Applications of cross-fit variance estimator for testing model specification, overidentification, and structural parameter hypotheses

Yukitoshi Matsushita, Taisuke Otsu and Keita Sunada

July 29, 2024

KEO Discussion Paper No. 181

KEO Discussion Paper No. 181

APPLICATIONS OF CROSS-FIT VARIANCE ESTIMATOR FOR TESTING MODEL SPECIFICATION, OVERIDENTIFICATION, AND STRUCTURAL PARAMETER HYPOTHESES

YUKITOSHI MATSUSHITA, TAISUKE OTSU, AND KEITA SUNADA

ABSTRACT. This paper advocates that power improvement based on the cross-fit variance estimator, proposed by Mikusheva and Sun (2022), is generally applicable to other econometric inference problems, where the statistics of interest are constructed by quadratic forms. We consider consistent specification testing for regression models, overidentifying restriction testing for linear instrumental variable regression models with many weak instruments, and parameter hypothesis testing with many weak instruments, develop the cross-fit variance estimators for those test statistics, and show that the resulting test statistics exhibit improved power properties. Numerical examples illustrate attractive finite sample properties of our cross-fitting approach.

1. INTRODUCTION

In a seminal paper, Newey and Robins (2018) introduced the cross-fitting approach to estimate functionals of nuisance functions, which cover various semiparametric econometric problems, to achieve faster remainder rates. Their key idea is to estimate the nuisance functions by using leave-one-out estimators, which eliminates "own observation" bias components contained in the conventional plug-in semiparametric estimators. Kline, Saggio and Sølvsten (2020) employed the cross-fitting approach to estimate quadratic forms of slope parameters for regression models with unrestricted heteroskedasticity, and showed that their cross-fit estimator exhibits excellent theoretical and finite sample properties even when the number of regressors grows in proportion to the number of observations. In a recent insightful paper, Mikusheva and Sun (2022) adopted the cross-fitting approach to estimate a variance component for their Anderson-Rubin type statistic to conduct inference on structural parameters in linear instrumental variable regression models with many weak instruments. Notable findings by Mikusheva and Sun (2022) are that the "own observation" bias term will inflate the limit of the conventional variance component under the alternative hypothesis and that their cross-fit variance estimator can circumvent such inflated variance to improve power properties of their Anderson-Rubin type test.

This paper advocates that Mikusheva and Sun's (2022) idea of power improvement by the cross-fit variance estimator is generally applicable to other econometric inference problems, where the statistics of interest are constructed by quadratic forms. In particular, we consider (i) consistent specification testing for regression models, (ii) overidentifying restriction testing for linear instrumental variable regression models with many weak instruments, and (iii) parameter hypothesis testing with many weak instruments, develop the cross-fit variance estimators for those test statistics, and show that the resulting test statistics exhibit improved power properties. For (i) and (ii), we consider the test statistics by Sun and Li (2006) and Chao *et al.* (2014), respectively, and conduct analogous bias corrections as Kline, Saggio and Sølvsten (2020) and Mikusheva and Sun (2022). For (iii), which is also studied by Mikusheva and Sun (2022), we consider the jackknife Lagrange multiplier (JLM) test statistic by Matsushita and Otsu (2022). Indeed, due to its more complicated form, the bias correction approach by Mikusheva and Sun (2022) is not directly applicable, and we develop an alternative cross-fit variance estimator by introducing an additional leave-one-out operation.

This paper is organized as follows. Section 2 studies the cross-fit variance estimation for specification testing. In Section 3, we consider instrumental variable regression with many weak instruments for testing overidentifying restrictions (Section 3.1) and parameter hypothesis testing (Section 3.2). Sections 4 and 5 illustrate finite sample properties of our cross-fitting approach.

2. Specification test for regression model

We first consider consistent specification testing of regression models. Let $\{y_i, x_i\}_{i=1}^n$ be a random sample of $(y, x) \in \mathbb{R} \times \mathbb{R}^G$. We wish to test whether the linear regression model is correctly specified:

$$H_0 : \mathbb{P}\{\mathbb{E}[y|x] = \gamma_0' x\} = 1 \quad \text{for some } \gamma_0 \in \Gamma,$$

against $H_1 : H_0$ is false. This section considers a consistent specification test proposed by Sun and Li (2006), which is a modified version of Hong and White (1995) to avoid a non-zero centering term. Let $b(x)$ be a K-dimensional vector of basis functions of $x \in \mathbb{R}^G$, where G is fixed but $K = K_n$ grows with the sample size n. Define an $n \times K$ matrix $B = (b(x_1), \ldots, b(x_n))'$ and $P = B(B'B)^{-1}B'$ whose (i,j)-th element is denoted by P_{ij}. Let $\hat{\gamma}$ be a consistent estimator of γ_0 under H_0 (typically the OLS estimator) and $\hat{u}_i = y_i - \hat{\gamma}' x_i$ be the residual. The specification test statistic by Sun and Li (2006) is written as

$$T = \frac{K^{-1/2} \sum_{i=1}^n \sum_{j \neq i} P_{ij} \hat{u}_i \hat{u}_j}{\sqrt{\hat{V}}}, \tag{2.1}$$

where $\hat{V}$ is an estimator of the variance $\mathbb{V}(K^{-1/2} \sum_{i=1}^n \sum_{j \neq i} P_{ij} \hat{u}_i \hat{u}_j)$ of the numerator defined as

$$\hat{V} = \frac{2}{K} \sum_{i=1}^n \sum_{j \neq i} P_{ij}^2 \hat{u}_i^2 \hat{u}_j^2.$$

This test statistic is constructed based on the fact that $\mathbb{E}[e\mathbb{E}[e|x]] \geq 0$ for $e = y - \mathbb{E}[y|x]$ with equality holding if and only if H_0 is true. The numerator of T is a sample analog of $\mathbb{E}[e\mathbb{E}[e|x]]$ using a series estimator for the conditional mean $\mathbb{E}[e|x]$. Sun and Li (2006) showed that $T \xrightarrow{d} N(0,1)$ under H_0, and $T \xrightarrow{p} +\infty$ under H_1. Moreover, they established the consistency of the variance estimator $\hat{V} - V \xrightarrow{p} 0$ under H_0, where $V = \frac{2}{K} \sum_{i=1}^n \sum_{j \neq i} P_{ij}^2 \mathbb{E}[e_i^2|x_i]\mathbb{E}[e_j^2|x_j]$.

