_0)} \\ Q_{X,X} \end{pmatrix} = \begin{pmatrix} 1 & 2\Delta & \Delta^2 \\ 0 & 1 & \Delta \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} \end{pmatrix}$.

where

$$\begin{aligned}
\Phi_1(\beta_0) &= \Delta^4\Upsilon + 4\Delta^3\tau + \Delta^2(4\Psi + 2\Phi_{13}) + 4\Delta\Phi_{12} + \Phi_1, \\
\Phi_{12}(\beta_0) &= \Delta^3\Upsilon + 3\Delta^2\tau + \Delta(2\Psi + \Phi_{13}) + \Phi_{12}, \\
\Phi_{13}(\beta_0) &= \Delta^2\Upsilon + 2\Delta\tau + \Phi_{13}, \\
\Psi(\beta_0) &= \Delta^2\Upsilon + 2\Delta\tau + \Psi, \\
\tau(\beta_0) &= \Delta\Upsilon + \tau.
\end{aligned} \tag{2.4}$$

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 B.1 in the Online Supplement. 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

[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 $\widehat{\Psi}(\beta_0) \xrightarrow{p} \Psi(\beta_0)$.

more details in Section B.2 in the Online Supplement. 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, in order for our jackknife CLC test proposed below to control size under both weak and strong identification, it suffices to require $\widehat{\gamma}(\beta_0)$ to be consistent under the null only. The power analyses in Lemmas 2.1 and 2.4 below, and subsequently, Theorems 4.1 and 4.2, only require the consistency of $\widehat{\gamma}(\beta_0)$ under strong identification with local alternatives and weak identification with fixed alternatives, respectively.

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

$$1\{AR(\beta_0) \geq z_\alpha\}, \quad AR(\beta_0) = \frac{Q_{e(\beta_0),e(\beta_0)}}{\widehat{\Phi}_1^{1/2}(\beta_0)}, \tag{2.5}$$

and Matsushita and Otsu (2022) consider the jackknife LM test

$$1\{LM^2(\beta_0) \geq \mathbb{C}_\alpha\}, \quad LM(\beta_0) = \frac{Q_{X,e(\beta_0)}}{\widehat{\Psi}^{1/2}(\beta_0)}. \tag{2.6}$$

Both tests are robust to weak identification, many instruments, and heteroskedasticity. Lemma 2.1 below characterizes the joint limit distribution of $(AR(\beta_0), LM(\beta_0))^\top$ under strong identification and local alternatives.

Lemma 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 $\mathcal{C} = \widetilde{\mathcal{C}}/d_n$ 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 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 2.1 implies, under strong identification and local alternatives,

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{[(1-\rho^2)\Psi]^{1/2}} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right). \tag{2.7}$$

Lemma 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 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 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 $\Delta \neq 0$,*

$$\begin{pmatrix} AR(\beta_0) \\ LM^*(\beta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{2.8}$$

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

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

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 (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 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 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 (2.7) and

[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.

(2.8) for their expressions in these two cases. Then, in the limit experiment, the linear combination test can be written as

$$\phi_{a_1,a_2,\infty} = 1\{a_1\mathcal{N}_1^2 + a_2(\tilde{\rho}\mathcal{N}_1 + (1-\tilde{\rho}^2)^{1/2}\mathcal{N}_2^*)^2 + (1-a_1-a_2)\mathcal{N}_2^{*2} \geq \mathbb{C}_\alpha(a_1,a_2;\tilde{\rho})\}, \tag{2.9}$$

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 2.1 and 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

$$\begin{pmatrix} a_1 + a_2\tilde{\rho}^2 & a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} \\ a_2\tilde{\rho}(1-\tilde{\rho}^2)^{1/2} & 1-a_1-a_2\tilde{\rho}^2 \end{pmatrix} = \mathcal{U}\begin{pmatrix} \nu_1(a_1,a_2) & 0 \\ 0 & \nu_2(a_1,a_2) \end{pmatrix}\mathcal{U}^\top \tag{2.10}$$

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

$$\begin{pmatrix} \widetilde{\mathcal{N}}_1 \\ \widetilde{\mathcal{N}}_2 \end{pmatrix} = \mathcal{U}^\top \begin{pmatrix} \mathcal{N}_1 \\ \mathcal{N}_2^* \end{pmatrix} \tag{2.11}$$

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

Theorem 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 $\theta_2 = m_2(\Delta)$ as in (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 (2.11). Then, for any $(a_1,a_2) \in \mathbb{A}_0$, $\phi_{a_1,a_2,\infty}$ defined in (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 $\mathbb{E}\phi \leq \mathbb{E}\phi_{a_1,a_2,\infty}$.*

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

(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 (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 2.4 is nonnegative, the one-sided test has good power. However, even in this case, the power curves simulation in Section 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 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 2.1. Fourth, under strong identification and fixed alternatives, our jackknife CLC test has asymptotic power equal to 1, as shown in Lemma 2.3 and Theorem 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 5.

Second, Theorem 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,

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

[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 $\sum_{i\in[n]} P_{ii}\Pi_i^2$ 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 $\mathcal{C} > 0$.

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 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 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 (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 other 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

$$a\mathcal{N}_1^{\dagger 2} + (1-a)\mathcal{N}_2^{\dagger 2}, \tag{2.12}$$

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 (2.12). Under strong identification and local alternatives, Lemma 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 new parametrization does not bring any additional power.

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

$$\phi_{a_1,a_2,\infty} = 1\left\{ \begin{array}{c} a_1 \mathcal{Z}_1^2(m_1(\Delta)) + a_2(\rho(\beta_0)\mathcal{Z}_1(m_1(\Delta)) + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2(m_2(\Delta)))^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(m_2(\Delta)) \geq \mathbb{C}_\alpha(a_1, a_2; \rho(\beta_0)) \end{array} \right\}, \tag{3.1}$$

where $m_1(\Delta)$ and $m_2(\Delta)$ are defined in Lemma 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 (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 2.4 can be written as

$$\begin{pmatrix} m_1(\Delta) \\ m_2(\Delta) \end{pmatrix} = \begin{pmatrix} C_1(\Delta) \\ C_2(\Delta) \end{pmatrix} \mu_D, \tag{3.2}$$

where

$$\begin{pmatrix} C_1(\Delta) \\ C_2(\Delta) \end{pmatrix} \equiv \begin{pmatrix} \Phi_1^{-1/2}(\beta_0)\Delta^2 \\ (1-\rho^2(\beta_0))^{-1/2}(\Psi^{-1/2}(\beta_0)\Delta - \rho(\beta_0)\Phi_1^{-1/2}(\beta_0)\Delta^2) \end{pmatrix}$$

$$\times \left[1 - (\Delta^2, \Delta) \left(\begin{pmatrix} \Phi_1(\beta_0) & \Phi_{12}(\beta_0) \\ \Phi_{12}(\beta_0) & \Psi(\beta_0) \end{pmatrix}^{-1} \begin{pmatrix} \Phi_{13}(\beta_0) \\ \tau(\beta_0) \end{pmatrix} \right) \right]^{-1}. \tag{3.3}$$

By (3.2), we see that $\phi_{a_1,a_2,\infty} = \phi_{a_1,a_2,\infty}(\Delta)$ defined in (2.9) can be written as

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

This motivates the definition that

$$\phi_{a_1,a_2,\infty}(\delta) = 1 \left\{ \begin{matrix} a_1 \mathcal{Z}_1^2(C_1(\delta)\mu_D) + a_2(\rho(\beta_0)\mathcal{Z}_1(C_1(\delta)\mu_D) + (1-\rho^2(\beta_0))^{1/2}\mathcal{Z}_2(C_2(\delta)\mu_D))^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(C_2(\delta)\mu_D) \geq \mathbb{C}_\alpha(a_1, a_2; \rho(\beta_0)) \end{matrix} \right\}. \tag{3.4}$$

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 5.2 and empirical application in Section 6. Specifying the parameter space is almost inevitable for any weak-identification-robust inference method, but additional simulation results in Section U of the Online Supplement 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 $\overline{a} < 1$,

$$\underline{a}(\mu_D, \gamma(\beta_0)) = \min\left(p_1, \frac{p_2 \mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}\right), \tag{3.5}$$

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 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,

$$(a_1(\mu_D, \gamma(\beta_0)), a_2(\mu_D, \gamma(\beta_0))) \in \underset{(a_1,a_2) \in \mathbb{A}(\mu_D, \gamma(\beta_0))}{\arg\min} \sup_{\delta \in \mathcal{D}(\beta_0)} (\mathcal{P}_{\delta,\mu_D} - \mathbb{E}\phi_{a_1,a_2,\infty}(\delta, \mu_D, \gamma(\beta_0))). \quad (3.6)$$

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 2.1(iii). Specifically, we require $p_1 \in (0, 1)$ and close to 0 and $p_2 > 1$. In the Online Supplement, we provide a detailed report on the finite sample performance of our CLC test for both simulation designs analyzed in Section 5 and the empirical application in Section 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 $\frac{1.1\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}$ may be larger than 0.01. In this case, we have $\mathbb{A}(\mu_D, \gamma(\beta_0)) = \{(a_1, a_2) \in \mathbb{A}_0, a_1 \in [0.01, 1]\}$. Third, under strong identification and local alternatives, $\frac{1.1\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}$ will converge to zero so that

$$\mathbb{A}(\mu_D, \gamma(\beta_0)) = \left\{ (a_1, a_2) \in \mathbb{A}_0, a_1 \in \left[\frac{1.1\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)c_{\mathcal{B}}(\beta_0)}{\Delta_*^4(\beta_0)\mu_D^2}, 1 \right] \right\}.$$

We show in Theorem 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 2.2, which is the UMP invariant and unbiased test. Furthermore, the minimax a_1 satisfies the requirement in Theorem 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 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

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

[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.

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 $\widehat{D}^2/\widehat{\sigma}_D^2 = D^2/\sigma_D^2 + o_p(1) \stackrel{d}{=} \mathcal{Z}^2(\mu_D/\sigma_D) + o_p(1)$ 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

$$\begin{aligned}
&\phi_{a_1,a_2,s}(\delta,\widehat{D},\widehat{\gamma}(\beta_0)) \\
&= 1\left\{\begin{array}{c} a_1\mathcal{Z}_1^2(\widehat{C}_1(\delta)f_s(\widehat{D},\widehat{\gamma}(\beta_0))) \\ +a_2\left[\widehat{\rho}(\beta_0)\mathcal{Z}_1(\widehat{C}_1(\delta)f_s(\widehat{D},\widehat{\gamma}(\beta_0))) + (1-\widehat{\rho}^2(\beta_0))^{1/2}\mathcal{Z}_2(\widehat{C}_2(\delta)f_s(\widehat{D},\widehat{\gamma}(\beta_0)))\right]^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2(\widehat{C}_2(\delta)f_s(\widehat{D},\widehat{\gamma}(\beta_0))) \geq \mathbb{C}_\alpha(a_1,a_2;\widehat{\rho}(\beta_0)) \end{array}\right\},
\end{aligned} \tag{3.7}$$

and $(\widehat{C}_1(\delta),\widehat{C}_2(\delta))$ are similarly defined as $(C_1(\delta),C_2(\delta))$ in (3.3) with $\gamma(\beta_0)$ replaced by $\widehat{\gamma}(\beta_0)$; that is,

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

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 (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

$$\mathcal{A}_s(\widehat{D},\widehat{\gamma}(\beta_0)) \in \mathop{\arg\min}_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))}\sup_{\delta\in\mathcal{D}(\beta_0)}(\mathcal{P}_{\delta,s}(\widehat{D},\widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(\delta,\widehat{D},\widehat{\gamma}(\beta_0))), \tag{3.8}$$

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

$$\widehat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} = 1\left\{\begin{array}{c} \mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0))AR^2(\beta_0) + \mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0))LM^2(\beta_0) \\ +(1 - \mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) - \mathcal{A}_{2,s}(\widehat{D}, \widehat{\gamma}(\beta_0)))LM^{*2}(\beta_0) \geq \mathbb{C}_\alpha(\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0)); \widehat{\rho}(\beta_0)) \end{array}\right\}. \tag{3.9}$$

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 (3.6) and (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 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* (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* (3.6) *with* μ_D *replaced by* $f_s(D,\gamma(\beta_0))$ *for* $l = 1, 2$.

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 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 $f_{krs}(D,\gamma(\beta_0))$. Second, Theorem 4.1 implies that our jackknife CLC test controls size asymptotically conditionally on $\widehat{D}$, and thus, unconditionally. Last, according to Theorem 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 (3.4), the two optimality results stated in Theorem 2.1(i) also hold asymptotically, conditional on $\widehat{D}$. To make this statement precise, we define the eigenvalue decomposition

$$\begin{aligned}&\begin{pmatrix}\mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) + \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}^2(\beta_0) & \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}(\beta_0)(1-\hat{\rho}^2(\beta_0))^{1/2} \\ \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}(\beta_0)(1-\hat{\rho}^2(\beta_0))^{1/2} & 1-\mathcal{A}_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) - \mathcal{A}_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\hat{\rho}^2(\beta_0)\end{pmatrix} \\ &= \mathcal{U}_s(\widehat{D},\widehat{\gamma}(\beta_0))\begin{pmatrix}\nu_{1,s}(\widehat{D},\widehat{\gamma}(\beta_0)) & 0 \\ 0 & \nu_{2,s}(\widehat{D},\widehat{\gamma}(\beta_0))\end{pmatrix}\mathcal{U}_s(\widehat{D},\widehat{\gamma}(\beta_0))^\top. \end{aligned} \tag{4.1}$$

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

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 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 $(\theta_1^2, \theta_2^2) = b \cdot (\nu_1(d, \gamma(\beta_0)), \nu_2(d, \gamma(\beta_0)))$ 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 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 (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

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

and

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

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

Theorem 4.2. *Suppose that Assumptions 1 and 2 hold. Further suppose that we are under strong identification and local alternatives as described in Lemma 2.1. Then, for $s \in \{pp, krs\}$, we have*

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

where $\mathcal{N}_2^* \stackrel{d}{=} \mathcal{N}\left(\frac{\widetilde{\Delta}\widetilde{\mathcal{C}}}{[(1-\rho^2)\Psi]^{1/2}}, 1\right)$. In addition, suppose $\breve{\phi}_n = \phi(AR(\beta_0), LM(\beta_0)) + o_P(1)$ 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 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 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 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 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 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 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 T in the Online Supplement, 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 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 4.1 and 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]

$$\left(\begin{array}{c} \{V_i, e_i\}_{i\in[n]} \text{ are independent, } \mathbb{E}e_i = \mathbb{E}V_i = 0, \\ \max_i \mathbb{E}e_i^4 + \max_i \mathbb{E}V_i^4 \leq C_1 < \infty, \\ \mathcal{C}_n = \frac{1}{\sqrt{K}}\sum_{i\in[n]}\sum_{j\neq i}\Pi_i P_{ij}\Pi_j \in \Re, \\ P_{ii} \leq C_2 < 1, \\ 0 < \kappa_1 \leq \text{mineig}(\mathbb{V}_n) \leq \text{maxeig}(\mathbb{V}_n) \leq \kappa_2 < \infty, \\ \text{where } C_1,\ C_2,\ \kappa_1, \text{ and } \kappa_2 \text{ are some fixed constants,} \\ \text{and Assumption 2 holds for } \beta_0 = \beta. \end{array}\right) \tag{4.2}$$

In Sections B.1 and B.2 of the Online Supplement, 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 4.3 shows that our jackknife CLC test has correct asymptotic size, under similar arguments as those in Andrews, Cheng, and Guggenberger (2020) and I.Andrews (2016).

Theorem 4.3. *Suppose Assumption 3 holds, $\{\Lambda_n\}_{n\geq 1}$ satisfies (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 (3.9) has asymptotic power 1 against fixed alternatives.

Theorem 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 (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 of the Online Supplement that the joint asymptotic normality (Assumption 1) holds in the case with exogenous controls.

5 Simulation

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 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 (2.5), (iii) the jackknife LM test defined in (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 (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 (2.4) as β ranges over $\mathcal{B}$ and generate $AR(\beta_0)$ and $LM(\beta_0)$ based on (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 1–4 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 U.1 of the Online Supplement, we further report power curves for alternative values of the tuning parameters (p_1, p_2) in (3.5) and of $\mathcal{C}$, and find that the overall patterns remain very similar.

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

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

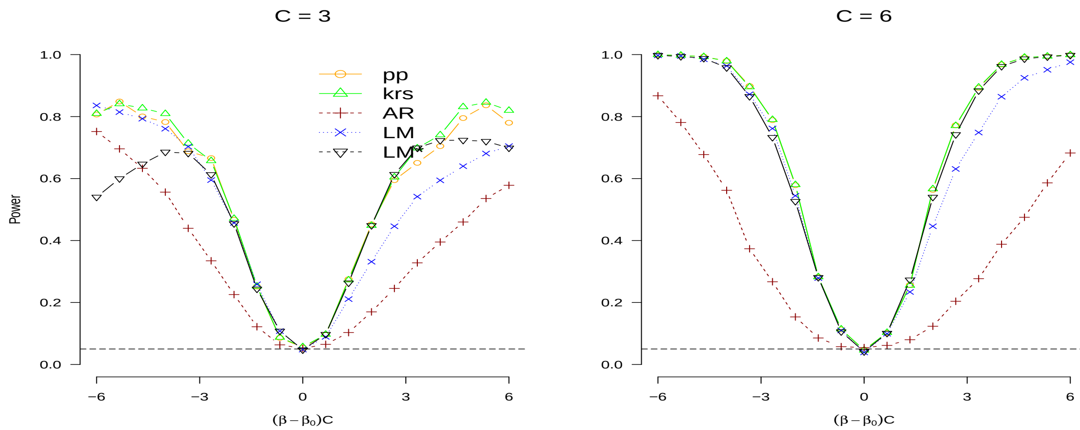

Figure 1: Power Curve for $\rho = 0.2$

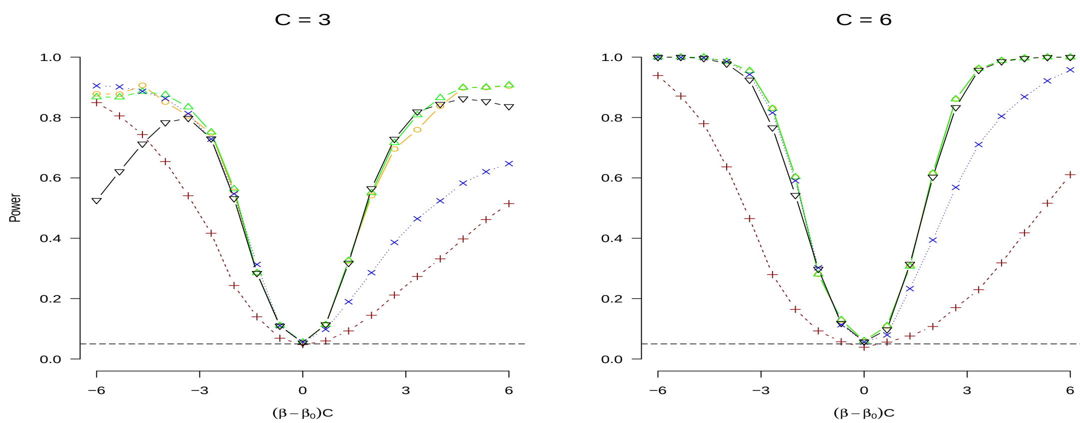

Figure 2: Power Curve for $\rho = 0.4$

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 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:

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

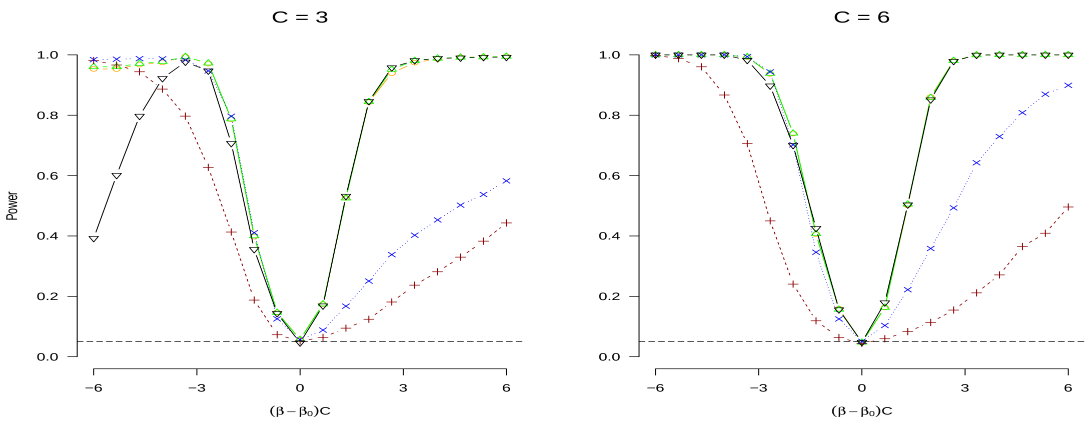

Figure 3: Power Curve for $\rho = 0.7$

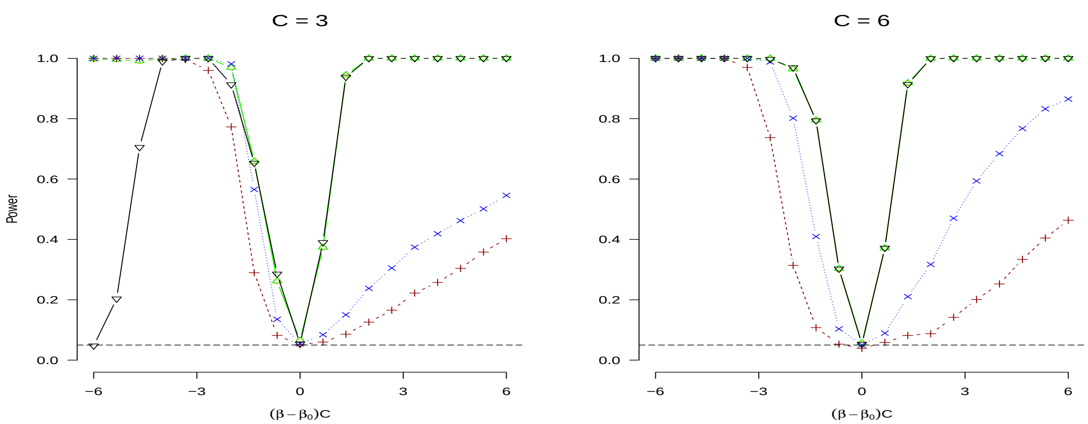

Figure 4: Power Curve for $\rho = 0.9$

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

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:

$$\widetilde{y}_i = \bar{y} + \beta\widetilde{s}_i + \omega(Q_i, C_i, P_i)(\nu_i + \kappa_2\xi_i) \tag{5.1}$$

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

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 (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 C in the Online Supplement. We set $(p_1, p_2) = (0.01, 1.1)$ in (3.5). Additional simulation results using other choices of (p_1, p_2) and $\mathcal{B}$ are reported in Section U.2 in the Online Supplement. 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 B.2 in the Online Supplement).

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.

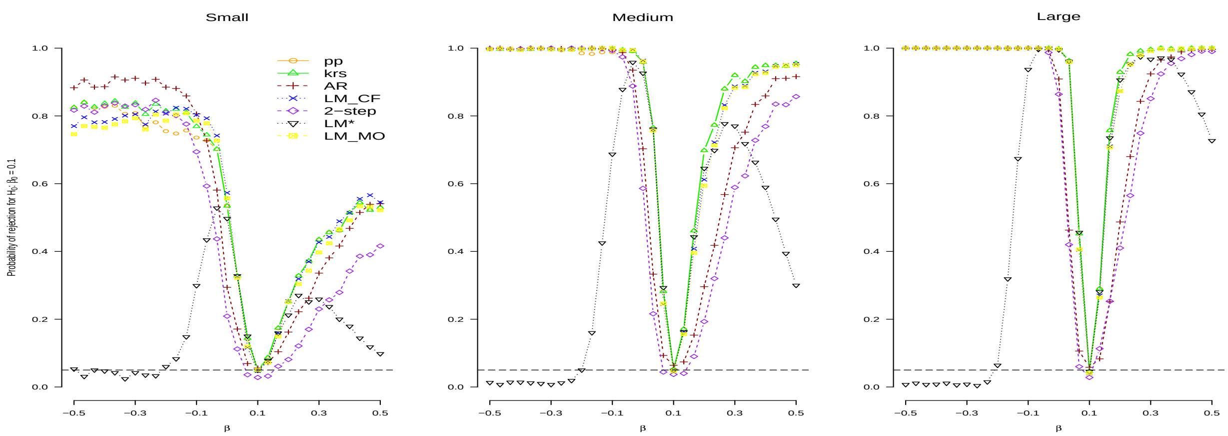

Figure 5: Power Curve for DGP 1

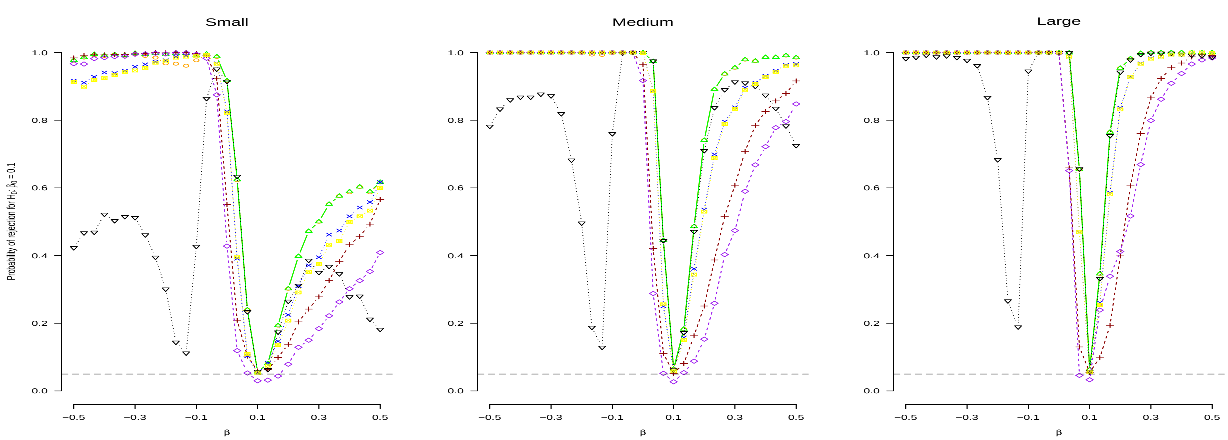

Figure 6: Power Curve for DGP 2

Figures 5 and 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 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 7 and 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, 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 5 and 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.

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

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

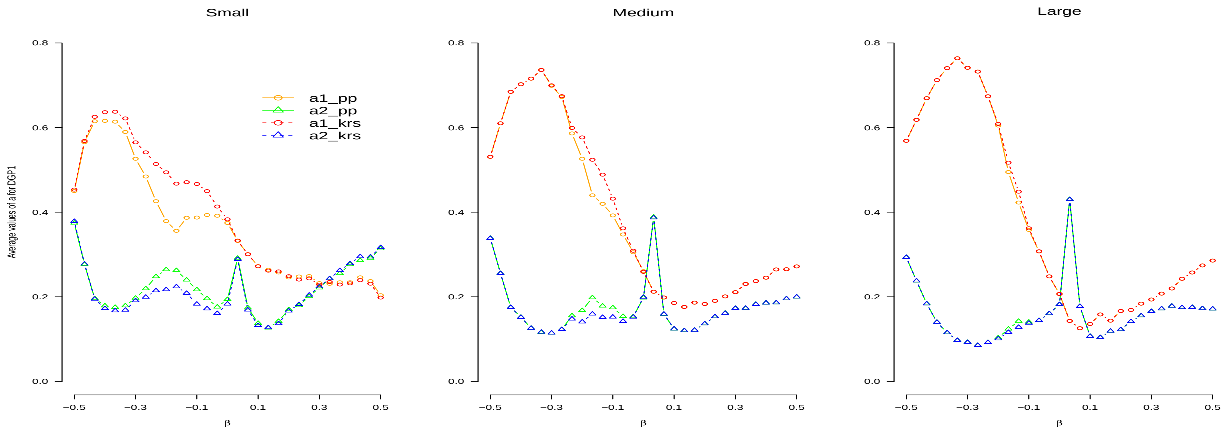

Figure 7: Average Values of a for DGP 1

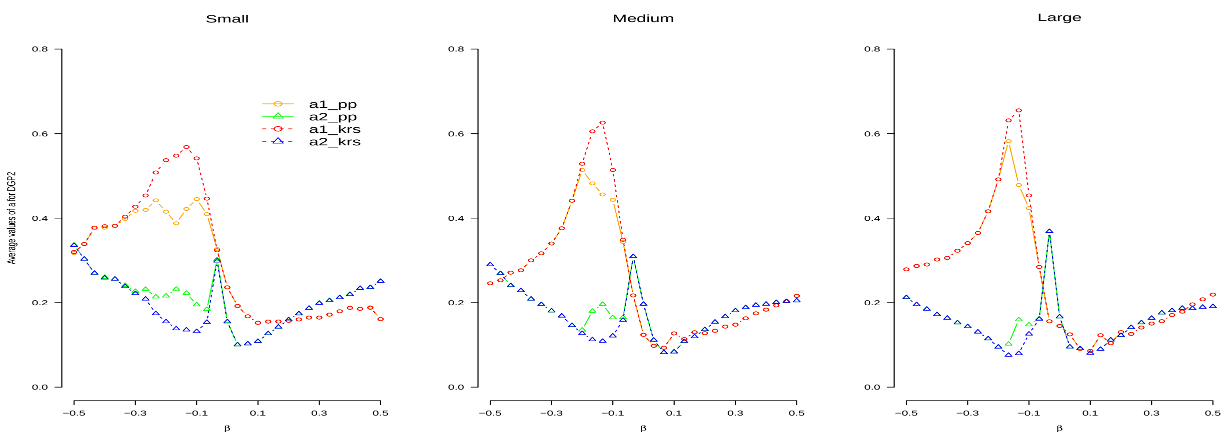

Figure 8: Average Values of a for DGP 2

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 D in the Online Supplement. The considered tests are similar to those in the previous section. The jackknife AR test is defined in (2.5) with $\widehat{\Phi}_1$ being the cross-fit estimator in Mikusheva and Sun (2022). The jackknife LM test is defined in (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 $\widetilde{F} > 9.98$ 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

[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.

(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 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 (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 V in the Online Supplement. 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 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 D in the Online Supplement.

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

A 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. *Suppose $\{\tilde{V}_i, \tilde{e}_i\}_{i\in[n]}$ are independent,* $\max_i \mathbb{E}\tilde{e}_i^4 + \max_i \mathbb{E}\tilde{V}_i^4 \le C < \infty$, $\max_i ||W_i||_2 \le C < \infty$, $\Pi^\top\Pi/K = O(1)$, *and* $0 < c \le mineig(W^\top W/n) \le maxeig(W^\top W/n) \le C < \infty$, *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 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 $(Q_{e,e}, Q_{X,e}, Q_{X,X} - Q_{\Pi,\Pi})$.

B Verifying Assumption 2

B.1 Standard Estimators

In this section, we maintain Assumption 5, which is stated below and just Mikusheva and Sun (2022, Assumption 1).

Assumption 5. *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:

$$\begin{pmatrix} Q_{e,e} \\ Q_{X,e} \\ Q_{X,X} - \mathcal{C} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{12} & \Psi & \tau \\ \Phi_{13} & \tau & \Upsilon \end{pmatrix}\right), \tag{B.1}$$

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

$$\begin{aligned}
\Phi_1 &= \lim_{n\to\infty} \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \sigma_i^2 \sigma_j^2, \\
\Phi_{12} &= \lim_{n\to\infty} \frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\gamma_j \sigma_i^2 + \gamma_i \sigma_j^2), \\
\Phi_{13} &= \lim_{n\to\infty} \frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \gamma_i \gamma_j, \\
\Psi &= \lim_{n\to\infty} \left[\frac{1}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 (\eta_i^2 \sigma_j^2 + \gamma_i \gamma_j) + \frac{1}{K} \sum_{i\in[n]} \omega_i^2 \sigma_i^2 \right], \\
\tau &= \lim_{n\to\infty} \left[\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \eta_i^2 \gamma_j + \frac{2}{K} \sum_{i\in[n]} \omega_i^2 \gamma_i \right], \quad \text{and} \\
\Upsilon &= \lim_{n\to\infty} \left[\frac{2}{K} \sum_{i\in[n]} \sum_{j\neq i} P_{ij}^2 \eta_i^2 \eta_j^2 + \frac{4}{K} \sum_{i\in[n]} \omega_i^2 \eta_i^2 \right].
\end{aligned}$$

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 6. *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 6 are in order. First, $\max_{i\in[n]} |\Pi_i| \leq C$ is mild because $\Pi_i = \mathbb{E}X_i$. Second, Assumption 6 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 B.1. *Suppose Assumptions 5 and 6 hold. Then Assumption 2 holds for Crudu et al.'s (2021) estimators defined above.*

B.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 7. *Suppose Assumption 6 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 7 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 B.1. *Suppose Assumptions 5 and 7 hold. Then, Lemmas 2, 3, S3.1, S3.2 in Mikusheva and Sun (2022) hold.*

Theorem B.2. *Suppose Assumptions 5 and 7 hold. Then, Assumption 2 holds for Mikusheva and Sun's (2022) cross-fit estimators defined above.*

C 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 (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 B.2 throughout the simulation.

D 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,[16] $YOB_{i,c}$ 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 $s \in \{51\ states\}$.[17] 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

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

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

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

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

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.

E Proof of Lemma 2.1

Under strong identification, by (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

$$Q_{e(\beta_0),e(\beta_0)} = Q_{e,e} + 2\Delta Q_{X,e} + \Delta^2 Q_{X,X} = Q_{e,e} + o_p(1),$$

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

This implies

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

F Proof of Lemma 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

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

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.

G Proof of Lemma 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.

H Proof of Lemma 2.4

Under weak identification, (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

$$\begin{aligned} &\Phi_1(\beta_0)/\Delta^4 \to \Upsilon, \\ &(\Phi_1(\beta_0)\Psi(\beta_0) - \Phi_{12}^2(\beta_0))/\Delta^4 \to \Upsilon\Psi - \tau^2, \\ &\frac{\Phi_1(\beta_0) - \Delta\Phi_{12}(\beta_0)}{\Delta^3} \to \tau. \end{aligned}$$

I Proof of Theorem 2.1

The first statement in Theorem 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 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 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 2.1(ii) is a direct consequence of Lemma 2.2. The "only if" part of Theorem 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 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 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 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.

J Proof of Theorem 4.1

We are under weak identification. By Lemma 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

$$\mathbb{E}\widehat{\phi}_{\mathcal{A}_s(\widehat{D}, \widehat{\gamma}(\beta_0))} \to \mathbb{E}\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)), a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(\Delta, \mu_D, \gamma(\beta_0)). \tag{J.1}$$

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

$$\begin{aligned} &\phi_{a_1(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)), a_2(f_s(D,\gamma(\beta_0)),\gamma(\beta_0)),\infty}(0, \mu_D, \gamma(\beta_0)) \\ &= 1\left\{ \begin{matrix} a_1(f_s(D, \gamma(\beta_0)), \gamma(\beta_0))\mathcal{Z}_1^2 + a_2(f_s(D, \gamma(\beta_0)), \gamma(\beta_0))(\rho(\beta_0)\mathcal{Z}_1 + (1 - \rho^2(\beta_0))^{1/2}\mathcal{Z}_2) \\ (1 - a_1(f_s(D, \gamma(\beta_0)), \gamma(\beta_0)) - a_2(f_s(D, \gamma(\beta_0)), \gamma(\beta_0)))\mathcal{Z}_2^2 \\ \geq \mathbb{C}_\alpha(a_1(f_s(D, \gamma(\beta_0)), \gamma(\beta_0)), a_2(f_s(D, \gamma(\beta_0)), \gamma(\beta_0)); \rho(\beta_0)) \end{matrix} \right\}, \end{aligned}$$

$\mathcal{Z}_1$ 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

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

K Proof of Corollary 4.1

By the continuous mapping theorem, we have

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

and

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

where, by construction, we have

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

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

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

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

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

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 2.1(i).