The naive variance estimator $\hat{V}$ is a natural sample counterpart for the population target V under the null hypothesis because $e_i = y_i - \gamma_0' x_i$ under H_0. However, using $\hat{V}$ may cause some loss in the power property. To see this point, suppose $\hat{\gamma}$ converges in probability to a pseudo true value γ_* under H_1, and the conditional mean $\theta_0(x) = \mathbb{E}[y|x]$ is well approximated by a basis expansion $\alpha_*' b(x)$ for the projection coefficients $\alpha_* = \mathbb{E}[b(x)b(x)']^{-1}\mathbb{E}[b(x)y]$. The residual $\hat{u}_i$ can be decomposed as

$$\hat{u}_i = \Delta_i + e_i + \rho_i,$$

where $\Delta_i = \alpha_*' b(x_i) - \gamma_*' x_i$ is a drift component under the alternative H_1, $e_i = y_i - \theta_0(x_i)$ is a mean zero stochastic error, and $\rho_i = \theta_0(x_i) - \alpha_*' b(x_i) - (\hat{\gamma} - \gamma_*)' x_i$ is an asymptotically negligible component under suitable conditions. Based on this decomposition, the naive variance estimator $\hat{V}$ is written as

$$\hat{V} = \frac{2}{K} \sum_{i=1}^{n} \sum_{j \neq i} P_{ij}^2 (\Delta_i + e_i + \rho_i)^2 (\Delta_j + e_j + \rho_j)^2,$$

which involves a term of order up to $O_p\left(\frac{1}{K}\sum_{i=1}^n \sum_{j\neq i} P_{ij}^2 \Delta_i^2 \Delta_j^2\right)$. Under global misspecification with larger values of Δ_i's, this positive bias term in $\hat{V}$ tends to be large, and the power of Sun and Li's (2006) test using T will deteriorate. One way to reduce this bias component is to employ a cross-fitting approach introduced by Newey and Robins (2018), Kline, Saggio and Sølvsten (2020), and Mikusheva and Sun (2022). In particular, we adapt Mikusheva and Sun's (2022) cross-fit variance estimator for a quadratic form with a double summation to the present setup.

Let $M = I - P$, M_{ij} be the (i,j)-th element of M, and m_i be the i-th column of M. We propose to estimate the variance term V by the following cross-fit estimator:

$$\hat{V}_{\text{cf}} = \frac{2}{K} \sum_{i=1}^{n} \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} \{\hat{u}_i (m_i' \hat{u})\}\{\hat{u}_j (m_j' \hat{u})\},$$

where $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_n)'$. It should be noted that the matrices P and M are constructed by the regressor matrix B for series estimation. The division by $M_{ii}M_{jj} + M_{ij}^2$ is due to the fact that $\mathbb{E}[e_i(m_i'e)e_j(m_j'e)|X] = (M_{ii}M_{jj} + M_{ij}^2)\mathbb{E}[e_i^2|x_i]\mathbb{E}[e_j^2|x_j]$ under independence of $\{e_i\}_{i=1}^n$ conditionally on $X = (x_1, \ldots, x_n)'$. The following theorem establishes consistency of $\hat{V}_{\text{cf}}$ for V.

Theorem 1. *Consider the setup of this section. Assume that (i) for every K, there is a nonsingular matrix Q such that the smallest eigenvalue of $\mathbb{E}[Qb(x)b(x)'Q']$ is bounded away from zero uniformly in K, (ii) there is a sequence of constants $\{\zeta_K\}$ satisfying $\sup_{x\in\mathbb{R}^G} ||b(x)|| \leq \zeta_K$ for all K such that $\zeta_K^2 K/n \to 0$ and $K \to \infty$ as $n \to \infty$, (iii) $\{y_i, x_i\}_{i=1}^n$ is iid and $\mathbb{E}[||(y_i, x_i)||^6] < \infty$, (iv) $\{e_i\}_{i=1}^n$ is mutually independent conditional on X, and (v) $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$ under H_0.*

Then under H_0,

$$\hat{V}_{\text{cf}} - V \xrightarrow{p} 0.$$

Assumptions (i) and (ii) are standard assumptions employed by Sun and Li (2006) as well. Sun and Li (2006) required that the fourth conditional moment is bounded, but Assumption (iii) imposes a higher moment condition due to cross-fitting. The mutual independence in Assumption (iv) is needed for the cross-fitting $\mathbb{E}[e_i(m_i'e)e_j(m_j'e)|X]$ to be valid as argued above. Assumption (v) is satisfied with the OLS estimator, for example.

This theorem guarantees that $T_{\text{cf}} = \frac{K^{-1/2}\sum_{i=1}^n \sum_{j\neq i} P_{ij}\hat{u}_i\hat{u}_j}{\sqrt{\hat{V}_{\text{cf}}}} \xrightarrow{d} N(0,1)$ under H_0, i.e., the both statistics T and T_{cf} have the identical asymptotic null distribution. On the other hand, under the alternative hypothesis, the variance estimators $\hat{V}$ and $\hat{V}_{\text{cf}}$ exhibit different asymptotic properties.

The rationale of our variance estimator $\hat{V}_{\text{cf}}$ can be explained as follows. Consider again the decomposition $\hat{u}_i = \Delta_i + e_i + \rho_i$ with $\Delta_i = \alpha_*' b(x_i) - \gamma_*' x_i$ under the alternative H_1. As far as the basis functions $b(x)$ contain x as a subvector, it holds $m_i'\Delta = 0$ for $\Delta = (\Delta_1, \ldots, \Delta_n)'$, and the cross-fit variance estimator satisfies

$$\hat{V}_{\text{cf}} = \frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i}\frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}\{(\Delta_i + e_i + \rho_i)(m_i'(e+\rho))\}\{(\Delta_j + e_j + \rho_j)(m_j'(e+\rho))\},$$

which only involves a term of order up to $O_p\left(\frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i}\frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}\Delta_i\Delta_j\right)$. Therefore, when Δ_i's take larger values under H_1, $\hat{V}_{\text{cf}}$ tends to be smaller than $\hat{V}$ and the associated test statistic T_{cf} is expected to be more powerful than T.

3. Statistical inference with many weak instruments

In this section, we redefine the notation and consider the instrumental variable regression model:

$$\begin{aligned} y_i &= \beta' x_i + u_i, \\ x_i &= \Pi' z_i + v_i, \end{aligned} \tag{3.1}$$

for $i = 1, \ldots, n$, where y_i is a scalar dependent variable, x_i is a G-dimensional vector of endogenous regressors, z_i is a K-dimensional vector of instruments, u_i and v_i are error terms, and β and Π are $G\times 1$ and $K \times G$ dimensional parameters, respectively. We are concerned with the setup of many weak instruments, where Π may decay with the sample size n and the number of instruments K may grow with n (although we suppress dependence of Π and K on n). Let $Z = (z_1, \ldots, z_n)'$, $P = Z(Z'Z)^{-1}Z'$, $M = I - P$, and P_{ij} and M_{ij} be the (i,j)-th element of P and M, respectively.

3.1. **Overidentifying restriction test.** We first study the overidentifying restriction test statistic proposed by Chao *et al.* (2014):

$$J = \frac{\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}\hat{u}_i\hat{u}_j}{\sqrt{\hat{\Phi}}} + K,$$

where $\hat{u}_i = y_i - \hat{\beta}' x_i$, $\hat{\beta}$ is an estimator for β, and

$$\hat{\Phi} = \frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^2\hat{u}_i^2\hat{u}_j^2.$$

The form of the test statistic J is close to that of the specification test statistic T we saw in the previous section. Chao *et al.* (2014) showed that $\hat{\Phi}$ is a consistent estimator for the variance component of the numerator in J, which is $\Phi = \frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^2\mathbb{E}[u_i^2]\mathbb{E}[u_j^2]$.

Let m_i be the i-th column of M. In this setup, we propose the following cross-fit variance estimator for Φ:

$$\hat{\Phi}_{\text{cf}} = \frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i}\frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}\{\hat{u}_i(m_i'\hat{u})\}\{\hat{u}_j(m_j'\hat{u})\}.$$

This estimator is similar to the cross-fit variance estimator proposed by Mikusheva and Sun (2022) for parameter hypotheses testing. In contrast to theirs, we need to use the residual $\hat{u}_i = y_i - \hat{\beta}' x_i$ instead of $u_i = y_i - \beta_0' x_i$ for the hypothetical value β_0 to be tested.

The consistency of our cross-fit variance estimator $\hat{\Phi}_{\mathrm{cf}}$ is obtained as follows.

Theorem 2. *Assume that (i) $\{u_i, v_i\}_{i=1}^n$ is a sequence of independent random variables with mean zero, and $\{z_i\}_{i=1}^n$ is a non-random sequence, (ii) there exists a constant C such that $\max_i \|\Pi' z_i\|^2 \leq C$, $\max_i \mathbb{E}[|v_i|^6] \leq C$, and $\max_i \mathbb{E}[|u_i|^6] \leq C$, (iii) there exists a constant δ such that $P_{ii} \leq \delta < 1$ for all i, and (iv) $\hat{\beta} \xrightarrow{p} \beta$.*

Then under the model in (3.1),

$$\hat{\Phi}_{\mathrm{cf}} - \Phi \xrightarrow{p} 0.$$

Assumptions (i) and (iii) are standard. Mikusheva and Sun (2022) do not impose $\max_i \|\Pi' z_i\|^2 \leq C$, but we need this in Assumption (ii) because our statistic J involves an estimator $\hat{\beta}$ instead of the fixed hypothetical value. Assumption (iv) is satisfied with the JIVE, HLIML, and HFUL estimators by Hausman *et al.* (2012) under suitable conditions (see their Theorem 1).

To see how the cross-fitting works for $\hat{\Phi}_{\mathrm{cf}}$, notice that the residual is written by $\hat{u}_i = (z_i'\Pi + v_i')\Delta + u_i$ with $\Delta = \beta - \hat{\beta}$, and the variance estimator proposed by Chao *et al.* (2014) is written as

$$\hat{\Phi} = \frac{1}{K} \sum_{i=1}^n \sum_{j \neq i} P_{ij}^2 \{(z_i'\Pi + v_i')\Delta + u_i\}^2 \{(z_j'\Pi + v_j')\Delta + u_j\}^2,$$

which involves a term of order up to $O_p\left(\frac{1}{K}\sum_{i=1}^n \sum_{j\neq i} P_{ij}^2 (z_i'\Pi\Delta)^2 (z_j'\Pi\Delta)^2\right)$. In contrast, our proposed cross-fit variance estimator is written as

$$\hat{\Phi}_{\mathrm{cf}} = \frac{1}{K} \sum_{i=1}^n \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} \{((z_i'\Pi + v_i')\Delta + u_i)(m_i'(V\Delta + u))\}\{((z_j'\Pi + v_j')\Delta + u_j)(m_j'(V\Delta + u))\},$$

which involves a term of order up to $O_p\left(\frac{1}{K}\sum_{i=1}^n \sum_{j\neq i} \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} (z_i'\Pi\Delta)(z_j'\Pi\Delta)\right)$. Notice that Δ does not converge to 0 in probability under the alternative hypothesis of Chao *et al.* (2014). If the degree of misspecification is large, which results in large $|\Delta|$, $\hat{\Phi}_{\mathrm{cf}}$ tends to be smaller than $\hat{\Phi}$. Hence the test statistic based on $\hat{\Phi}_{\mathrm{cf}}$ is expected to be more powerful than J.

3.2. **Parameter hypothesis test.** We next consider the JLM test statistic by Matsushita and Otsu (2022) for testing the parameter hypothesis $H_0 : \beta = b$ against $H_1 : \beta \neq b$, that is

$$S = (u_0' P^* X)\hat{\Psi}^{-1}(X' P^* u_0),$$

where $u_0 = (u_{01}, \ldots, u_{0n})'$, $u_{0i} = y_i - x_i' b$, P^* is defined as $P_{ij}^* = P_{ij}$ for $i \neq j$ and $P_{ii}^* = 0$ for all i, and

$$\hat{\Psi} = X' P^* \Sigma_0 P^* X + \sum_{i=1}^n \sum_{j=1}^n x_i x_j' u_{0i} u_{0j} P_{ij}^{*2},$$

with $\Sigma_0 = \mathrm{diag}(u_{01}^2, \ldots, u_{0n}^2)$. As shown in Matsushita and Otsu (2022), $\hat{\Psi}$ is a valid estimator for the variance component

$$\Psi = \sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}[x_i P_{ik} u_{0k}^2 P_{kj} x_j' | Z] + \sum_{i=1}^{n} \sum_{j\neq i} \mathbb{E}[x_i x_j' u_{0i} u_{0j} P_{ij}^2 | Z],$$

to guarantee $S \xrightarrow{d} \chi_G^2$ under H_0. It should be noted that the variance term Ψ takes a more complicated form than the ones in the previous section or Mikusheva and Sun (2022), so it is not trivial how to construct a cross-fit estimator for Ψ.

As an unbiased estimator of Ψ, we propose the following cross-fit variance estimators

$$\begin{aligned}
\hat{\Psi}_{\mathrm{cf},1} &= \sum_{i,j,k,i\neq k,j\neq k}^{n} x_i P_{ik} \frac{\tilde{u}_{0k} u_{0k}}{M_{kk}} P_{kj} x_j' + \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j' u_{0i} u_{0j} P_{ij}^{*2}, \\
\hat{\Psi}_{\mathrm{cf},2} &= \sum_{i,j,k,i\neq k,j\neq k}^{n} x_i P_{ik} \frac{\tilde{u}_{0k} u_{0k}}{M_{kk}} P_{kj} x_j' + \sum_{i=1}^{n} \sum_{j\neq i} x_i \tilde{x}_j' \frac{\bar{u}_{0i}^{(j)} u_{0j}}{M_{ii} M_{jj}} P_{ij}^2,
\end{aligned}$$

where $\tilde{u}_{0k} = \sum_{l=1}^{n} M_{kl} u_{0l}$, $\tilde{x}_j' = \sum_{k=1}^{n} M_{jk} x_k'$, and $\bar{u}_{0i}^{(j)} = \sum_{m\neq j} M_{im} u_{0m}$. The difference between $\hat{\Psi}_{\mathrm{cf},1}$ and $\hat{\Psi}_{\mathrm{cf},2}$ is that $\hat{\Psi}_{\mathrm{cf},2}$ has the "leave-one-out" cross-fit term $\bar{u}_{0i}^{(j)}$, which does not appear in Mikusheva and Sun's (2022) cross-fit variance estimator for their heteroskedasticity robust Anderson-Rubin type statistic. Indeed a "usual" cross-fit estimator using $\tilde{u}_i$ instead of $\bar{u}_{0i}^{(j)}$ is biased because the moments of $v_i^2 u_{0j}^2$ remain for $i \neq j$. Our leave-one-out cross-fitting successfully removes these terms, and ensures $\hat{\Psi}_{\mathrm{cf},2}$ to be unbiased.