L Proof of Theorem 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
&|\widehat{\Delta}_*(\beta_0)-\Delta_*|\leq c\varepsilon, \quad \inf_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}|d_n\widetilde{\delta}|\geq(1-\varepsilon)(|\Delta_*|-c\varepsilon),\\
&|\widehat{\Phi}_1(\beta_0)-\Phi_1|\leq c\varepsilon, \quad |d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))-\widetilde{\mathcal{C}}^2|\leq c\varepsilon,\\
&\sup_{\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left[1-(d_n^2\widetilde{\delta}^2,d_n\widetilde{\delta})\left(\begin{pmatrix}\widehat{\Phi}_1(\beta_0) & \widehat{\Phi}_{12}(\beta_0)\\ \widehat{\Phi}_{12}(\beta_0) & \widehat{\Psi}(\beta_0)\end{pmatrix}^{-1}\begin{pmatrix}\widehat{\Phi}_{13}(\beta_0)\\ \widehat{\tau}(\beta_0)\end{pmatrix}\right)\right]^2\\
&\leq\left[1-(\Delta_*^2,\Delta_*)\left(\begin{pmatrix}\Phi_1 & \Phi_{12}\\ \Phi_{12} & \Psi\end{pmatrix}^{-1}\begin{pmatrix}\Phi_{13}\\ \tau\end{pmatrix}\right)\right]^2+c\varepsilon\leq c_{\mathcal{B}}+c\varepsilon,\\
&|\widehat{c}_{\mathcal{B}}(\beta_0)-c_{\mathcal{B}}|\leq c\varepsilon.
\end{aligned} \tag{L.3}$$

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 (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

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

In addition, by Lemma 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

$$
\begin{aligned}
&\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)} \Big| \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\}\Big| \\
&\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),
\end{aligned}
$$

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

$$
\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,1}(\varepsilon)} \left|Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})\right| \xrightarrow{p} 0. \tag{L.4}
$$

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

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

By (L.3), on $\mathcal{E}_n(\varepsilon)$, there exists a constant $c > 0$ such that

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

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

and

$$
\begin{aligned}
&\underline{a}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0))\widehat{C}_1^2(d_n\widetilde{\delta})f_s^2(\widehat{D},\widehat{\gamma}(\beta_0)) \\
&\geq \frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0)d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))}\widehat{C}_1^2(d_n\widetilde{\delta})(d_n f_s(\widehat{D},\widehat{\gamma}(\beta_0)))^2 \\
&\geq \frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))(\Phi_1-c\varepsilon)(c_{\mathcal{B}}-c\varepsilon)}{(|\Delta_*|+c\varepsilon)^4(\widetilde{\mathcal{C}}^2+c\varepsilon)}\frac{(\Phi_1(\beta_0)+c\varepsilon)^{-1}(1-\varepsilon)^4(|\Delta_*|-c\varepsilon)^4(\widetilde{\mathcal{C}}^2-c\varepsilon)}{c_{\mathcal{B}}+c\varepsilon} \\
&\geq (1.1-c\varepsilon)\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)),
\end{aligned}
$$

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

$$
\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0)) \geq \mathbb{P}^*(o_p(1)+(1.1-c\varepsilon)\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)) \geq \mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))) \to 1.
$$

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

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

and thus,

$$
\begin{aligned}
&\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)} \left[\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D},\widehat{\gamma}(\beta_0))-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right] \\
&\leq \sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)} \left[1-\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right] \stackrel{p}{\longrightarrow} 0. \qquad \text{(L.5)}
\end{aligned}
$$

Furthermore, note that $a_1+a_2 \leq \overline{a} < 1$ and when $\widetilde{\delta} \in \widetilde{\mathcal{D}}_{n,2}(\varepsilon)$, on $\mathcal{E}_n(\varepsilon)$, (L.3) implies $\widetilde{\delta}^2 \to \infty$. Therefore, we have

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

which further implies

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

and

$$\begin{aligned}\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}&\Bigg[\mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\geq\mathbb{C}_\alpha\}\\&-\mathbb{E}1\left\{\begin{array}{c}a_1\mathcal{Z}_1^2+a_2\left(\rho\mathcal{Z}_1+(1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2\\+(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\geq\mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\Bigg]\xrightarrow{p}0.\end{aligned}\tag{L.6}$$

Combining (L.5) and (L.6), we have

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,2}(\varepsilon)}\left|Q_n(a_1,a_2,\widetilde{\delta})-Q(a_1,a_2,\widetilde{\delta})\right|\to0.\tag{L.7}$$

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

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

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

$$\begin{aligned}\mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))&\geq\mathbb{P}^*((1-\overline{a})\mathcal{Z}_2^2(\widehat{C}_2(d_n\widetilde{\delta})f_s(\widehat{D},\widehat{\gamma}(\beta_0)))\geq\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)))\\&\geq\mathbb{P}^*\left((1-\overline{a})\mathcal{Z}_2^2\left(\frac{M_1(\varepsilon)\varepsilon|\widetilde{\mathcal{C}}|}{(2(1-\rho^2)\Psi c_{\mathcal{B}})^{1/2}}\right)\geq\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\right)\\&\geq\mathbb{P}^*\left((1-\overline{a})\mathcal{Z}_2^2\left(\frac{M_1(\varepsilon)\varepsilon|\widetilde{\mathcal{C}}|}{(2(1-\rho^2)\Psi c_{\mathcal{B}})^{1/2}}\right)\geq\mathbb{C}_{\alpha,\max}(\rho)+c\varepsilon\right)-\varepsilon\geq1-2\varepsilon,\end{aligned}$$

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 (L.2) and the fact that $\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0)) \xrightarrow{p} \mathbb{C}_{\alpha,\max}(\rho)$. This implies, on $\mathcal{E}_n(\varepsilon)$,

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)} \left[\mathcal{P}_{d_n\widetilde{\delta},s}(\widehat{D},\widehat{\gamma}(\beta_0)) - \mathbb{E}^*\phi_{a_1,a_2,s}(d_n\widetilde{\delta},\widehat{D},\widehat{\gamma}(\beta_0))\right] \leq 2\varepsilon. \tag{L.8}$$

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

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

and

$$\begin{aligned} &\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)} \Bigg[\mathbb{E}1\{\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha\} \\ &- \mathbb{E}1\left\{\begin{array}{l} a_1\mathcal{Z}_1^2 + a_2\left(\rho\mathcal{Z}_1 + (1-\rho^2)^{1/2}\mathcal{Z}_2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}})\right)^2 \\ +(1-a_1-a_2)\mathcal{Z}_2^2((1-\rho^2)^{-1/2}\Psi^{-1/2}\widetilde{\delta}\widetilde{\mathcal{C}}) \geq \mathbb{C}_\alpha(a_1,a_2;\rho)\end{array}\right\}\Bigg] \leq \varepsilon. \end{aligned} \tag{L.9}$$

Combining (L.8) and (L.9), we have, on $\mathcal{E}_n(\varepsilon)$,

$$\sup_{(a_1,a_2)\in\mathbb{A}(f_s(\widehat{D},\widehat{\gamma}(\beta_0)),\widehat{\gamma}(\beta_0)),\widetilde{\delta}\in\widetilde{\mathcal{D}}_{n,3}(\varepsilon)} \left|Q_n(a_1,a_2,\widetilde{\delta}) - Q(a_1,a_2,\widetilde{\delta})\right| \leq 3\varepsilon. \tag{L.10}$$

Combining (L.4), (L.7), and (L.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

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

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 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 4.2 and $\mathcal{N}_2^* = (1-\rho^2)^{-1/2}(\mathcal{N}_2 - \rho\mathcal{N}_1)$. Then, the result holds by Theorem 2.1(ii).

M Proof of Theorem 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 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 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 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. (2020) and then we can establish the result by using Corollary 2.1 in their paper.

N Proof of Theorem 4.4

We consider strong identification with fixed alternatives. By construction, we have $\mathcal{A}_{1,s}(\widehat{D}, \widehat{\gamma}(\beta_0)) \geq \frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))}$. By Theorem 2.1(iii), it suffices to show that, w.p.a.1,

$$\frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))} \geq \frac{\tilde{q}\Psi^2(\beta_0)\rho^4(\beta_0)}{\mathcal{C}^2\Phi_1(\beta_0)},$$

or equivalently,

$$\frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0) d_n^2 f_s^2(\widehat{D},\widehat{\gamma}(\beta_0))} \geq \frac{\tilde{q}\Psi^2(\beta_0)\rho^4(\beta_0)}{\widetilde{\mathcal{C}}^2\Phi_1(\beta_0)} = \frac{\tilde{q}\Phi_1(\beta_0)}{\widetilde{\mathcal{C}}^2\Delta_*^4(\beta_0)}, \tag{N.1}$$

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,

$$\begin{aligned}\frac{1.1\mathbb{C}_{\alpha,\max}(\widehat{\rho}(\beta_0))\widehat{\Phi}_1(\beta_0)\widehat{c}_{\mathcal{B}}(\beta_0)}{\widehat{\Delta}_*^4(\beta_0)d_n^2 f_s^2(\widehat{D}, \widehat{\gamma}(\beta_0))} &\geq \frac{1.1(\mathbb{C}_{\alpha,\max}(\rho(\beta_0)) - c\varepsilon)(c_{\mathcal{B}}(\beta_0) - c\varepsilon)(\Phi_1(\beta_0) - c\varepsilon)}{(\Delta_*^4(\beta_0) + c\varepsilon)(c_{\mathcal{B}}(\beta_0) + \varepsilon)\widetilde{\mathcal{C}}^2}\\ &\geq \frac{(1.1 - c\varepsilon)\mathbb{C}_{\alpha,\max}(\rho(\beta_0))\Phi_1(\beta_0)}{\Delta_*^4(\beta_0)\widetilde{\mathcal{C}}^2},\end{aligned}$$

where the second inequality holds because ε can be arbitrarily small. Then, we can let $\tilde{q}$ in (N.1) be $(1.1 - c\varepsilon)\mathbb{C}_{\alpha,\max}(\rho(\beta_0))$ which is greater than $\mathbb{C}_{\alpha,\max}(\rho(\beta_0))$. This concludes the proof.

O Proof of Theorem A.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 S.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

$$Q_{X,e} = Q_{\overline{X}-W\hat{\gamma}_V,\tilde{e}-W\hat{\gamma}_e}$$

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

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

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

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

$$
Var\left(\sum_{i\in[n]} \frac{(G_i+\omega_i)\tilde{e}_i}{\sqrt{K}} \right) = \sum_{i\in[n]} \frac{(G_i+\omega_i)^2 \tilde{\sigma}_i^2}{K}
$$

$$\leq C \sum_{i\in[n]} \frac{(G_i+\omega_i)^2}{K}$$
$$\leq C\left[\frac{(\sum_{i\in[n]}\Pi_i P_{ii} W_i^\top)(\sum_{i\in[n]} W_i W_i^\top)^{-1}(\sum_{i\in[n]}\Pi_i P_{ii} W_i)}{K} + \frac{\Pi^\top\Pi}{K}\right]$$
$$\leq C\left[p_n^2 \frac{\Pi^\top\Pi}{K} + \frac{\Pi^\top\Pi}{K}\right] = O(1)$$

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, we note that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i\in[n]} G_i\tilde{e}_i/\sqrt{K}\right)^2 &\leq C\sum_{i\in[n]} G_i^2/K \\
&= C(\sum_{i\in[n]}\Pi_i P_{ii} W_i^\top)(\sum_{i\in[n]} W_i W_i^\top)^{-1}(\sum_{i\in[n]}\Pi_i P_{ii} W_i)/K \\
&\leq C\sum_{i\in[n]}\Pi_i^2 P_{ii}^2/K \\
&\leq C\Pi^\top\Pi p_n^2/K.
\end{aligned}$$

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

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

P Proof of Theorem B.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,

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

$$= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 V_i^2(\Delta)U_j^2(\Delta) + \Delta\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\Pi_i V_i(\Delta)U_j^2(\Delta) + \Pi_j U_j(\Delta)V_i^2(\Delta))$$
$$+ \Delta^2\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\Pi_i^2 U_j^2(\Delta) + \Pi_j^2 V_i^2(\Delta) + 4\Pi_i\Pi_j V_i(\Delta)U_j(\Delta))$$
$$+ \Delta^3\frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\Pi_i^2\Pi_j U_j(\Delta) + \Pi_j^2\Pi_i V_i(\Delta)) + \Delta^4\frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\Pi_i^2\Pi_j^2$$
$$\equiv \sum_{l=0}^{4}\Delta^l T_l.$$

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

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

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

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

and

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

Therefore, we have

$$
\begin{aligned}
T_0 &= \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \mathbb{E}(V_i^2(\Delta)U_j^2(\Delta)) + o_p(1) \\
&= \Delta^4 \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \eta_i^2\eta_j^2 + \Delta^3 \frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\eta_i^2\gamma_j + \eta_j^2\gamma_i) + \Delta^2 \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\eta_i^2\sigma_j^2 + \eta_j^2\sigma_i^2 + 4\gamma_i\gamma_j) \\
&+ \Delta \frac{4}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2(\gamma_i\sigma_j^2 + \gamma_j\sigma_i^2) + \frac{2}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\sigma_i^2\sigma_j^2 + o_p(1) \\
&= \Phi_1(\beta_0) + o_p(1).
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 X_i e_i(\beta_0) X_j e_j(\beta_0)) \\
&= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i e_i(\beta_0)\Pi_j e_j(\beta_0)) + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 \Pi_i e_i(\beta_0) V_j e_j(\beta_0)) \\
&+ \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 V_i e_i(\beta_0)\Pi_j e_j(\beta_0)) + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 V_i e_i(\beta_0) V_j e_j(\beta_0)) \\
&= \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2 (\gamma_i + \Delta\eta_i^2)(\gamma_j + \Delta\eta_j^2) + o_p(1). \qquad \text{(P.1)}
\end{aligned}
$$

In addition, we have

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

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 6. Combining (P.1) and (P.2), we have

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

Q Proof of Theorem B.2

Given Lemma B.1, Lemmas 2 and 3 in Mikusheva and Sun (2022) hold under Assumptions 5 and 7. 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 B.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 B.1, Lemma S3.1 in Mikusheva and Sun (2022) holds under Assumptions 5 and 7, 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 B.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 B.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

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXX_je_iM_ie = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2\sigma_i^2 + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXe_jX_iM_ie = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\gamma_j\gamma_i + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXe_je_iM_iX = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\gamma_j\gamma_i + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXX_jX_iM_ie = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2\gamma_i + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXX_je_iM_iX = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2\gamma_i + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXe_jX_iM_iX = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\gamma_j\eta_i^2 + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 M_jXX_jX_iM_iX = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^2\eta_j^2\eta_i^2 + o_p(1).$$

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

$$\frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i}\widetilde{P}_{ij}^2 X_i M_i X M_j X e_j(\beta_0) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\eta_i^2\gamma_j + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\eta_i^2\eta_j^2\Delta + o_p(1),$$
$$\frac{1}{K}\sum_{i\in[n]}(\sum_{j\neq i} P_{ij}X_j)^2\left(\frac{e_i(\beta_0)M_i X}{2M_{ii}} + \frac{X_i M_i e(\beta_0)}{2M_{ii}}\right) = \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\eta_i^2\gamma_j + \frac{1}{K}\sum_{i\in[n]}\sum_{j\neq i} P_{ij}^2\eta_i^2\eta_j^2\Delta + o_p(1),$$

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.

R Proof of Lemma B.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

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

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

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

$$\frac{C\Delta^4}{K^2}\left(\sum_{i\in[n]}\lambda_i^2\left(\sum_{j\in[n]}P_{ij}^2\right)\Pi^\top\Pi + \lambda^\top\lambda\left(\sum_{j\in[n]}P_{jj}|\Pi_j|\right)^2\right) \le \frac{C\Delta^4}{K^2}\Big(p_n(\lambda^\top\lambda)(\Pi^\top\Pi) + (\lambda^\top\lambda)(p_nK)(\Pi^\top\Pi)\Big)$$

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

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

and

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

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

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

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

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 $o(1)$:

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

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

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

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

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

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

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

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

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

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

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

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

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

where $\sum_{k\in[n]}|M_{jk}M_{j'k}| \leq 1$ 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.,

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

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

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

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

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

where $\xi_{1,i}, \xi_{2,i}$ stay for either e_i or V_i, 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 $o(1)$:

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

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

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

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

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

$$\leq \frac{C}{K^2}\left(p_n^2(\Pi^\top\Pi)^2 + p_n^2(\Pi^\top\Pi)^2 + p_n^2(\Pi^\top\Pi) + p_n^2(\Pi^\top\Pi)\right) = o(1),$$

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

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},

$$\frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}^2 a_i^2 + \sum_{i\in[n]}\sum_{j\in[n]}P_{ij}^2\,|a_i a_j|\right) \leq \frac{C}{K^2}\left(p_n^2 a^\top a + p_n a^\top a\right) = o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\omega_i^4\frac{\lambda_i^2}{M_{ii}^2} \leq \frac{C}{K^2}\left(\max_{i\in[n]}\omega_i^2\right)^2\sum_{i\in[n]}\lambda_i^2 \leq Cp_n^2\left(\frac{\Pi^\top\Pi}{K}\right)^3 = o(1),$$

$$\frac{C}{K^2}\sum_{i\in[n]}\omega_i^4\Pi_i^2 \leq \frac{C}{K^2}\sum_{i\in[n]}\omega_i^4 \leq \frac{C}{K^2}p_n\left(\Pi^\top\Pi\right)^2 = o(1),\ \text{where we have used } \max_{i\in[n]}|\Pi_i| \leq C,$$

$$\frac{C}{K^2}\sum_{i\in[n]}\sum_{j\neq i}P_{ij}^4\left(a_i^2 + |a_i|\,|a_j|\right) \leq \frac{C}{K^2}\left(p_n^2 a^\top a + p_n^2 a^\top a\right) = o(1),$$

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

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

$$\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\,|\Pi_i|\right)^2 + \left(\sum_{i\in[n]}|\omega_i\Pi_i|\right)^2\right) \leq \frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\right)^2 + \left(\sum_{i\in[n]}|\omega_i\Pi_i|\right)^2\right)$$

$$\leq \frac{C}{K^2}\left(\left(\Pi^\top\Pi\right)^2 + \left(\sum_{i\in[n]}\omega_i^2\right)\left(\Pi^\top\Pi\right)\right) \leq \frac{C}{K^2}\left((\Pi^\top\Pi)^2 + (\Pi^\top\Pi)^2\right) = o(1),$$

When $a_i = \frac{\lambda_i}{M_{ii}}$, we have

$$\frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^2\left|\frac{\lambda_i}{M_{ii}}\right|\right)^2 + \left(\sum_{i\in[n]}\left|\omega_i\frac{\lambda_i}{M_{ii}}\right|\right)^2\right) \leq \frac{C}{K^2}\left(\left(\sum_{i\in[n]}\omega_i^4\right)(\lambda^\top\lambda) + \left(\sum_{i\in[n]}\omega_i^2\right)(\lambda^\top\lambda)\right)$$

$$\leq \frac{C}{K^2}\left(p_n(\Pi^\top\Pi)^2(\lambda^\top\lambda)+(\Pi^\top\Pi)(\lambda^\top\lambda)\right)=o(1).$$

Furthermore, we can show that

$$\frac{C}{K^2}\left(\sum_{i\in[n]}|\omega_i a_i|\right)^2 \leq \frac{C}{K^2}(\Pi^\top\Pi)(a^\top a)=o(1),$$

$$\frac{C}{K}\sum_{i\in[n]}P_{ii}\,|a_i| \leq \frac{C}{K}\left(\sum_{i\in[n]}P_{ii}^2\right)^{1/2}\left(a^\top a\right)^{1/2} \leq \frac{C}{K}(p_nK)^{1/2}\left(a^\top a\right)^{1/2}=o(1),$$

$$\frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}\,|a_i|\right)^2 \leq \frac{C}{K^2}\left(\sum_{i\in[n]}P_{ii}^2\right)\left(a^\top a\right) \leq \frac{C}{K^2}p_nK\left(a^\top a\right)=o(1).$$

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.