As shown in Appendix C, our cross-fit estimators $\hat{\Psi}_{\mathrm{cf},1}$ and $\hat{\Psi}_{\mathrm{cf},2}$ are unbiased:

$$\mathbb{E}[\hat{\Psi}_{\mathrm{cf},1} | Z] = \mathbb{E}[\hat{\Psi}_{\mathrm{cf},2} | Z] = \Psi. \tag{3.2}$$

Also by adapting the argument in Matsushita and Otsu (2022), we can show that for $q = 1, 2$,

$$S_{\mathrm{cf},q} = (u_0' P^* X) \hat{\Psi}_{\mathrm{cf},q}^{-1} (X' P^* u_0) \xrightarrow{d} \chi_G^2,$$

under H_0.

To see how the cross-fitting works for $\hat{\Psi}_{\mathrm{cf},1}$, notice that we can write $u_{0i} = z_i' \Pi \Delta + \eta_i$, where $\eta_i = u_i + v_i' \Delta$ under the alternative $H_1 : \beta = b + \Delta$. Then the first term of the variance component of S can be written as

$$\sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}[x_i P_{ik} (z_k' \Pi \Delta + \eta_k)^2 P_{kj} x_j' | Z].$$

On the other hand, the first term of the conditional mean $\mathbb{E}[\hat{\Psi}_{\mathrm{cf},1} | Z]$ for our cross-fit estimator is characterized as

$$\sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}\left[x_i P_{ik} \eta_k (z_k' \Pi \Delta + \eta_k) P_{kj} x_j' \middle| Z \right].$$

Then by the similar argument as we saw in the previous sections, when Δ_i's take larger values under H_1, $\hat{\Psi}_{\mathrm{cf},1}$ tends to be smaller than $\hat{\Psi}$ and the resulting test statistic S_{cf} is expected to be more powerful than S. Similar comments apply to $\hat{\Psi}_{\mathrm{cf},2}$.

4. Simulation

4.1. **Specification testing.** We first consider specification testing discussed in Section 2. Based on Tripathi and Kitamura (2003), we consider two data generating processes

$$\begin{aligned} \text{DGP 1} &: y_i = \beta_0 + \beta_1 x_i + u_i, \\ \text{DGP 2} &: y_i = \beta_0 + \beta_1 x_i + \frac{c}{\sqrt{2\pi}} \exp(-(x_i - 2.5)^2) + u_i, \end{aligned}$$

where $\beta_0 = \beta_1 = 1$, $\log x_i \sim N(0, 1)$ with 5% upper tail truncation, and the structural error u_i is specified as $u_i = \varepsilon_i \sqrt{0.1 + 0.2x_i + 0.3x_i^2}$ with $\varepsilon_i \sim N(0, 1)$ independently from x_i. The linear regression model is correctly specified under DGP 1 and misspecified under DGP 2, where we control the magnitude of misspecification by the value of c. The sample size is set as $n = 400$. We employ cubic splines as basis functions $b(x)$, and its dimension is set to $K = \lceil n^{1/3} \rceil = 8$. We focus on the rejection frequencies at a nominal 5% significance level of Sun and Li's (2006) test statistic T in (2.1) and our proposed statistic T_{cf}, which replaces $\hat{V}$ in (2.1) with the cross-fit variance estimator $\hat{V}_{\text{cf}}$. The number of Monte Carlo replications in each experiment is 1000.

The null rejection frequencies for T and T_{cf} are 3.1% and 4.5%, respectively under DGP 1. We investigate the power properties using DGP 2, and the results are reported in Table 1. This table shows that T_{cf} is more powerful than T for different values of c.

Table 1. Estimated powers under DGP 2

c	T	T_{cf}
1.5	9.5%	12.0%
2.0	19.7%	23.9%
2.5	33.3%	35.4%
3.0	47.4%	49.7%
3.5	64.5%	65.5%
4.0	80.5%	81.0%

4.2. **Structural parameter hypothesis testing.** We next consider parameter hypothesis testing on a structural parameter in an instrumental variable regression model discussed in Section 3.2. We mostly follow the simulation design by Matsushita and Otsu (2022). The data generating process is specified as

$$\begin{aligned} y_i &= \beta x_i + u_i, \\ x_i &= z_i' \pi + v_i, \end{aligned}$$

where $\pi = (d, \ldots, d)'$ and $z_i = (z_{1i}, z_{1i}^2, z_{1i}^3, z_{2i}')'$ with $z_{1i} \sim N(0, 1)$ and $z_{2i} \sim N(0, I_{K-4})$. The error term is generated by

$$(u_i, v_i) = ((1 + \phi z_{21i})\varepsilon_{1i}, \rho u_i + \sqrt{1 - \rho^2}\varepsilon_{2i}),$$

where ε_{1i} and ε_{2i} are independently drawn from $N(0, 1)$. We set $n = 200$ for the sample size in all cases, and set $\beta = \gamma = 1$, $\rho = 0.2$, and $\phi = 0.2$. Note that the error terms are heteroskedastic. The number of instruments is $K = 30$. For each Monte Carlo replication, we set the value of d to fix the value of the concentration parameter $\delta^2 = \frac{\pi' Z Z' \pi}{Var(v_i)}$. For each concentration parameter value $\delta^2 = 60$, 30, and 10, we calculate the powers of the tests for (i) the heteroskedasticity robust version of Anderson-Rubin type test with the naive variance estimator (AR), (ii) the AR test with the cross-fit variance estimator by Mikusheva and Sun (2022) (cross-AR), (iii) the jackknife Lagrange multiplier test by Matsushita and Otsu (2022) (JLM), and (iv) the JLM test with the proposed cross-fit variance estimators $\hat{\Psi}_{\text{cf},1}$ and $\hat{\Psi}_{\text{cf},2}$ (cross-JLM and cross-JLM*, respectively). The number of Monte Carlo replications in each experiment is 1000.

Figures 4.1–4.3 display the power curves at the nominal 5% significance level. From these figures, we find that: (i) cross-JLM and cross-JLM* carry the same feature of JLM's power improvement for small $|\Delta|$ compared with cross-AR, (ii) cross-JLM and cross-JLM* improve the power property of JLM not only for large values of $|\Delta|$ but also for small $|\Delta|$ as shown in Figures 4.2 and 4.3, (iii) when identification is strong or moderate, cross-JLM and cross-JLM* outperform other methods in terms of local power as shown in Figures 4.1 and 4.2, and (iv) the leave-one-out cross-fit feature of cross-JLM* does not exhibit significant improvement from cross-JLM.

These results for the power properties naturally lead to the question whether the test inversion of cross-JLM yields a shorter confidence interval. We compare the coverage probability and median length of the confidence intervals based on cross-AR, JLM, and cross-JLM in Table 4.2. "Infinite CI" reports the frequency that the length of the confidence interval exceeds the predetermined length, 10. From Table 4.2, we find that (i) cross-JLM tends to have the shorter confidence interval than the other two when the identification is strong or moderate, and (ii) cross-AR exhibits fewer infinite-length confidence intervals than the other two due to its excellent power performance for the distant alternatives.

FIGURE 4.1. Power curves: $\delta^2 = 60$.

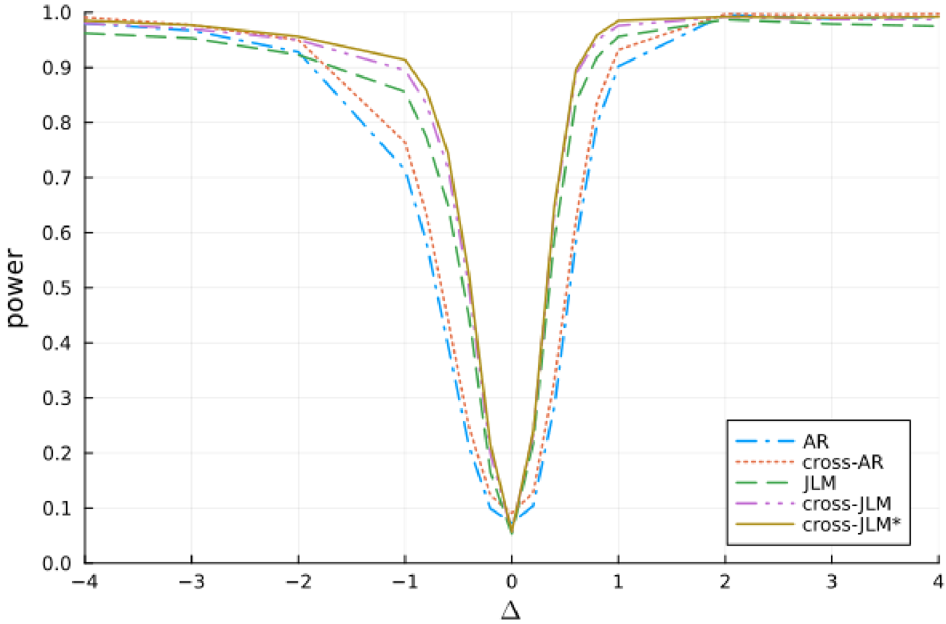

FIGURE 4.2. Power curves: $\delta^2 = 30$.

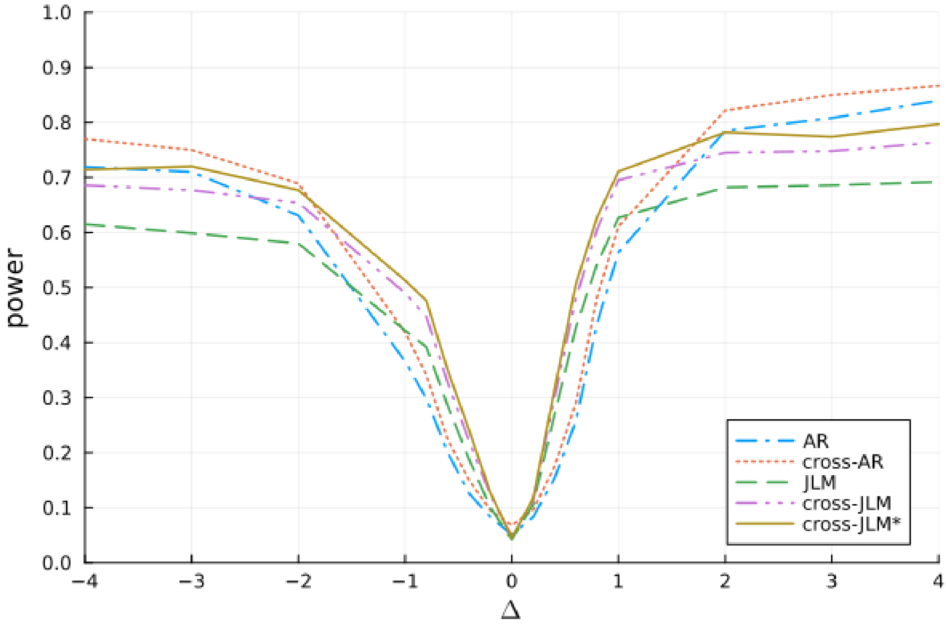

FIGURE 4.3. Power curves: $\delta^2 = 10$.

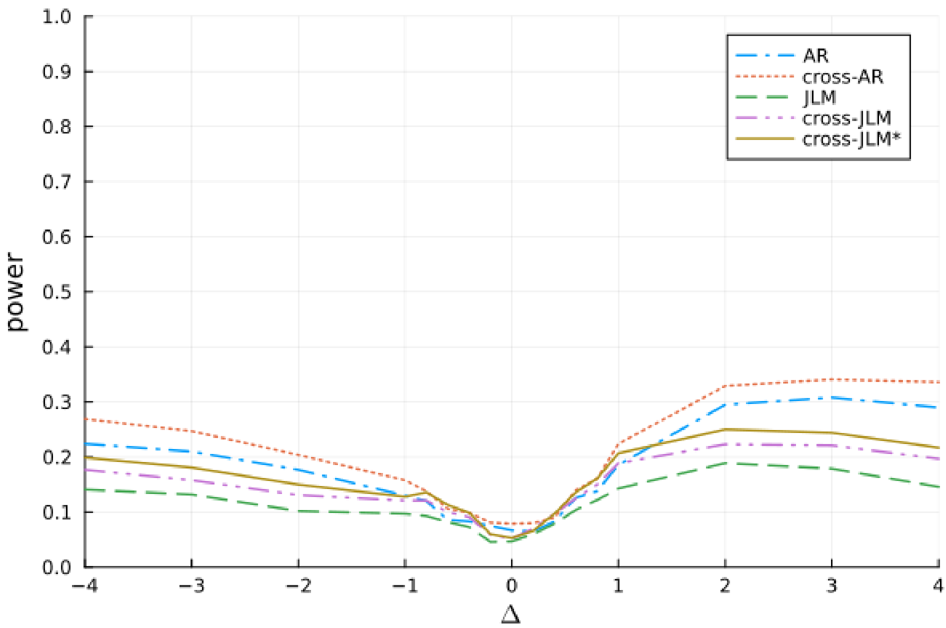

Table 2. Coverage probability and median length of confidence intervals

	Coverage	Median length	Infinite CI
		$\delta^2 = 60$	
cross-AR	97.8%	1.22	0.2%
JLM	95.4%	0.82	2.3%
cross-JLM	94.8%	0.75	0.5%
		$\delta^2 = 30$	
cross-AR	96.7%	2.34	18.8%
JLM	96.2%	2.65	35.5%
cross-JLM	95.8%	1.75	26.9%
		$\delta^2 = 10$	
cross-AR	98.4%	10.0	70.1%
JLM	98.5%	10.0	85.0%
cross-JLM	98.3%	10.0	80.0%

5. Illustrations based on Angrist and Krueger (1991)

Angrist and Krueger (1991) is a canonical empirical example of many instrumental variables, where the returns to education are estimated by regressions of the log weekly wage on the year of schooling with many instruments, such as the quarter of birth (QOB), year of birth (YOB), and state of birth (SOB) dummy variables from the 1980 US census of 329,509 men born in 1930–39. To assess performance of the proposed cross-fit JLM test in a more empirically relevant setting, we conduct another simulation study that preserves the structure of Angrist and Krueger's (1991) data considered by Angrist and Frandsen (2022) and Mikusheva and Sun (2022). For a detailed description of this simulation exercise, see Mikusheva and Sun (2022, pp.2684–2685).