S Lemma S.1 and Its Proof

Lemma S.1. *Suppose assumptions in Theorem A.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

$$\mathbb{E}Q_{\tilde{e},W}Q_{\tilde{e},W}^\top \leq C \sum_{i\in[n]} (\sum_{j\neq i} P_{ij}W_j)^\top (\sum_{j\neq i} P_{ij}W_j)/K = C \sum_{i\in[n]} P_{ii}^2 W_i^\top W_i/K \leq C,$$

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 $\sum_{i\in[n]} W_i W_i^\top/n = O_P(1)$ 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

$$\hat{\gamma}_e^\top Q_{W,W^\top} \hat{\gamma}_e = o_P(1), \quad \text{and} \quad \hat{\gamma}_e^\top Q_{W,W^\top} \hat{\gamma}_V = o_P(1).$$

□

T Comparison with HLIM Estimator under Strong Identification

We consider the model in Section A 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

$$(\hat{\beta}^{HLIM}, \hat{\gamma}^{HLIM}) = \arg\min_{b,r} \mathcal{Q}(b,r), \quad \mathcal{Q}(b,r) = \frac{\sum_{i\in[n]}\sum_{j\neq i}(\tilde{Y}_i - \tilde{X}_i b - W_i^\top r)\tilde{P}_{ij}(\tilde{Y}_i - \tilde{X}_i b - W_i^\top r)}{\sum_{i\in[n]}(\tilde{Y}_i - \tilde{X}_i b - W_i^\top r)^2},$$

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, $X_i = \overline{X}_i - W_i^\top \hat{\gamma}_V$, $\overline{e}_i(\beta_0) = e_i(\beta_0) + W_i^\top \hat{\gamma}_e - W_i^\top \hat{\pi}_x(\beta_0 - \beta)$, 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

$$\begin{aligned}
&(\mu_n^2/\sqrt{K})(\hat{\beta}^{HLIM} - \beta_0) \\
&= (\mu_n^2/\sqrt{K})L(\hat{\delta} - \delta_0) \\
&= -(\mu_n^2/\sqrt{K})L\left(\frac{\partial \hat{D}(\overline{\delta})}{\partial \delta}\right)^{-1}\hat{D}(\delta_0)
\end{aligned}$$

$$
\begin{aligned}
&= -(\mu_n^2/\sqrt{K})L(S_n^\top)^{-1}\left(S_n^{-1}\frac{\partial \hat{D}(\overline{\delta})}{\partial \delta}(S_n^\top)^{-1}\right)^{-1} S_n^{-1}\hat{D}(\delta_0) \\
&= -(\mu_n^2/\sqrt{K})(0,1/\mu_n)(H^{-1}+o_P(1))\mathrm{diag}(1/\sqrt{n},\cdots,1/\sqrt{n},1/\mu_n)\begin{pmatrix} I_d & 0 \\ -\pi_x^\top & 1\end{pmatrix}\hat{D}(\delta_0) \\
&= -\frac{\mu_n}{\sqrt{K}}\left(\left((H^{21}+o_P(1))/\sqrt{n}-\pi_x^\top(H^{22}+o_P(1))/\mu_n,(H^{22}+o_P(1))/\mu_n\right)\right)\hat{D}(\delta_0) \\
&= (H^{22}+o_P(1))(-\pi_x^\top,1)\hat{D}(\delta_0)/\sqrt{K} \\
&= (H^{22}+o_P(1))\frac{\sum_{i\in[n]}\sum_{j\neq i}\left[\overline{X}_i\tilde{P}_{ij}\overline{e}_j(\beta_0)-\overline{e}_i(\beta_0)\tilde{P}_{ij}\overline{e}_j(\beta_0)\frac{\overline{X}^\top\overline{e}(\beta_0)}{\overline{e}^\top(\beta_0)\overline{e}(\beta_0)}\right]}{\sqrt{K}},
\end{aligned}
$$

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 S.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

$$\sum_{i\in[n]} X_i e_i(\beta_0) P_{W,ii} = \frac{1}{n}\sum_{i\in[n]} X_i e_i(\beta_0) W_i^\top \left(W^\top W/n\right)^{-1} W_i = O_P(1).$$

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 $W_h(\beta) = \frac{Q_{X,e(\beta_0)} - \tilde{\rho} Q_{e(\beta_0),e(\beta_0)}}{(\Psi - 2\tilde{\rho}\Phi_{12} + \tilde{\rho}^2\Phi_1)^{1/2}} + o_P(1)$.

U Additional Simulation Results

U.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 2.

For Figures 9–40, all the settings remain the same as those in Section 5.1 in the main paper except we use alternative values of the tuning parameters for (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 5. Specifically, Figures 9–12 report the results for $(0.01, 1.5)$, Figures 13–16 report those for $(0.01, 2)$, Figures 17–20 report those for $(0.001, 1.1)$, Figures 21–24 report those for $(0.001, 1.5)$, Figures 25–28 report those for $(0.001, 2)$, Figures 29–32 report those for $(0.1, 1.1)$, Figures 33–36 report those for $(0.1, 1.5)$, and Figures 37–40 report those for $(0.1, 2)$, respectively. We find the results are very similar to those reported in the main paper.

Furthermore, Figures 41–44 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 41–44, the tuning parameters are set as $(p_1, p_2) = (0.01, 1.1)$, which are same as those in Section 5 of the main text. The results for other values of p_1 and p_2 remain very similar and thus are omitted for brevity.