First we investigate power properties of cross-AR, JLM, and cross-JLM. To vary identification strength, we vary the sample size of the simulated data to be 1.5%, 1%, and 0.5% of the original sample size, which correspond to strong, moderate, and weak identification, respectively, as noted by Mikusheva and Sun (2022). We also allow that the number of available instruments to vary according to the simulated sample size. The number of Monte Carlo replications in each experiment is 1000. Figures 5.1-5.3 display the power curves at the nominal 5% significance level. We find similar patterns as in the previous simple Monte Carlo exercise: (i) cross-JLM carries the same feature of JLM's power improvement for small $|\Delta|$ compared with cross-AR, (ii) cross-JLM improves the power property of JLM for large values of $|\Delta|$, but the improvements are small for small $|\Delta|$ as shown in 5.2 and 5.3.

Table 3 shows the results for the confidence intervals. We find that (i) in terms of the median lengths, JLM and cross-JLM have much shorter confidence intervals than cross-AR, but (ii) cross-AR exhibits fewer infinite-length confidence intervals than the other two due to its excellent power performance for the distant alternatives. Our results suggest that the proposed cross-JLM test can be a useful complement to the existing cross-AR by Mikusheva and Sun (2022).

Figure 5.1. Power curves under the setting of simulation design by AF22

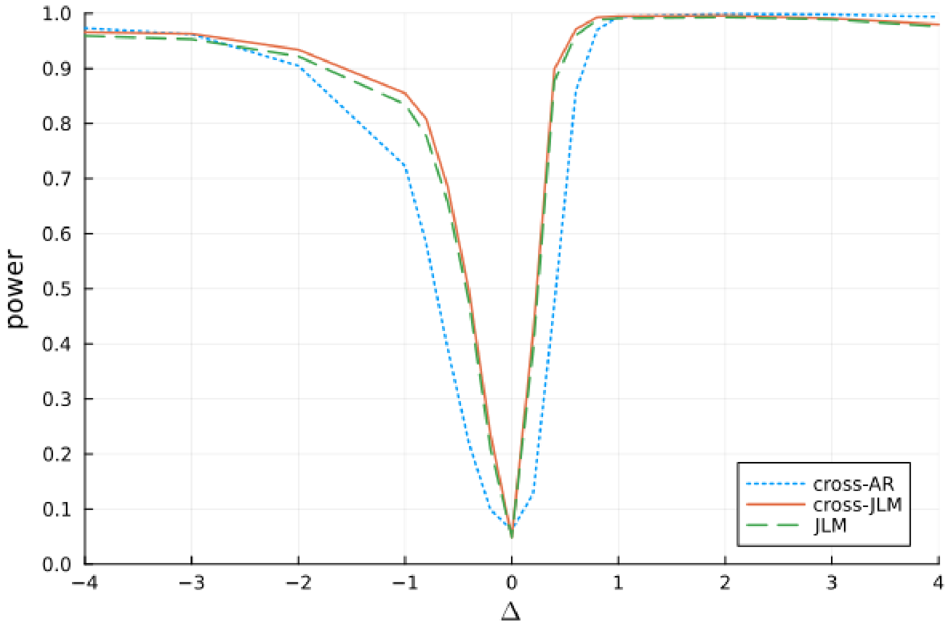

Figure 5.2. Power curves under the setting of simulation design by AF22

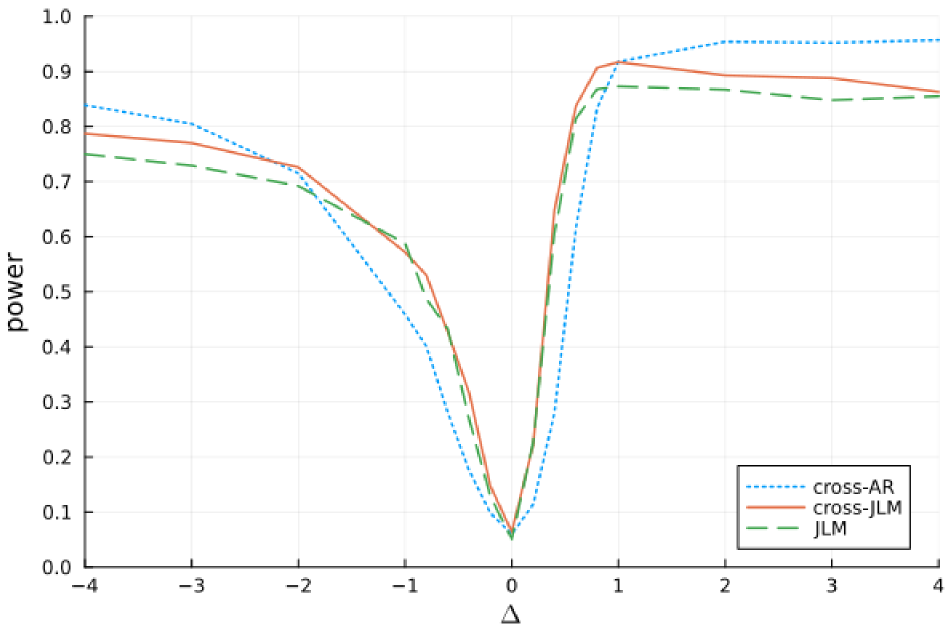

FIGURE 5.3. Power curves under the setting of simulation design by AF22

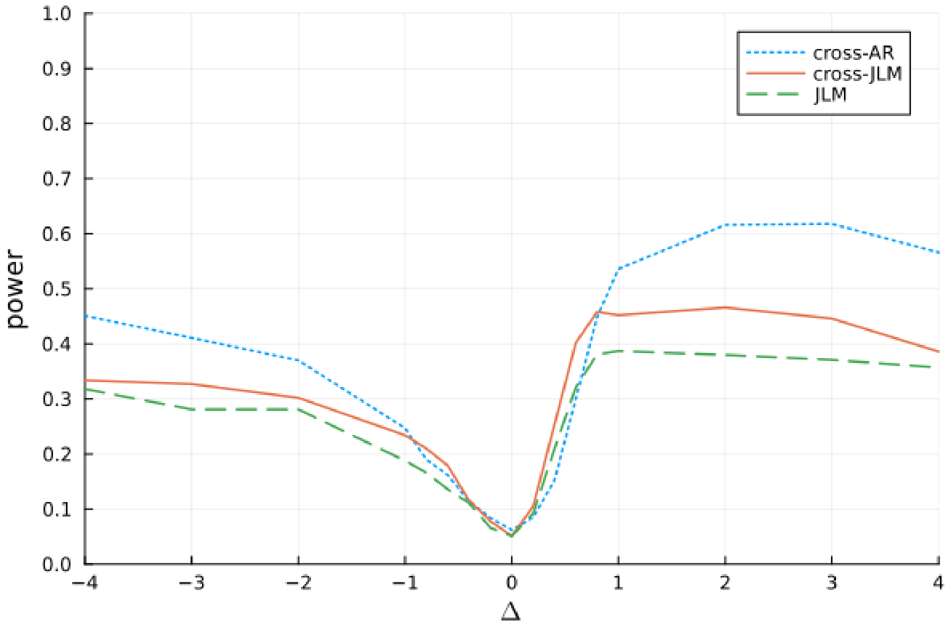

TABLE 3. Coverage probability and median length of confidence intervals

	Coverage	Median length	Infinite CI
		$n = 4923,\ k = 154$	
cross-AR	95.0%	1.14	0.1%
JLM	95.7%	0.61	1.1%
cross-JLM	94.9%	0.60	0.7%
		$n = 3209,\ k = 135$	
cross-AR	97.5%	1.61	11.7%
JLM	96.4%	1.08	20.0%
cross-JLM	96.0%	1.1	19.5%
		$n = 1599,\ k = 111$	
cross-AR	97.9%	9.95	50.0%
JLM	98.2%	10.0	65.5%
cross-JLM	97.1%	10.0	62.7%