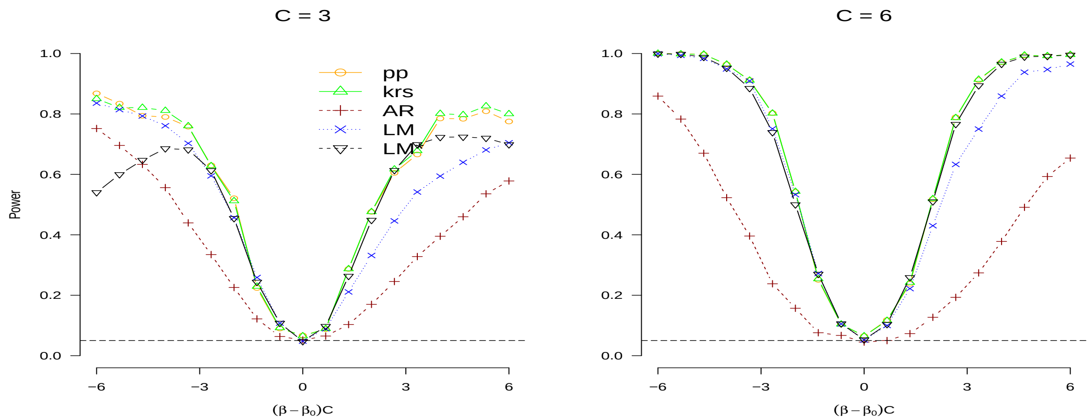

Figure 9: Power Curve for $\rho = 0.2$, $p_1 = 0.01$, and $p_2 = 1.5$

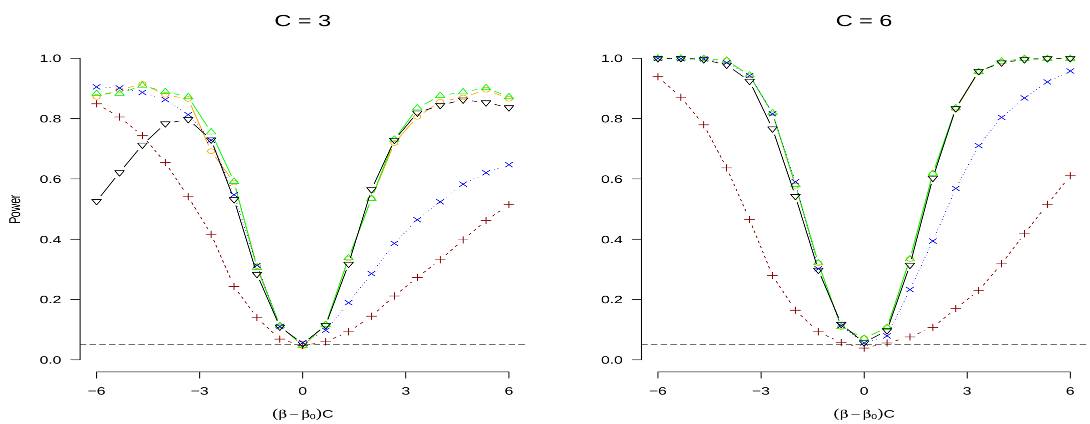

Figure 10: Power Curve for $\rho = 0.4$, $p_1 = 0.01$, and $p_2 = 1.5$

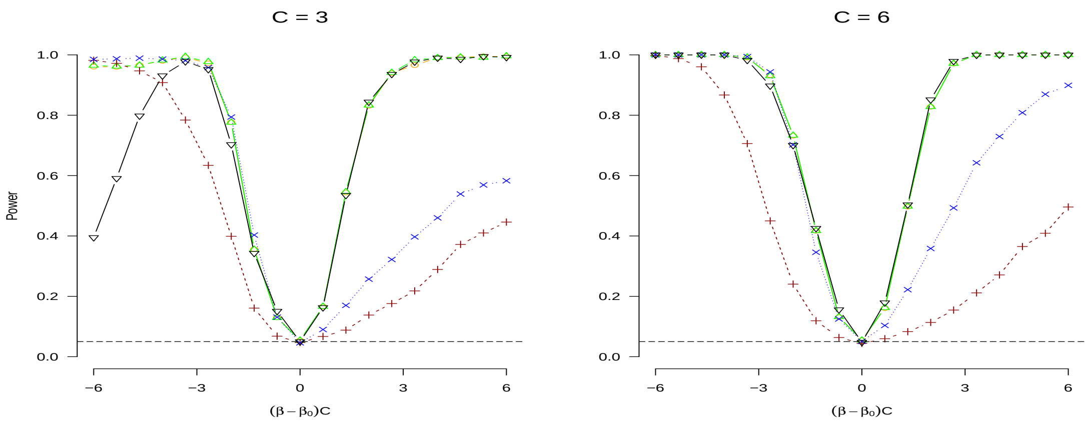

Figure 11: Power Curve for $\rho = 0.7$, $p_1 = 0.01$, and $p_2 = 1.5$

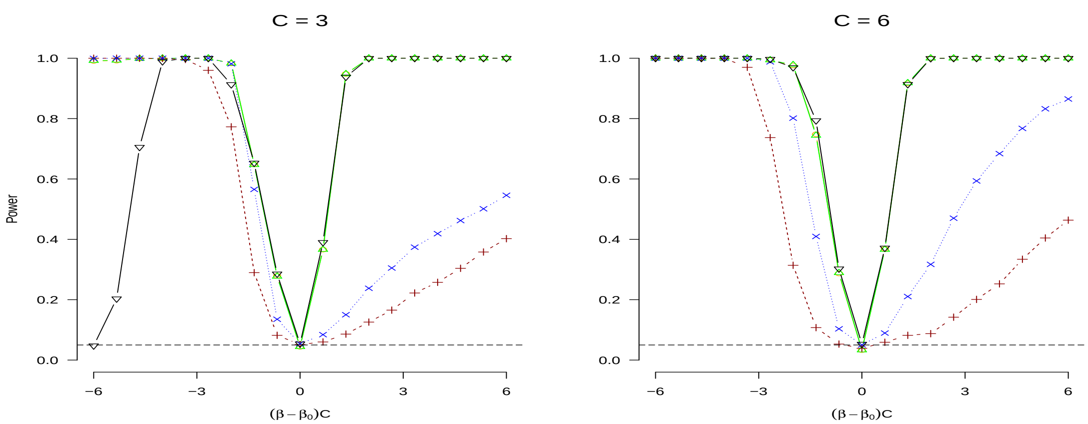

Figure 12: Power Curve for $\rho = 0.9$, $p_1 = 0.01$, and $p_2 = 1.5$

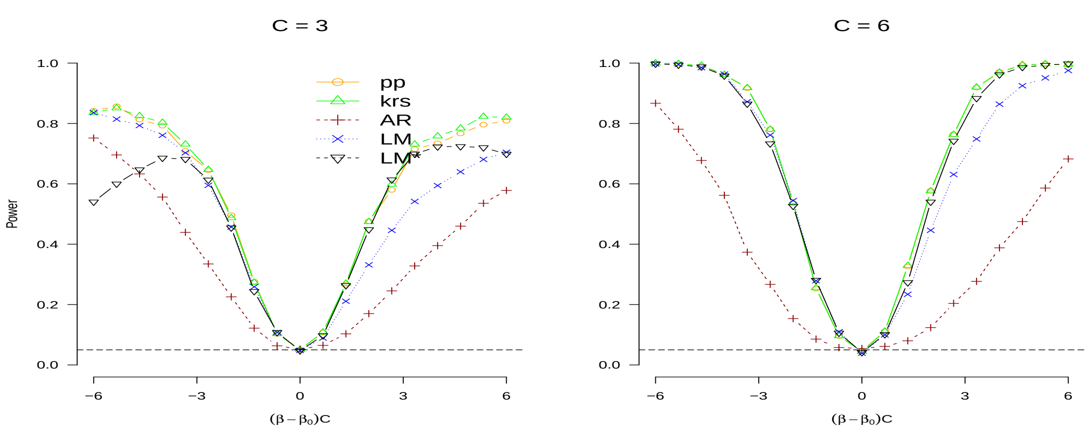

Figure 13: Power Curve for $\rho = 0.2$, $p_1 = 0.01$, and $p_2 = 2$

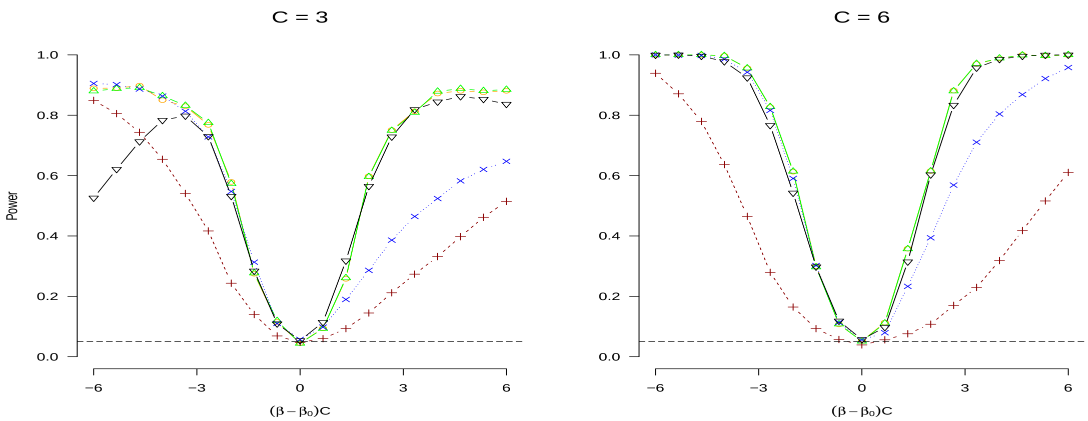

Figure 14: Power Curve for $\rho = 0.4$, $p_1 = 0.01$, and $p_2 = 2$

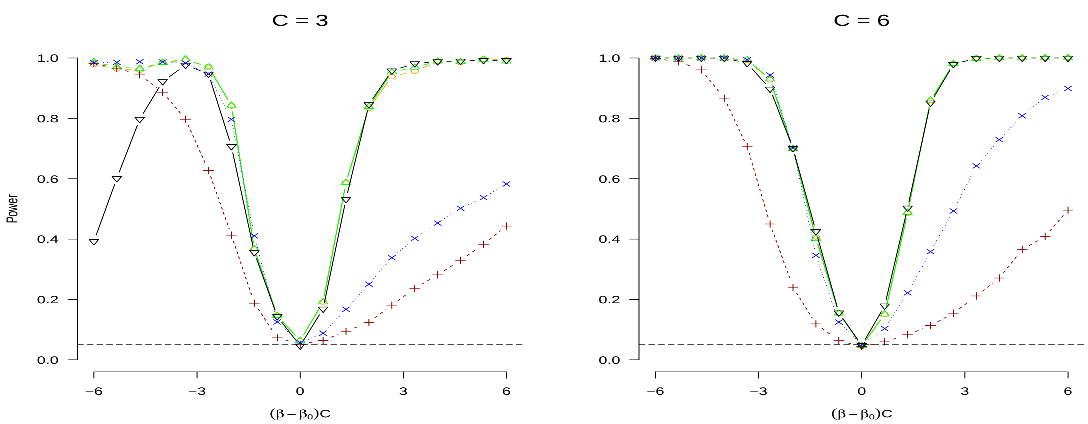

Figure 15: Power Curve for $\rho = 0.7$, $p_1 = 0.01$, and $p_2 = 2$

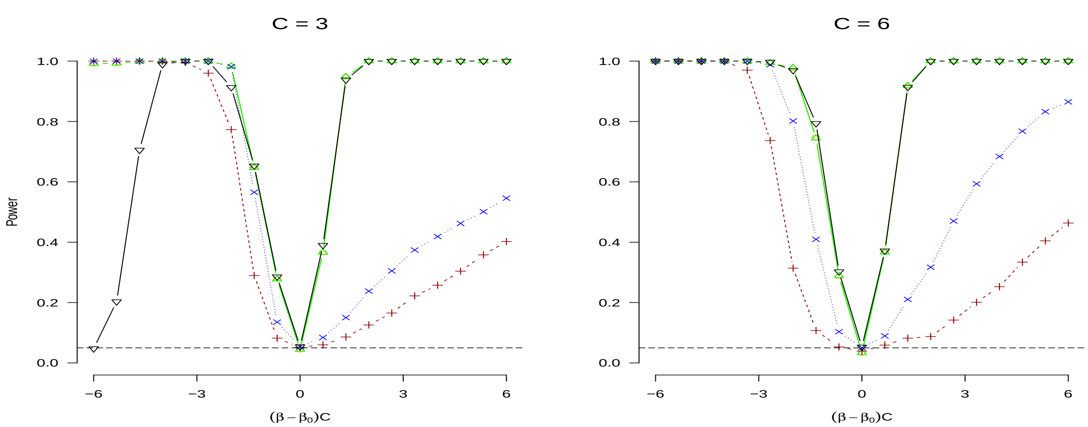

Figure 16: Power Curve for $\rho = 0.9$, $p_1 = 0.01$, and $p_2 = 2$

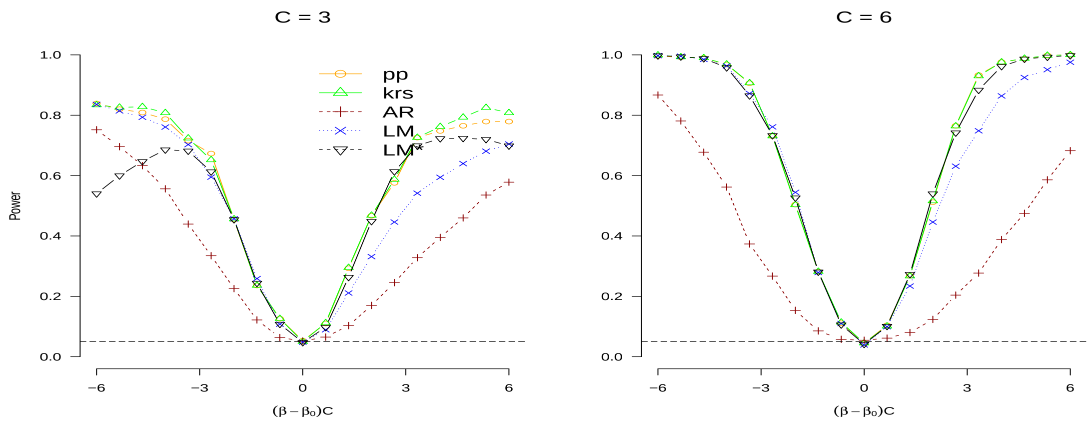

Figure 17: Power Curve for $\rho = 0.2$, $p_1 = 0.001$, and $p_2 = 1.1$

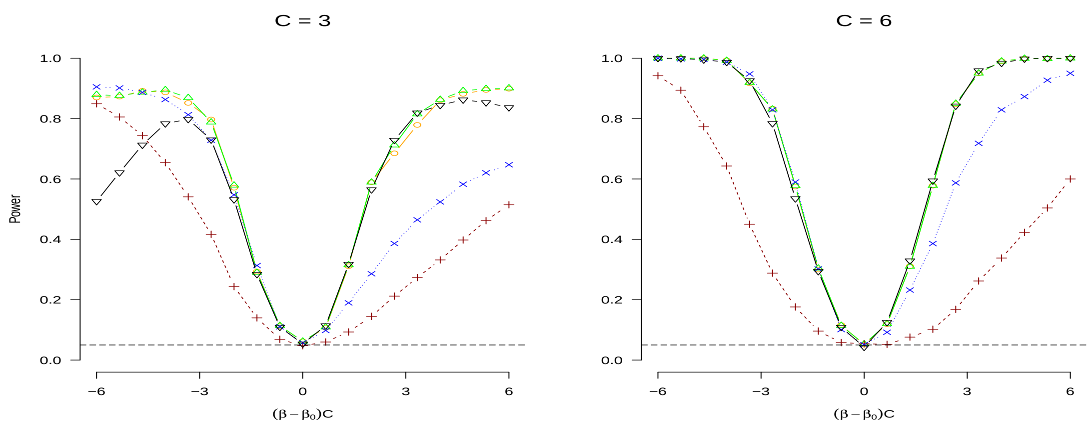

Figure 18: Power Curve for $\rho = 0.4$, $p_1 = 0.001$, and $p_2 = 1.1$

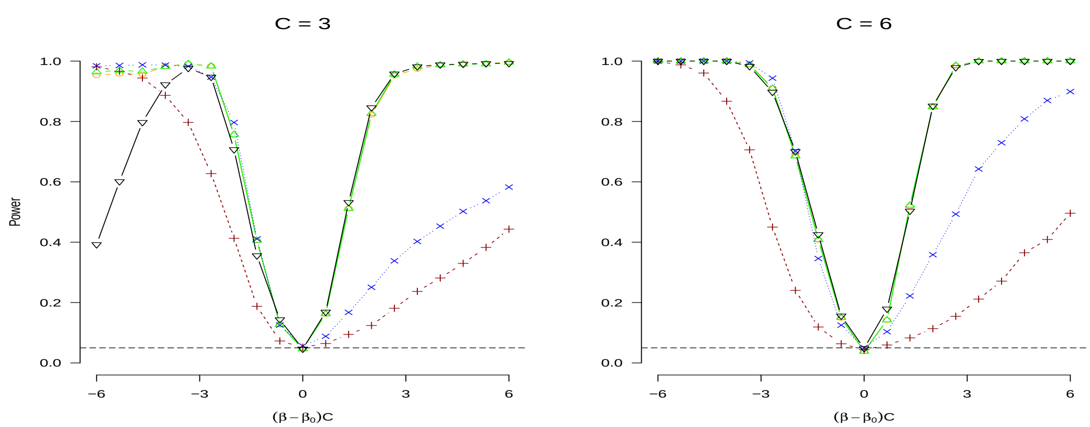

Figure 19: Power Curve for $\rho = 0.7$, $p_1 = 0.001$, and $p_2 = 1.1$

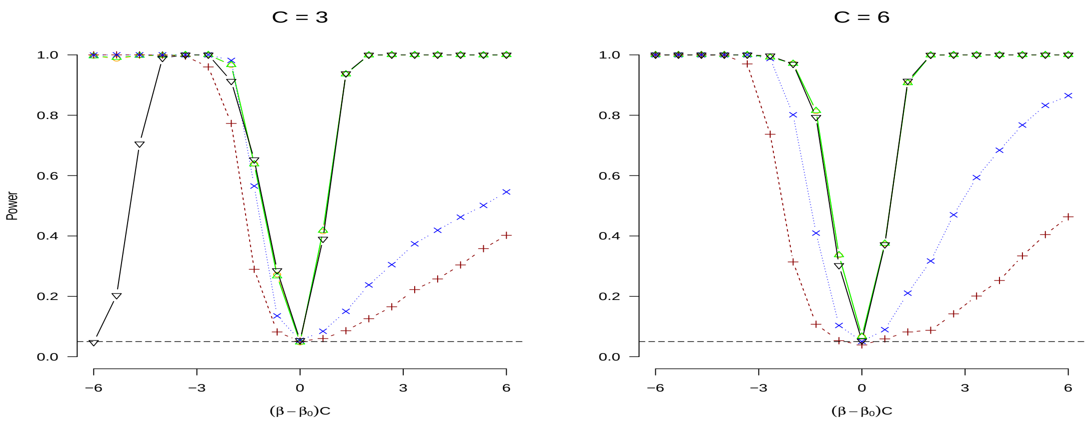

Figure 20: Power Curve for $\rho = 0.9$, $p_1 = 0.001$, and $p_2 = 1.1$

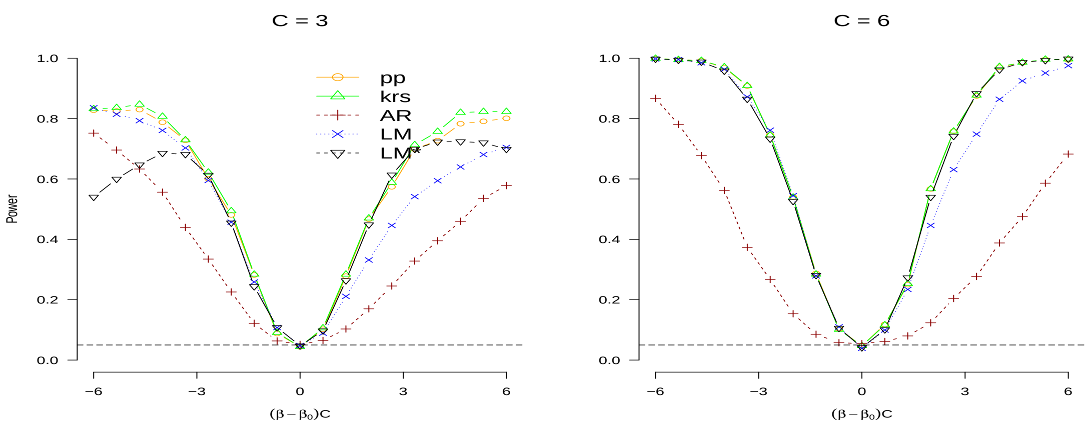

Figure 21: Power Curve for $\rho = 0.2$, $p_1 = 0.001$, and $p_2 = 1.5$