We close this section with an empirical illustration for the original Angrist and Krueger's (1991) dataset. In Table 4, we apply cross-AR, JLM, and cross-JLM to construct the 95% confidence sets using 180 and 1530 instruments as in Mikusheva and Sun (2022). In this application, the confidence sets based on JLM and cross-JLM are narrower than those based on cross-AR. This difference may be attributed to the better power property of the JLM and cross-JLM tests for small values of $|\Delta|$ as illustrated in the simulation study above. In this illustration, the cross-JLM confidence interval is slightly longer than the JLM confidence interval.

Table 4. 95% confidence interval using Angrist and Krueger (1991) data

	cross-AR	JLM	cross-JLM
180 instruments	$[0.008, 0.201]$	$[0.067, 0.133]$	$[0.067, 0.133]$
1530 instruments	$[-0.047, 0.202]$	$[0.025, 0.123]$	$[0.013, 0.124]$

Appendix A. Proof of Theorem 1

Hereafter C means a generic positive constant. To simplify the presentation, we provide the proof for the case of $G = 1$ (i.e., x is scalar). Let $e_i = y_i - \mathbb{E}[y_i|x_i]$. Under H_0, $\hat{u}_i$ is written as

$$\hat{u}_i = y_i - x_i\hat{\gamma} = x_i(\gamma_0 - \hat{\gamma}) + e_i,$$

and we have

$$\begin{aligned}\{\hat{u}_i(m_i'\hat{u})\}\{\hat{u}_j(m_j'\hat{u})\} &= (\gamma_0 - \hat{\gamma})^2 x_i(m_i'e)x_j(m_j'e) + (\gamma_0 - \hat{\gamma})e_i(m_i'e_i)x_j(m_j'e) \\ &\quad +(\gamma_0 - \hat{\gamma})e_i(m_i'e)e_j(m_j'e) + e_i(m_i'e)e_j(m_j'e).\end{aligned}$$

Let $\tilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$ and $X = (x_1, \ldots, x_n)'$. Note that V can be written as

$$V = \frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2\mathbb{E}[\{e_i(m_i'e)\}\{e_j(m_j'e)\}|X].$$

Thus, the estimation error can be decomposed as

$$\begin{aligned}\hat{V}_{\text{cf}} - V &= \frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2\{e_i(m_i'e)e_j(m_j'e) - \mathbb{E}[e_i(m_i'e)e_j(m_j'e)|X]\} \\ &\quad +(\gamma_0 - \hat{\gamma})^2\frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2 x_i(m_i'e)x_i(m_j'e) \\ &\quad +(\gamma_0 - \hat{\gamma})\frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2\{x_i(m_i'e_i)e_j(m_j'e) + e_i(m_i'e)x_j(m_j'e)\}.\end{aligned}$$

By Mikusheva and Sun (2022, Lemma 2), the first term is $o_p(1)$. Since $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$ under H_0 (Assumption (iv)), it is sufficient for the conclusion to show that

$$T_1 := \frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2 x_i(m_i'e)x_i(m_j'e) = o_p(n), \tag{A.1}$$

$$T_2 := \frac{2}{K}\sum_{i=1}\sum_{j\neq i}\tilde{P}_{ij}^2 x_i(m_i'e_i)e_j(m_j'e) = o_p(n^{1/2}). \tag{A.2}$$

We first show (A.1). Let $\phi_{ij} = x_i(m_i'e)x_j(m_j'e)$. Then we have

$$\begin{aligned}\mathbb{E}[T_1^2] &= \frac{4}{K^2}\mathbb{E}\left[\sum_{i=1}^n\sum_{j\neq i}\tilde{P}_{ij}^4\mathbb{E}[\phi_{ij}^2 \mid X]\right] + \frac{4}{K^2}\mathbb{E}\left[\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{ki}^2\mathbb{E}[\phi_{ij}\phi_{ki}|X]\right] \\ &\quad +\frac{4}{K^2}\mathbb{E}\left[\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{kj}^2\mathbb{E}[\phi_{ij}\phi_{kj}|X]\right] + \frac{2}{K^2}\mathbb{E}\left[\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\tilde{P}_{ij}^2\tilde{P}_{kl}^2\mathbb{E}[\phi_{ij}\phi_{kl}|X]\right] \\ &=: A_1 + A_2 + A_3 + A_4.\end{aligned}$$

Note that ϕ_{ij} is decomposed as

$$\begin{aligned}
\phi_{ij} &= x_i x_j (M_{ii}M_{jj}e_i e_j + M_{ii}M_{ij}e_i^2 + M_{ij}M_{jj}e_j^2 + M_{ij}^2 e_i e_j) \\
&\quad + x_i x_j \left\{ \sum_{a\neq i,j} (M_{ii}M_{ja}e_i e_a + M_{ij}M_{ja}e_j e_a + M_{jj}M_{ia}e_j e_a + M_{ij}M_{ia}e_i e_a) \right\} \\
&\quad + x_i x_j \left\{ \sum_{a\neq i,j} \sum_{b\neq i,j} M_{ia}M_{jb}e_a e_b \right\} \\
&=: \phi_{ij}^a + \phi_{ij}^b + \phi_{ij}^c.
\end{aligned}$$

Assumption (iii) implies that $\mathbb{E}[(\phi_{ij}^a)^2|X] \leq Cx_i^2 x_j^2$. By applying the same argument as in Lemma 1 below, it holds that $\mathbb{E}[(\phi_{ij}^b)^2|X]$ and $\mathbb{E}[(\phi_{ij}^c)^2|X]$ are bounded by $Cx_i^2 x_j^2$. Thus, we have

$$\mathbb{E}[\phi_{ij}^2|X] \leq Cx_i^2 x_j^2.$$

Let $\lambda_{\min}(A)$ be the minimum eigenvalue of a matrix A. $\tilde{P}_{ij}^2$ can be bounded as

$$\tilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} \leq \frac{P_{ij}^2}{(1-P_{ii})(1-P_{jj})} \leq \frac{P_{ij}^2}{\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^2}, \tag{A.3}$$

where the first inequality follows from the properties of the matrices P and M, and the second inequality follows from

$$P_{ii} \leq \lambda_{\min}^{-1}(n^{-1}B'B)n^{-1} \max_i ||b(x_i)||^2 \leq \lambda_{\min}^{-1}(n^{-1}B'B)n^{-1}\zeta_K^2, \tag{A.4}$$

by using Assumptions (i)-(ii). Furthermore, Mikusheva and Sun (2022, Lemma S1.3 (b)) implies

$$\sum_{i=1}^n \sum_{j\neq i} P_{ij}^4 x_i^2 x_j^2 \leq \sum_{i=1}^n \sum_{j=1}^n P_{ij}^2 x_i^2 x_j^2 \leq \sum_{i=1}^n x_i^4. \tag{A.5}$$