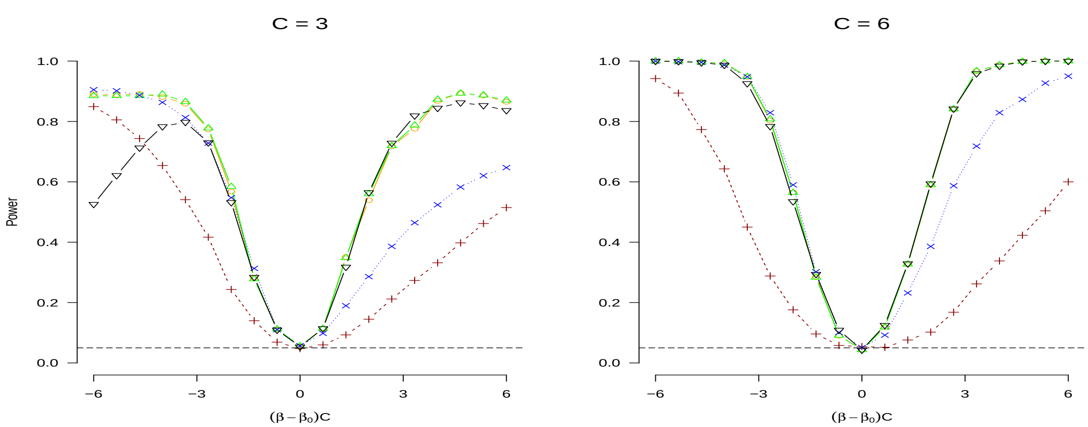

Figure 22: Power Curve for $\rho = 0.4$, $p_1 = 0.001$, and $p_2 = 1.5$

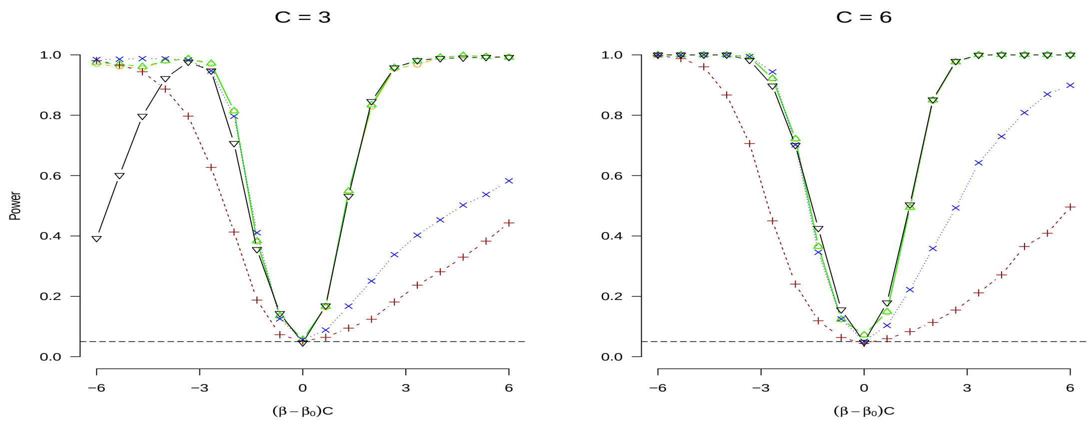

Figure 23: Power Curve for $\rho = 0.7$, $p_1 = 0.001$, and $p_2 = 1.5$

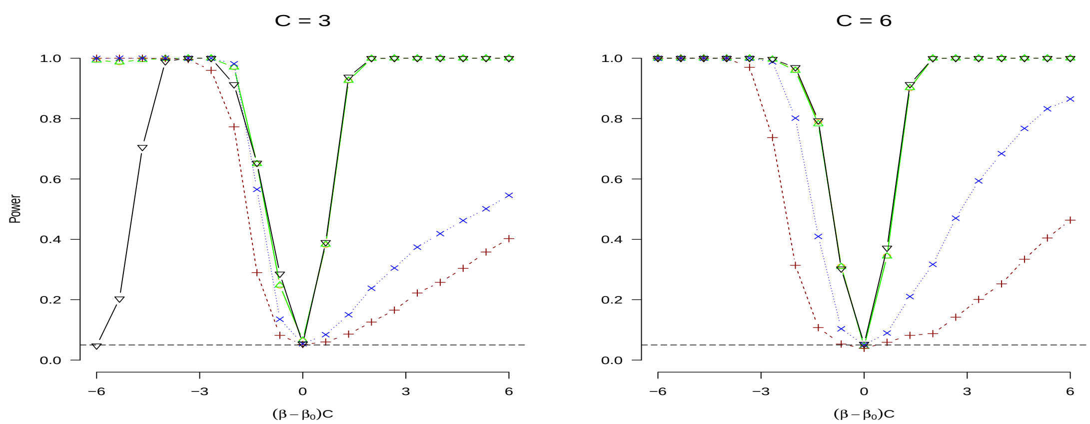

Figure 24: Power Curve for $\rho = 0.9$, $p_1 = 0.001$, and $p_2 = 1.5$

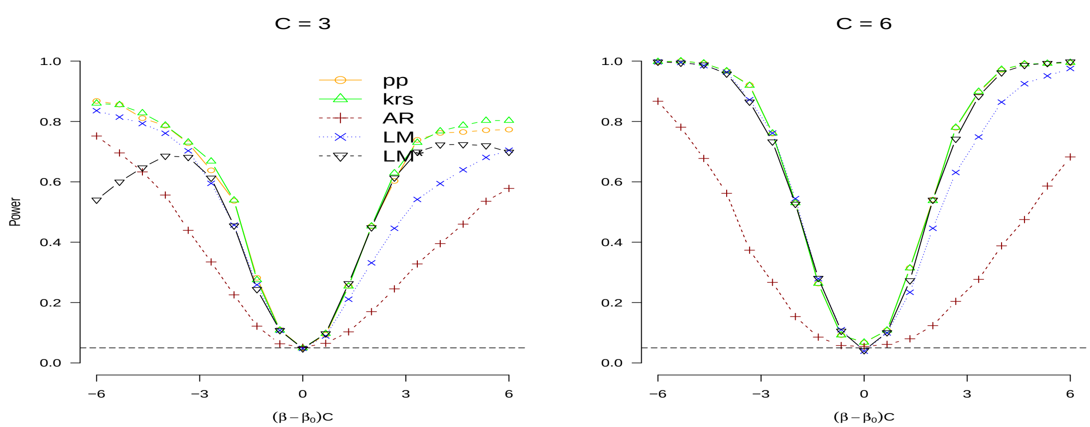

Figure 25: Power Curve for $\rho = 0.2$, $p_1 = 0.001$, and $p_2 = 2$

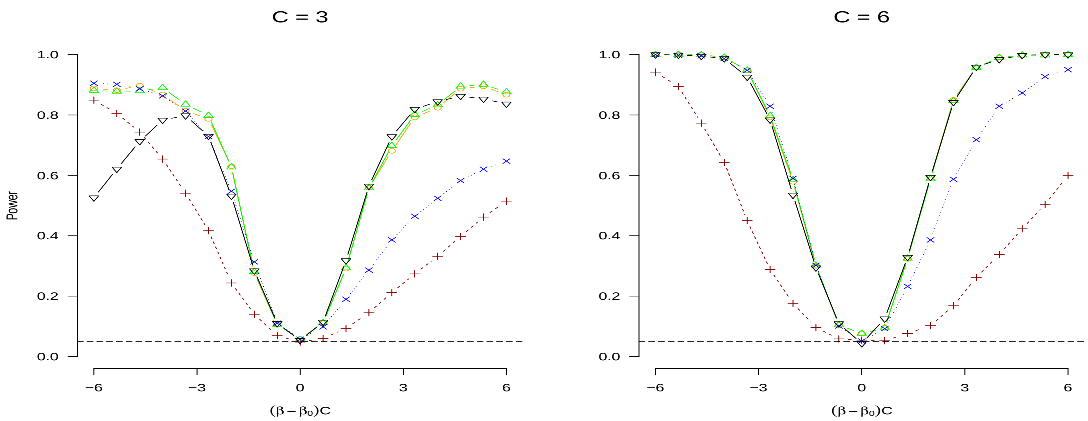

Figure 26: Power Curve for $\rho = 0.4$, $p_1 = 0.001$, and $p_2 = 2$

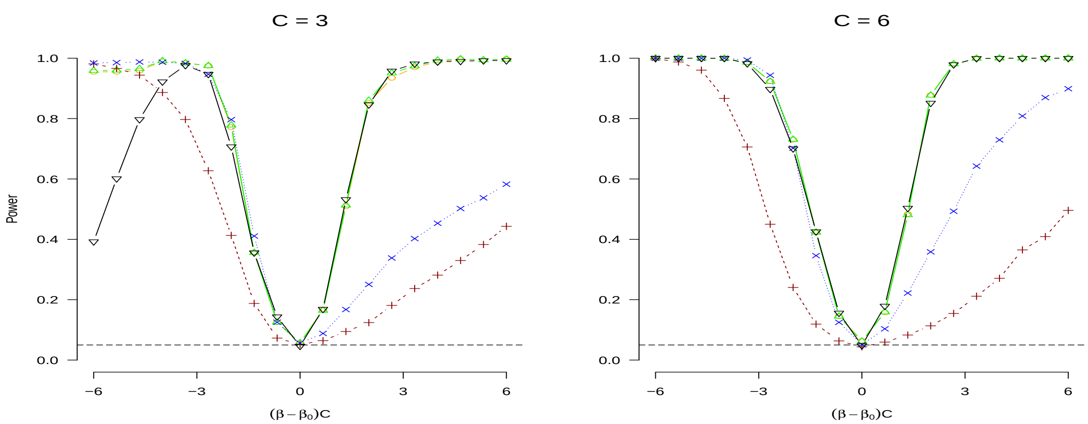

Figure 27: Power Curve for $\rho = 0.7$, $p_1 = 0.001$, and $p_2 = 2$

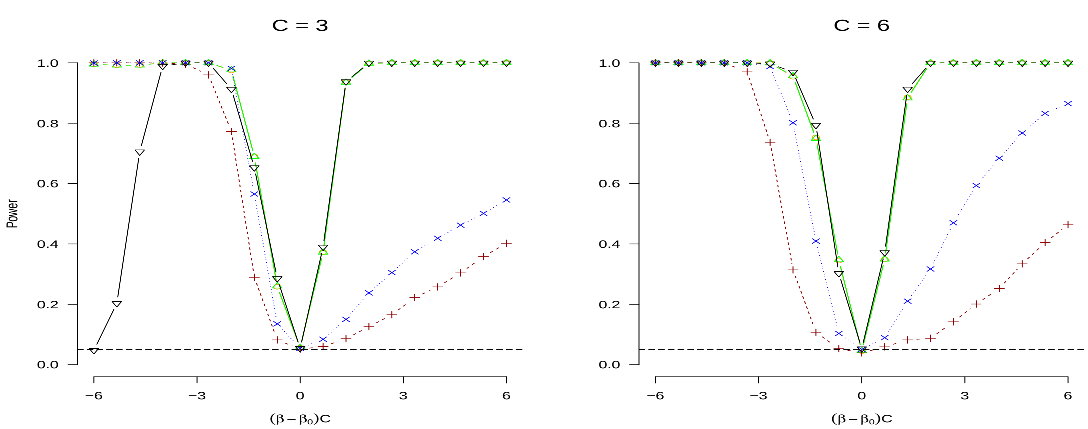

Figure 28: Power Curve for $\rho = 0.9$, $p_1 = 0.001$, and $p_2 = 2$

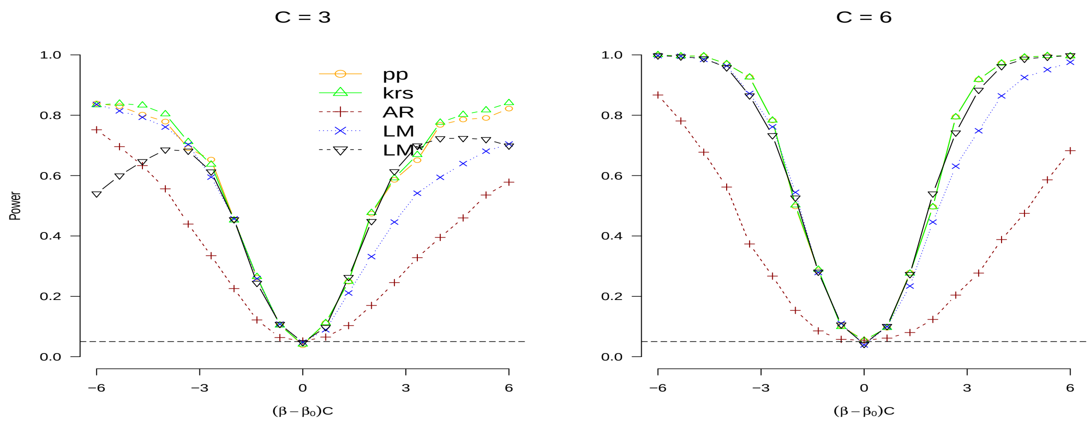

Figure 29: Power Curve for $\rho = 0.2$, $p_1 = 0.1$, and $p_2 = 1.1$

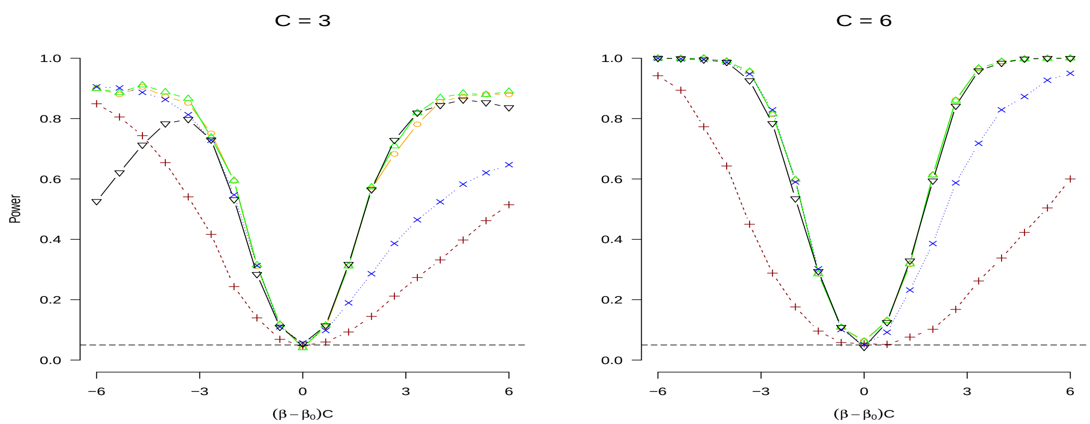

Figure 30: Power Curve for $\rho = 0.4$, $p_1 = 0.1$, and $p_2 = 1.1$

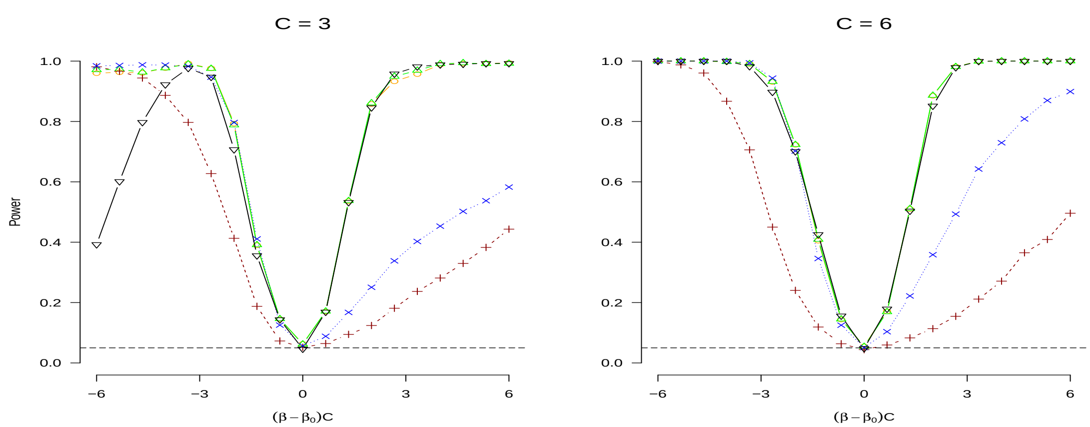

Figure 31: Power Curve for $\rho = 0.7$, $p_1 = 0.1$, and $p_2 = 1.1$

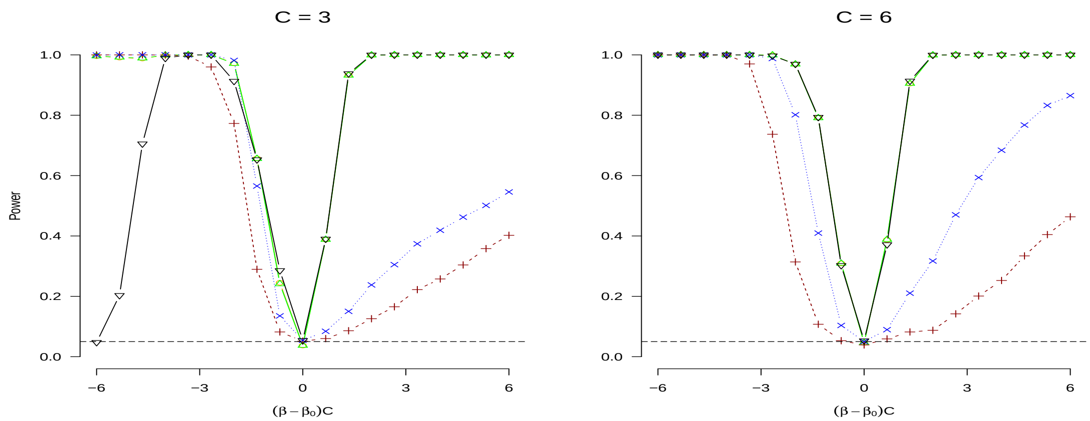

Figure 32: Power Curve for $\rho = 0.9$, $p_1 = 0.1$, and $p_2 = 1.1$

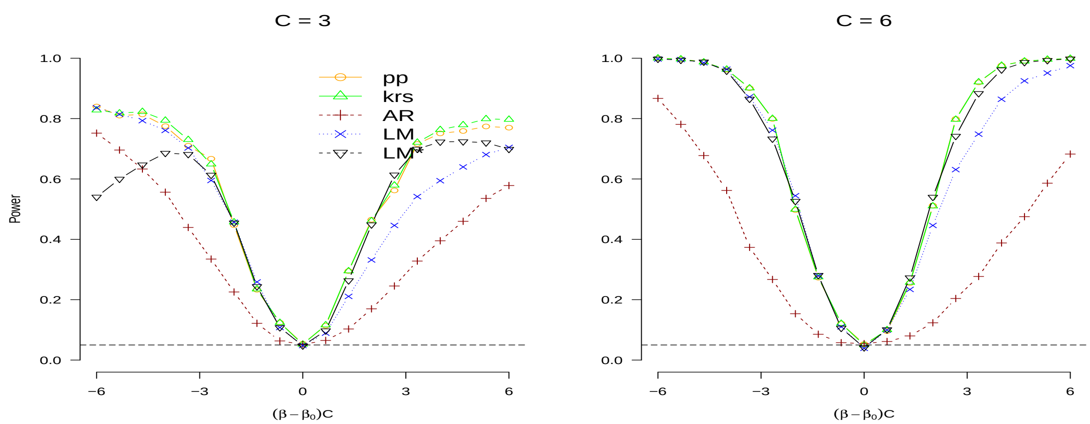

Figure 33: Power Curve for $\rho = 0.2$, $p_1 = 0.1$, and $p_2 = 1.5$

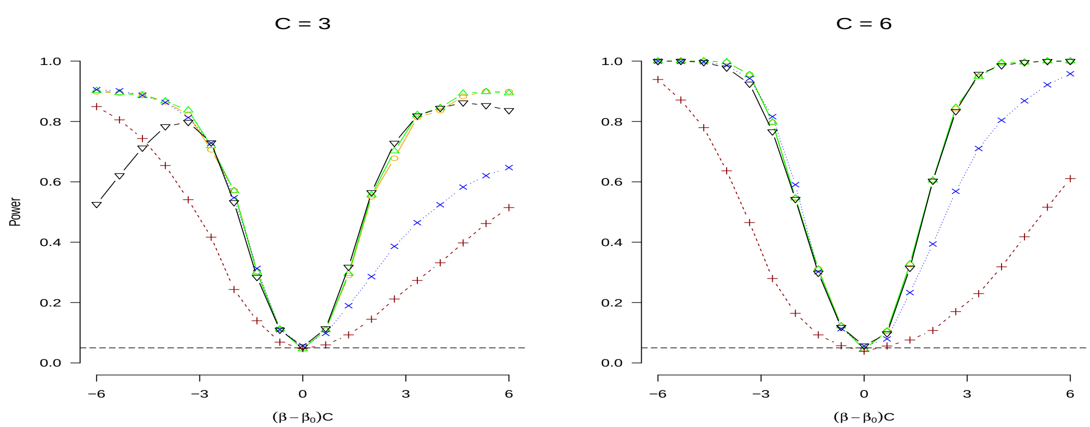

Figure 34: Power Curve for $\rho = 0.4$, $p_1 = 0.1$, and $p_2 = 1.5$

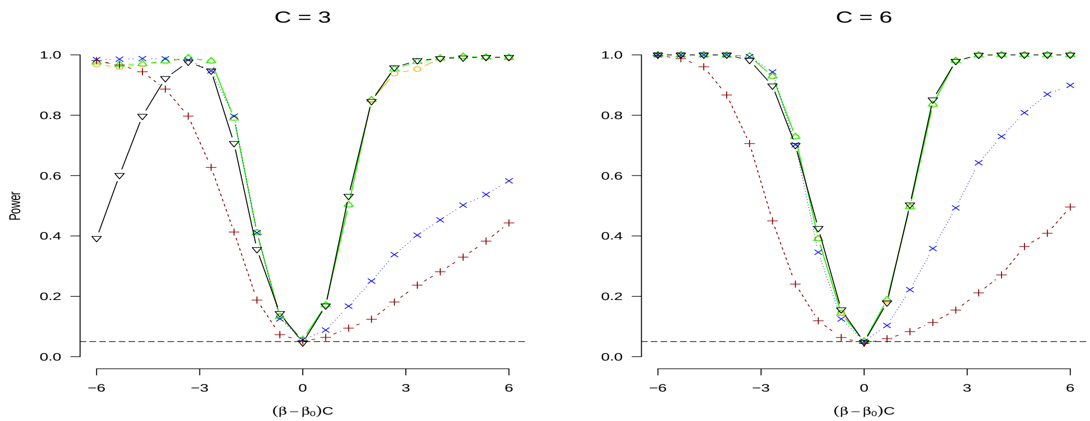

Figure 35: Power Curve for $\rho = 0.7$, $p_1 = 0.1$, and $p_2 = 1.5$

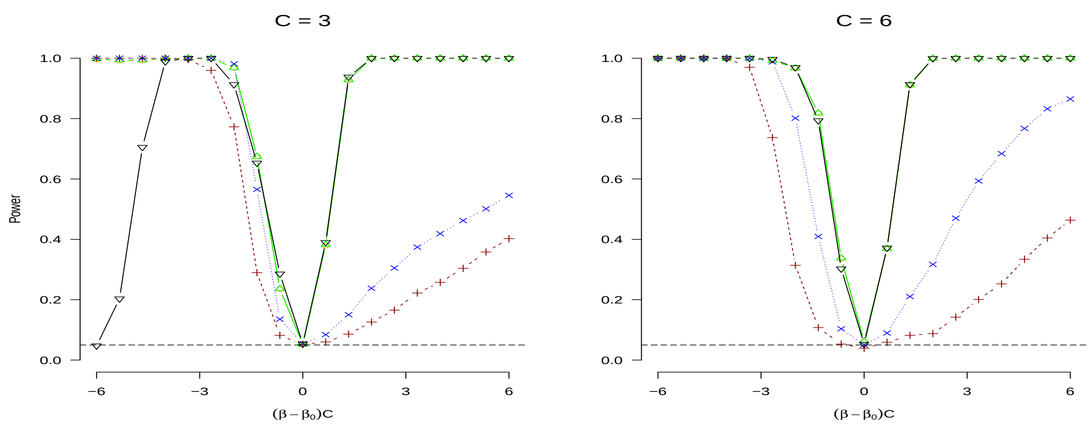

Figure 36: Power Curve for $\rho = 0.9$, $p_1 = 0.1$, and $p_2 = 1.5$

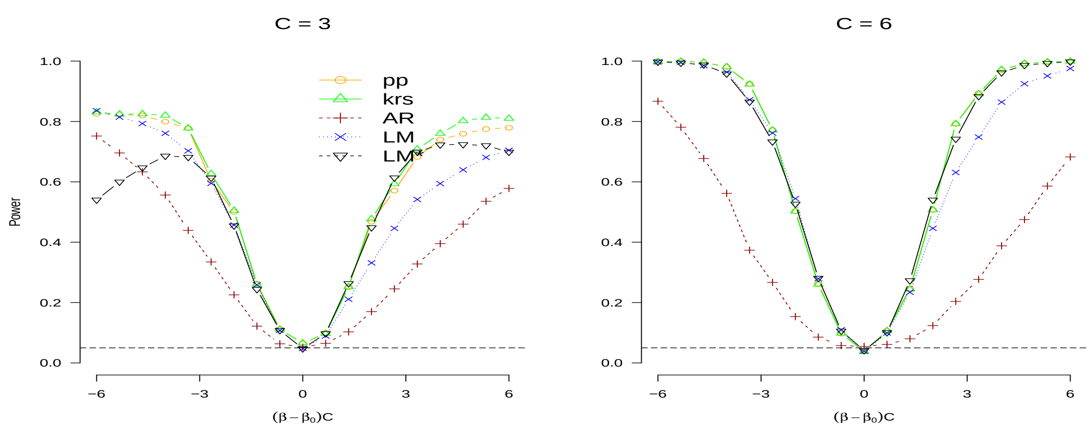

Figure 37: Power Curve for $\rho = 0.2$, $p_1 = 0.1$, and $p_2 = 2$

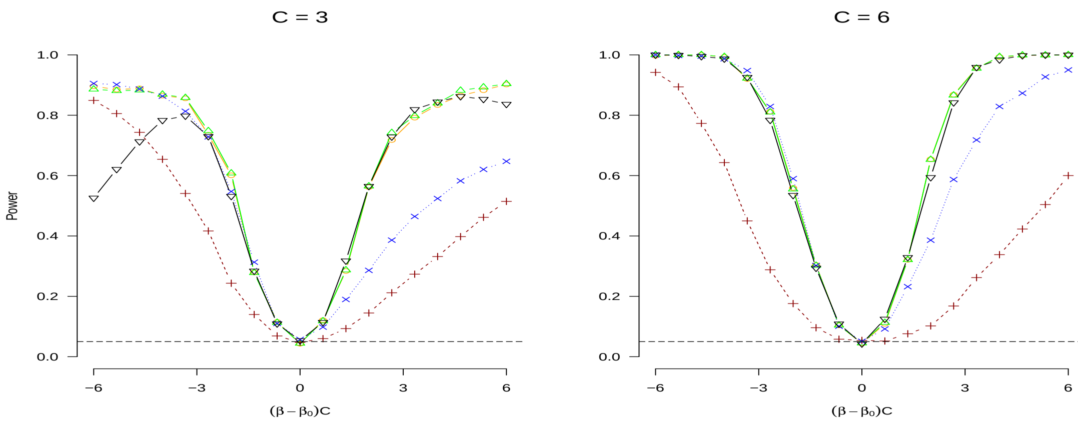

Figure 38: Power Curve for $\rho = 0.4$, $p_1 = 0.1$, and $p_2 = 2$

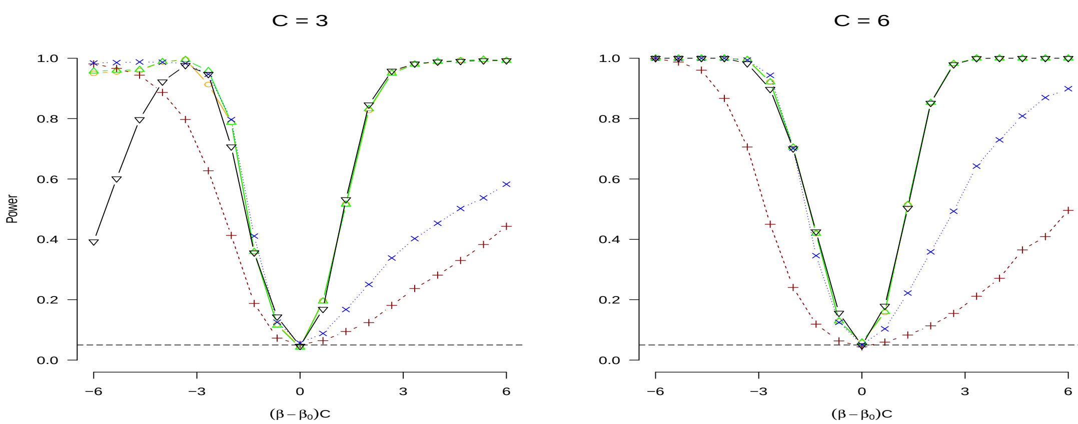

Figure 39: Power Curve for $\rho = 0.7$, $p_1 = 0.1$, and $p_2 = 2$

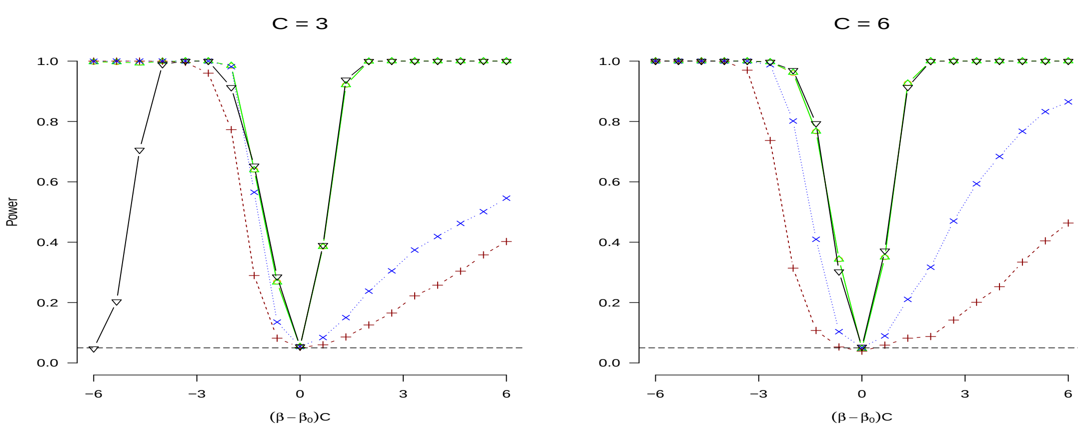

Figure 40: Power Curve for $\rho = 0.9$, $p_1 = 0.1$, and $p_2 = 2$

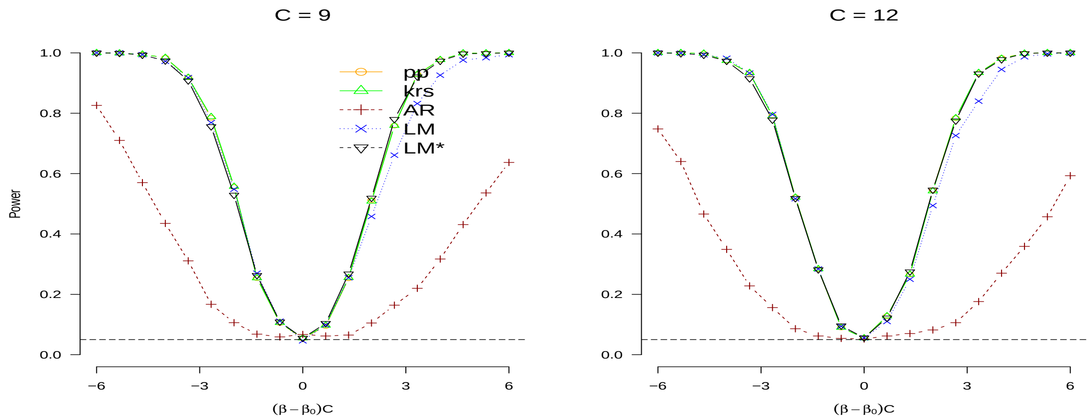

Figure 41: Power Curve for $\rho = 0.2$ with $C = 9$ or 12

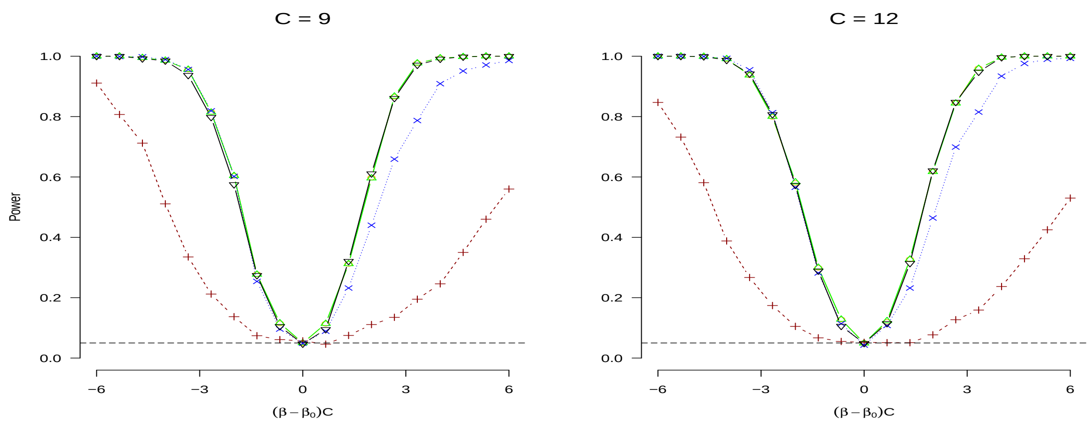

Figure 42: Power Curve for $\rho = 0.4$ with $C = 9$ or 12

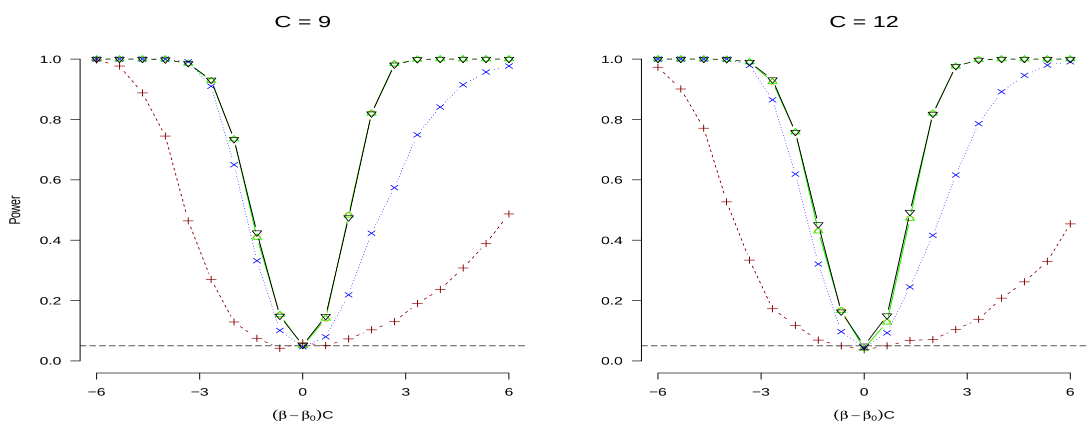

Figure 43: Power Curve for $\rho = 0.7$ with $C = 9$ or 12

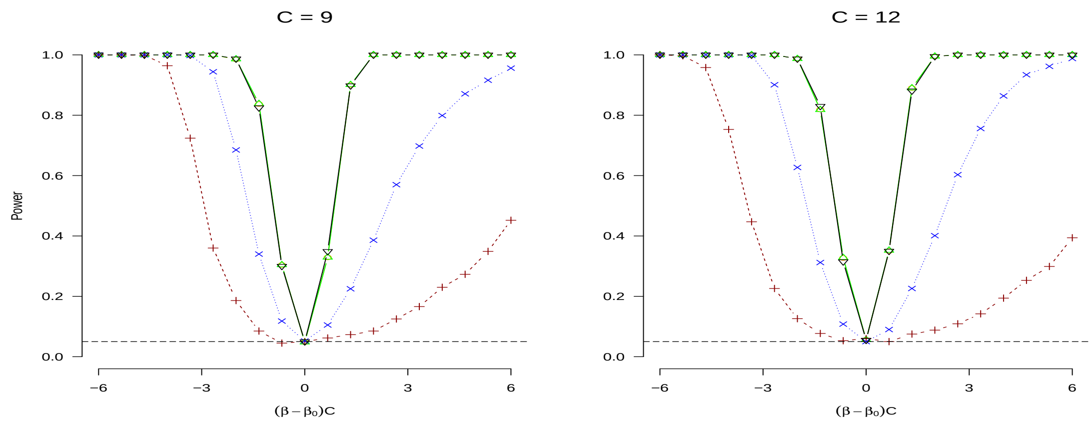

Figure 44: Power Curve for $\rho = 0.9$ with $C = 9$ or 12

U.2 Additional Simulation Results Based on the Calibrated Data

We run two sets of robustness checks. 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 45–52 are results for DGP 1, while Figures 53–60 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 61 and 62, while DGP 2 is reported in Figures 63 and 64.

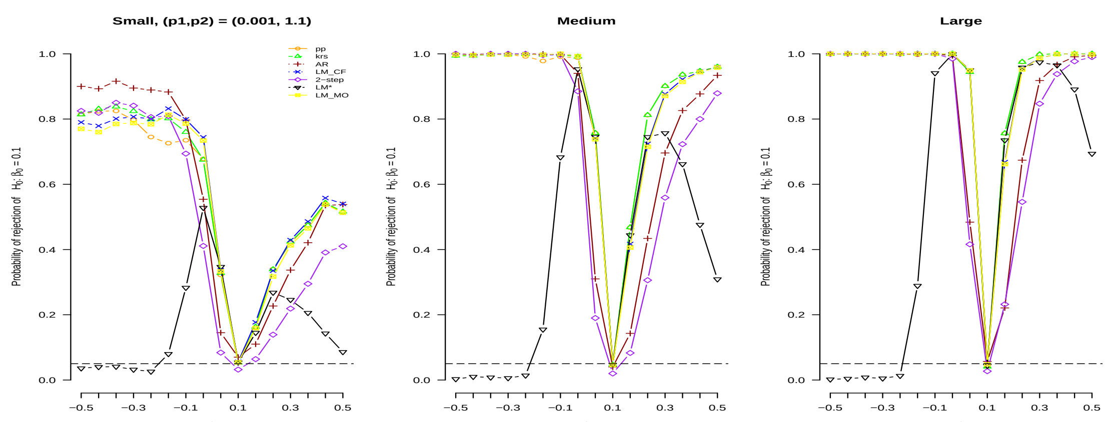

Figure 45: Power Curve for DGP 1 with $(p_1, p_2) = (0.001, 1.1)$

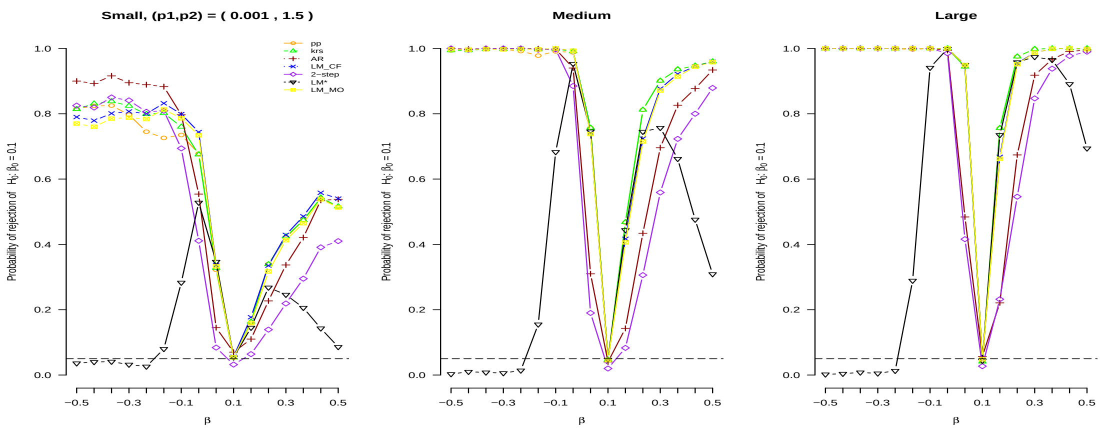

Figure 46: Power Curve for DGP 1 with $(p_1, p_2) = (0.001, 1.5)$

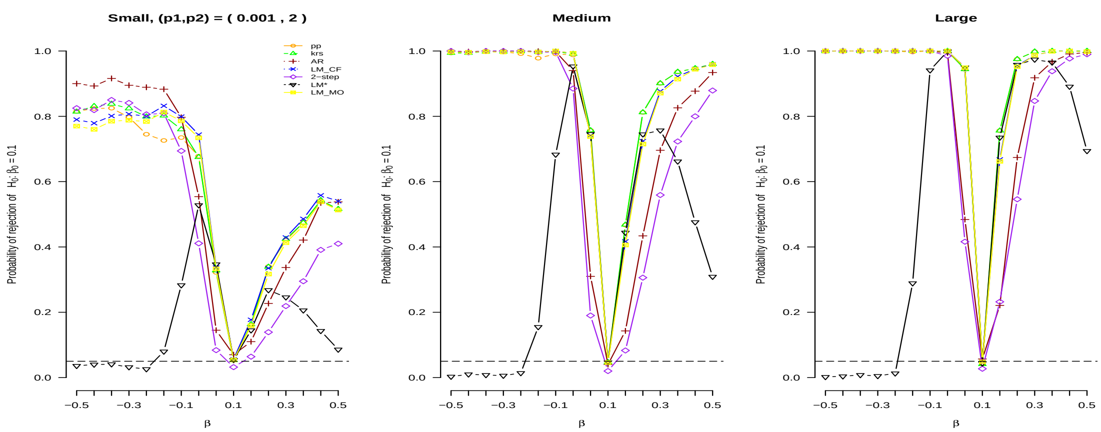

Figure 47: Power Curve for DGP 1 with $(p_1, p_2) = (0.001, 2)$

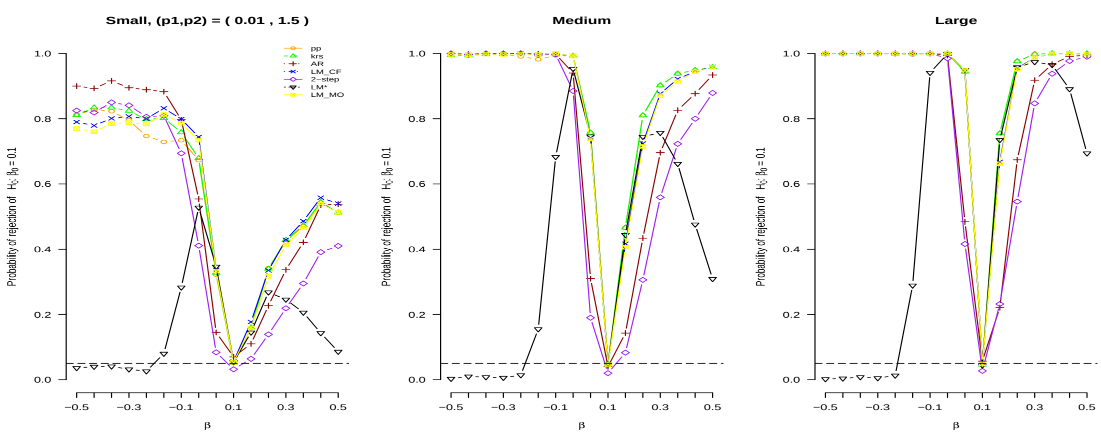

Figure 48: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 1.5)$

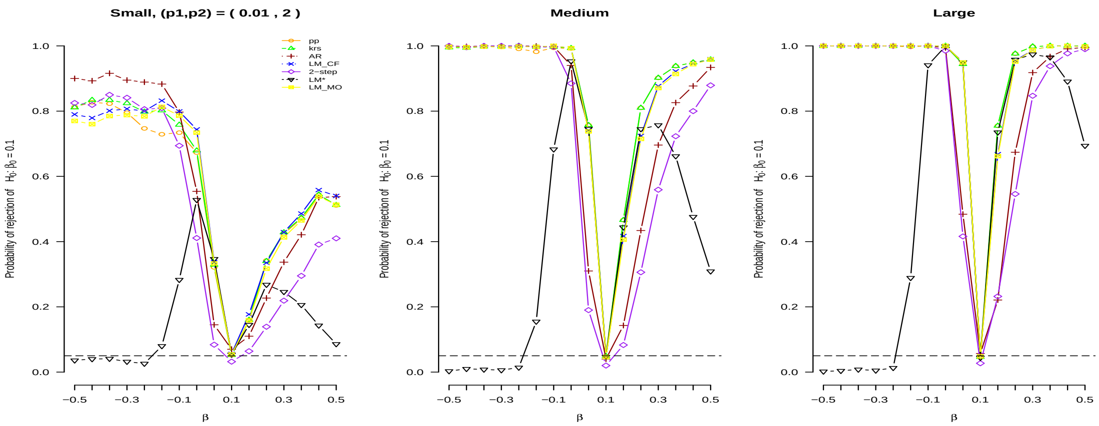

Figure 49: Power Curve for DGP 1 with $(p_1, p_2) = (0.01, 2)$

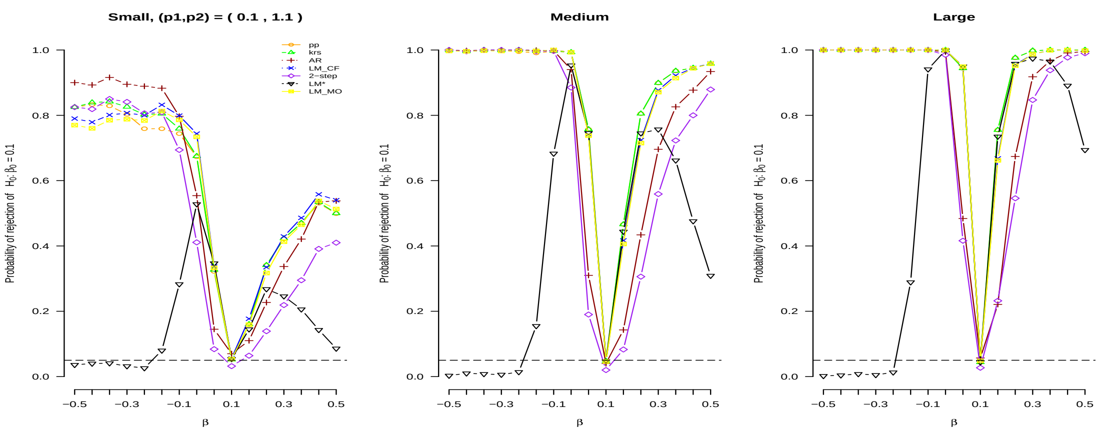

Figure 50: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 1.1)$

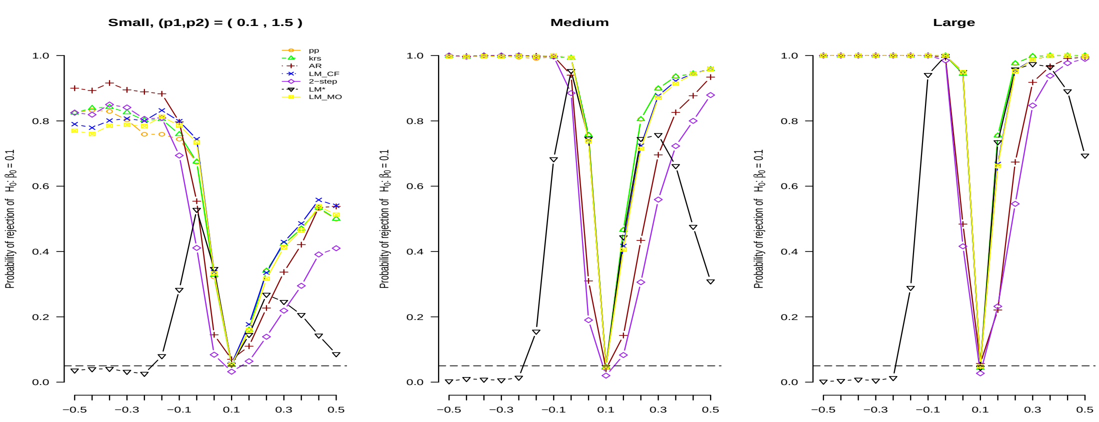

Figure 51: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 1.5)$

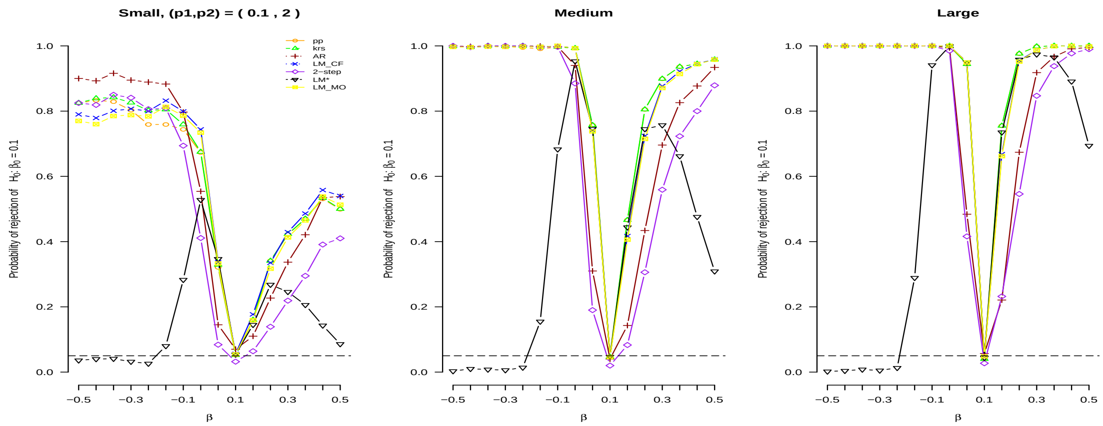

Figure 52: Power Curve for DGP 1 with $(p_1, p_2) = (0.1, 2)$

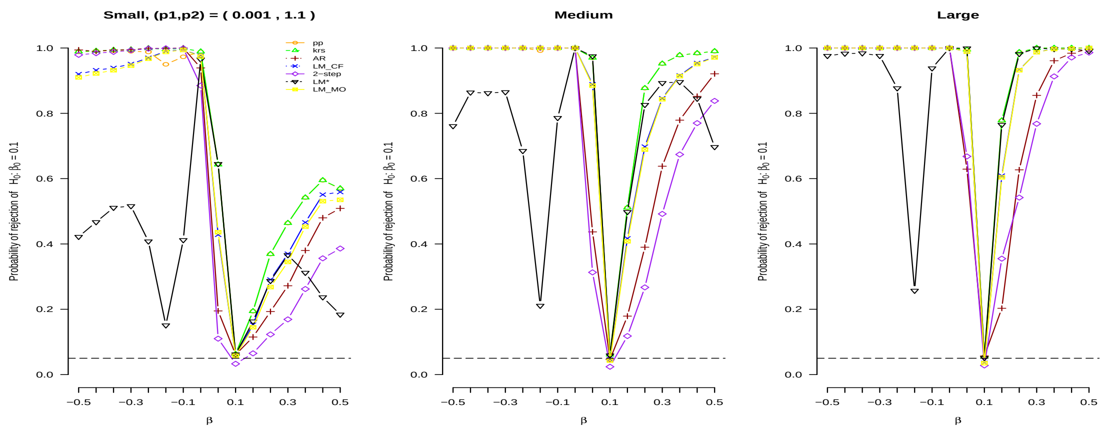

Figure 53: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 1.1)$

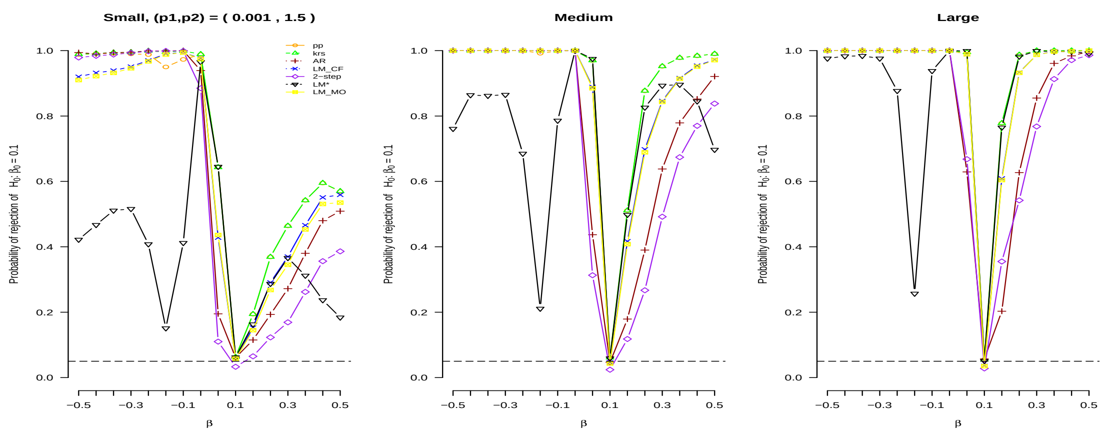

Figure 54: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 1.5)$

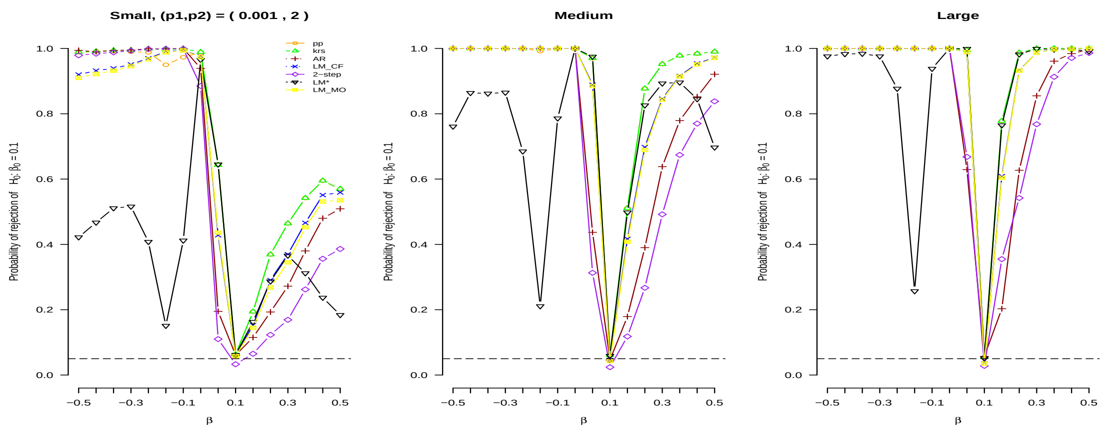

Figure 55: Power Curve for DGP 2 with $(p_1, p_2) = (0.001, 2)$

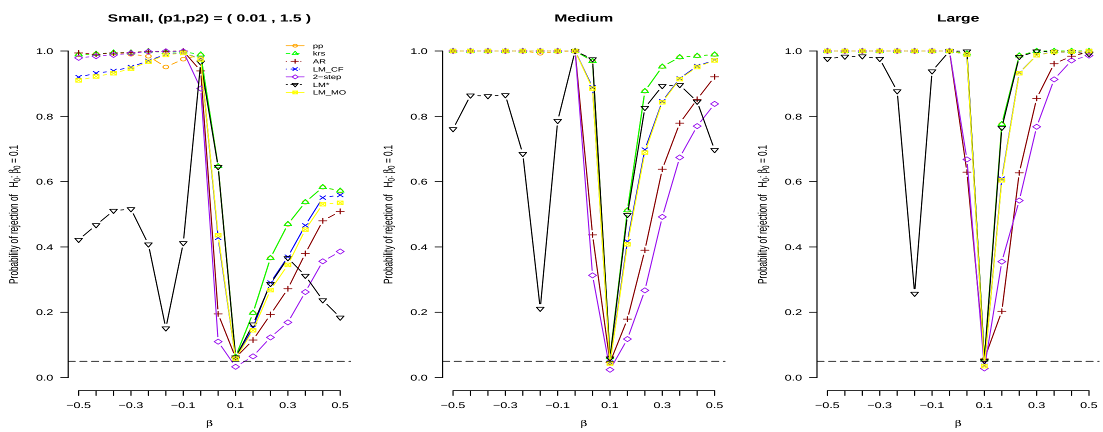

Figure 56: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 1.5)$

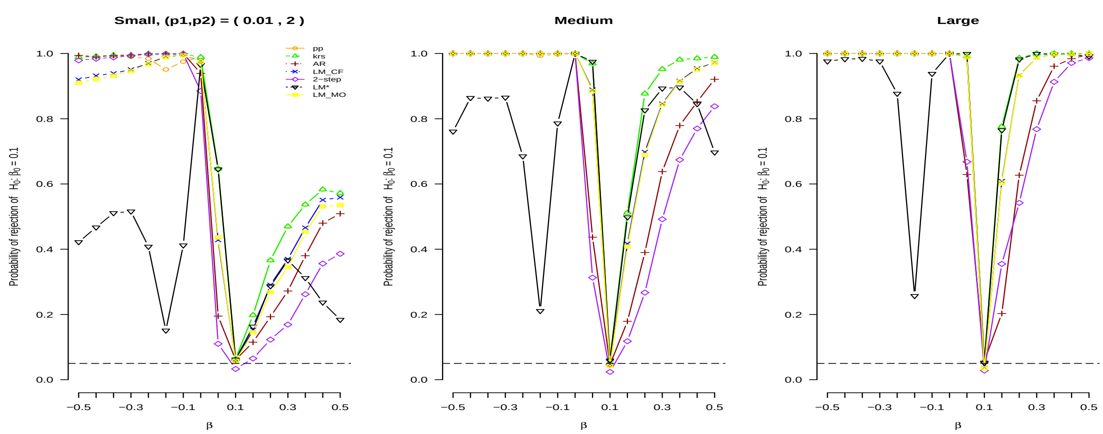

Figure 57: Power Curve for DGP 2 with $(p_1, p_2) = (0.01, 2)$

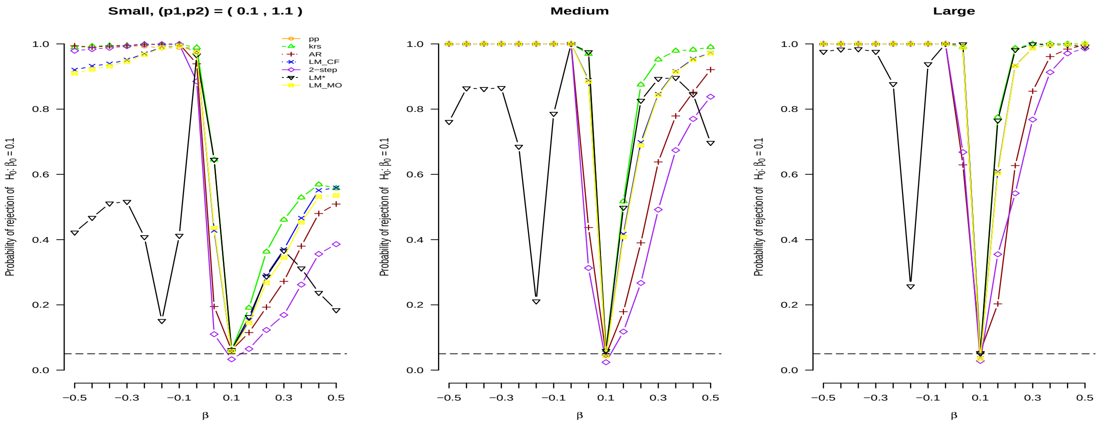

Figure 58: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 1.1)$

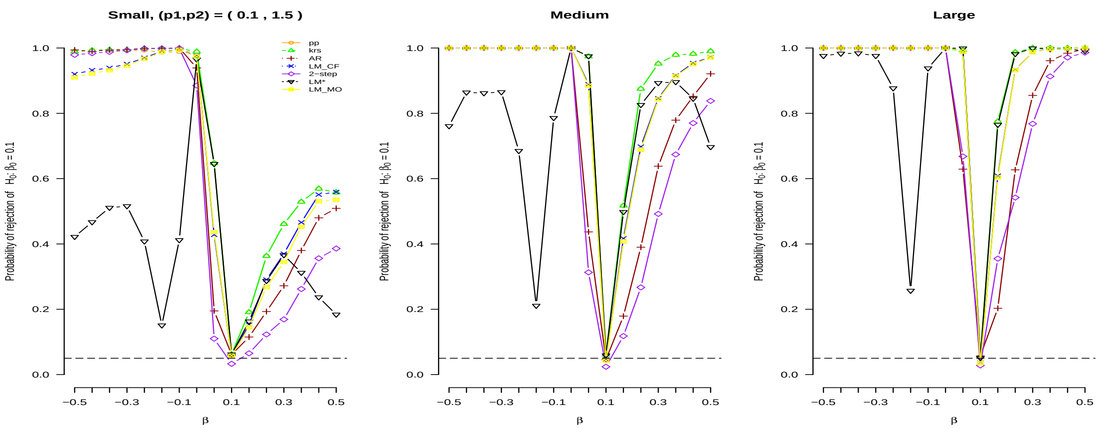

Figure 59: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 1.5)$

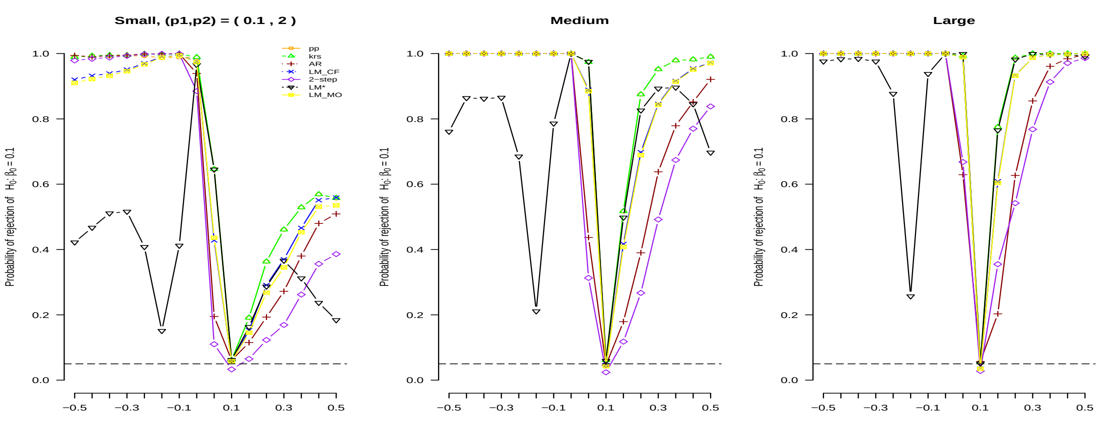

Figure 60: Power Curve for DGP 2 with $(p_1, p_2) = (0.1, 2)$

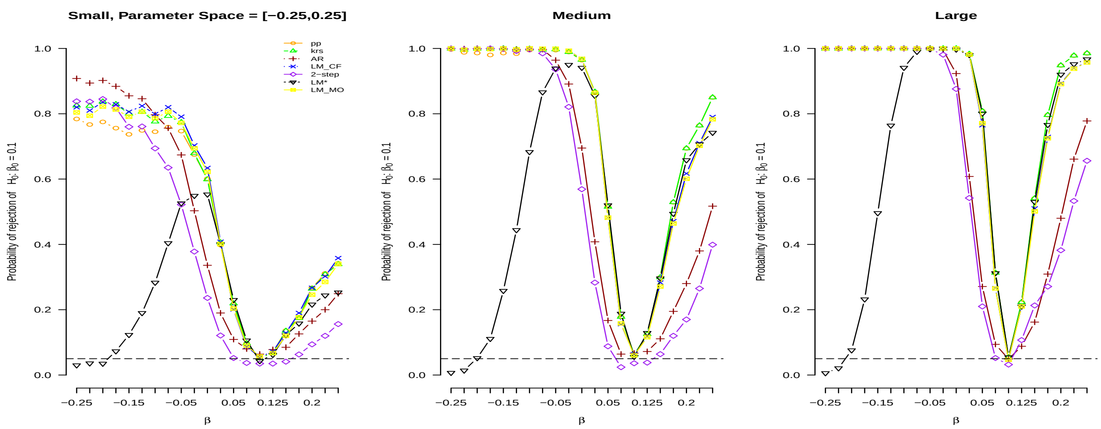

Figure 61: Power Curve for DGP 1 with Parameter Space = $[-0.25, 0.25]$

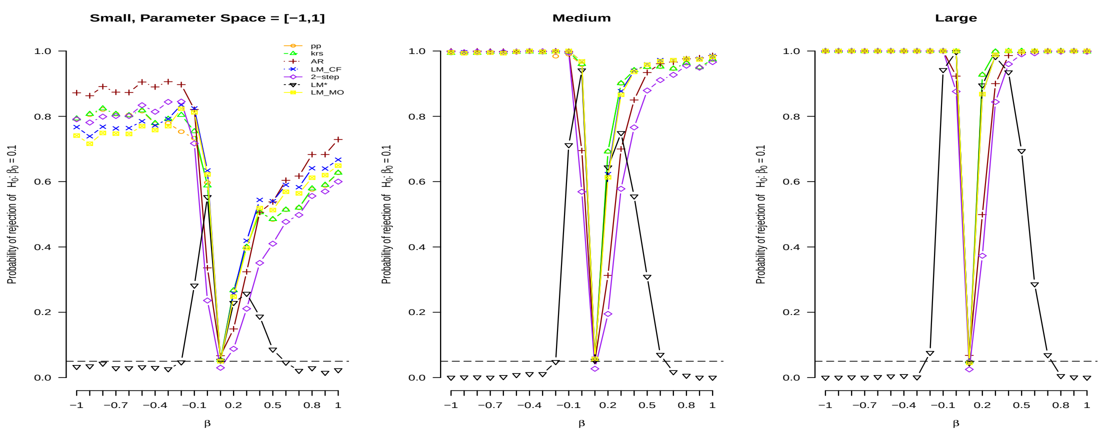

Figure 62: Power Curve for DGP 1 with Parameter Space = $[-1, 1]$

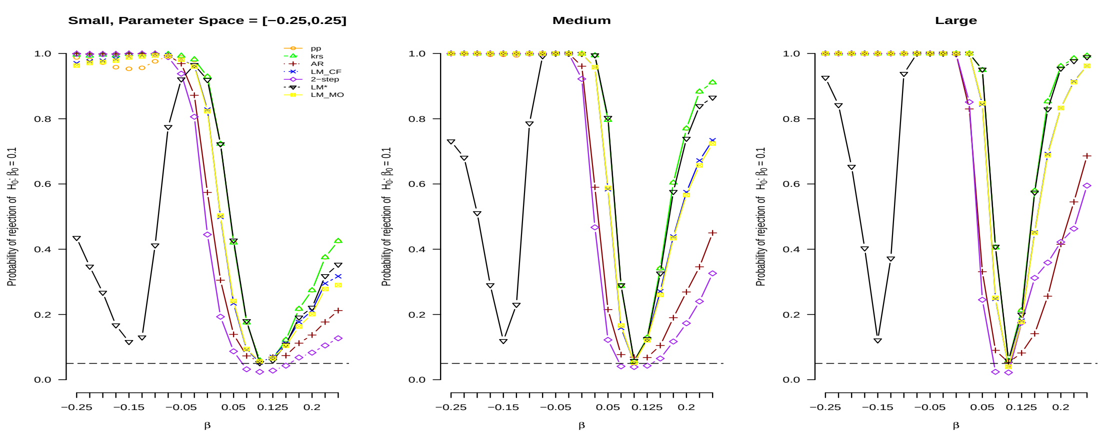

Figure 63: Power Curve for DGP 2 with Parameter Space = $[-0.25, 0.25]$

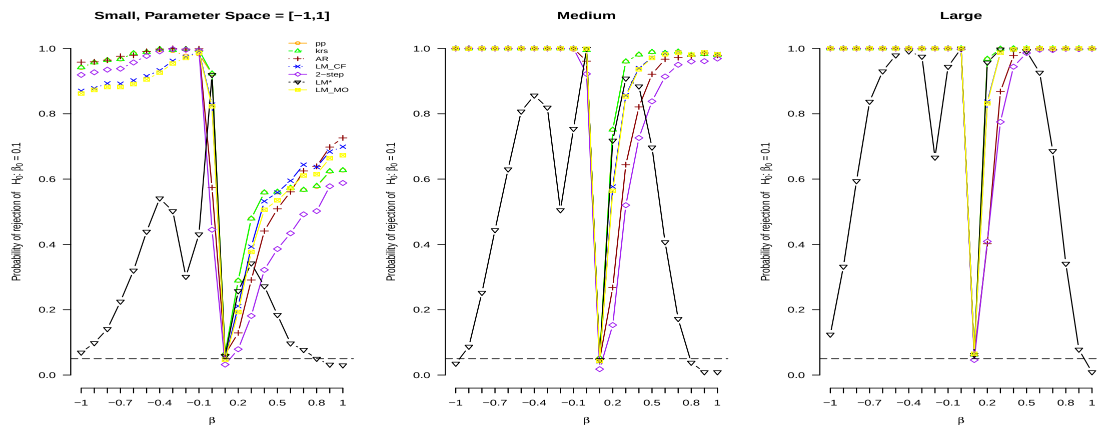

Figure 64: Power Curve for DGP 2 with Parameter Space = $[-1, 1]$

V 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 3. 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 3: 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 4 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 4: Confidence Intervals under $(p_1, p_2) = (0.01, 1.1)$ with varying Parameter Space $\mathcal{B}_2$ and $\mathcal{B}_3$

References

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.

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 (2020). 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. 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.

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 (2023). Jackknife estimation of a cluster-sample iv regression model with many weak instruments. *Journal of Econometrics*.

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.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

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.

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.

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 112*(10), 3260–90.

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

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 (2021). Jackknife empirical likelihood: small bandwidth, sparse network and high-dimensional asymptotics. *Biometrika 108*(3), 661–674.

Matsushita, Y. and T. Otsu (2022). Jackknife lagrange multiplier test with many weak instruments. *Econometric Theory forthcoming.*

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.

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.

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.

Wang, W. and M. Kaffo (2016). Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics 192*(1), 231–268.