Combining these results, A_1 is bounded as

$$\begin{aligned}
A_1 &\leq \frac{4C}{K^2}\mathbb{E}\left[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} \sum_{i=1}^n \sum_{j\neq i} P_{ij}^4 x_i^2 x_j^2 \right] \\
&\leq \frac{4C}{K^2} \sum_{i=1}^n \mathbb{E}[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^4] = O(nK^{-2}),
\end{aligned}$$

where the first inequality follows from (A.3), the second inequality follows from (A.5), and the equality follows from the Hölder inequality and Assumptions (i)-(iii).

For A_2, the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\phi_{ij}\phi_{ik}|X]| \leq Cx_i^2 x_j^2.$$

Also Mikusheva and Sun (2022, Lemmas S1.1 (iii) and S1.3 (e)) imply

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n P_{ij}^2 P_{ki}^2 x_i^2 x_j^2 \leq K \sum_{i=1}^n x_i^4.$$

By applying the same argument as A_1, we have

$$A_2 \leq \frac{4C}{K}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}x_i^4] = O(nK^{-1}).$$

Similarly, we have $A_3 = O(nK^{-1})$.

For A_4, note that

$$\mathbb{E}[\phi_{ij}\phi_{kl}|X] = \mathbb{E}[(\phi_{ij}^a+\phi_{ij}^b+\phi_{ij}^c)(\phi_{kl}^a+\phi_{kl}^b+\phi_{kl}^c)|X].$$

Since the indices in A_4 are distinct, $\mathbb{E}[\phi_{ij}^q\phi_{kl}^r|X]=0$ for $q\neq r$. Also we have $\mathbb{E}[\phi_{ij}^s\phi_{kl}^s|X] \leq C|x_ix_jx_kx_l|$ for $s\in\{a,b,c\}$. By Mikusheva and Sun (2022, Lemma S1.3 (b)),

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}P_{ij}^2P_{kl}^2|x_ix_jx_kx_l| \leq \left(\sum_{i=1}^{n}x_i^2\right)^2.$$

Combining these results yields

$$\begin{aligned}
A_4 &= \frac{2}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\tilde{P}_{ij}^2\tilde{P}_{kl}^2\{\mathbb{E}[\phi_{ij}^a\phi_{kl}^a|X]+\mathbb{E}[\phi_{ij}^b\phi_{kl}^b|X]+\mathbb{E}[\phi_{ij}^c\phi_{kl}^c|X]\}\right]\\
&\leq \frac{C}{K^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2n^{-1}\}^{-4}x_i^2x_j^2] = O(n^2K^{-2}).
\end{aligned}$$

Therefore, the Markov inequality yields $T_1 = O_p(\max\{n^{1/2}K^{-1}, n^{1/2}K^{-1/2}, nK^{-1}\})$ and Assumption (ii) guarantees (A.1).

We next show (A.2). Let $\psi_{ij} = x_i(m_i'e)e_j(m_j'e)$. Then we have

$$\begin{aligned}
\mathbb{E}[T_2^2] &= \frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\tilde{P}_{ij}^4\mathbb{E}[\psi_{ij}^2|X]\right]+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\tilde{P}_{ij}^4\mathbb{E}[\psi_{ij}\psi_{ji}|X]\right]\\
&\quad+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{il}^2\mathbb{E}[\psi_{ij}\psi_{il}|X]\right]+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{jl}^2\mathbb{E}[\psi_{ij}\psi_{jl}|X]\right]\\
&\quad+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{ki}^2\mathbb{E}[\psi_{ij}\psi_{ki}|X]\right]+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\tilde{P}_{ij}^2\tilde{P}_{kj}^2\mathbb{E}[\psi_{ij}\psi_{kj}|X]\right]\\
&\quad+\frac{1}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\tilde{P}_{ij}^2\tilde{P}_{kl}^2\mathbb{E}[\psi_{ij}\psi_{kl}|X]\right]\\
&=: C_1+C_2+C_3+C_4+C_5+C_6+C_7.
\end{aligned}$$

Note that ψ_{ij} is decomposed as

$$
\begin{aligned}
\psi_{ij} &= x_i e_j (M_{ii}M_{jj}e_i e_j + M_{ii}M_{ij}e_i^2 + M_{ij}M_{jj}e_j^2 + M_{ij}^2 e_i e_j) \\
&\quad + x_i e_j \left\{ \sum_{a \neq i,j} (M_{ii}M_{ja}e_i e_a + M_{ij}M_{ja}e_j e_a + M_{jj}M_{ia}e_j e_a + M_{ij}M_{ia}e_i e_a) \right\} \\
&\quad + x_i e_j \left\{ \sum_{a \neq i,j} \sum_{b \neq i,j} M_{ia}M_{jb}e_a e_b \right\} \\
&=: \psi_{ij}^a + \psi_{ij}^b + \psi_{ij}^c.
\end{aligned}
$$

Assumption (iii) implies that $\mathbb{E}[(\psi_{ij}^a)^2|X] \leq Cx_i^2$. By applying the same argument as in Lemma 1 below, it holds that $\mathbb{E}[(\psi_{ij}^b)^2|X]$ and $\mathbb{E}[(\psi_{ij}^c)^2|X]$ are bounded by Cx_i^2. Thus, we have

$$\mathbb{E}[\psi_{ij}^2|X] \leq Cx_i^2.$$

Furthermore, Mikusheva and Sun (2022, Lemma S1.3 (a)) implies

$$\sum_{i=1}^n \sum_{j=1}^n P_{ij}^4 x_i^2 \leq \sqrt{K} \sum_{i=1}^n x_i^2.$$

Combining these results with (A.4), C_1 is bounded as

$$
\begin{aligned}
C_1 &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} \sum_{i=1}^n \sum_{j \neq i} P_{ij}^4 x_i^2 \right] \\
&\leq \frac{2C}{K^2}\sqrt{K} \sum_{i=1}^n \mathbb{E}[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^2] = O(nK^{-3/2}).
\end{aligned}
$$

For C_2, notice that the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\psi_{ij}\psi_{ji}|X]| \leq C|x_i x_j|.$$

Also Mikusheva and Sun (2022, Lemma S1.3 (b)) implies

$$\sum_{i=1}^n \sum_{j=1}^n P_{ij}^4 |x_i x_j| \leq \sum_{i=1}^n x_i^2.$$

By applying the same argument as C_1, we have

$$
\begin{aligned}
C_2 &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} \sum_{i=1}^n \sum_{j=1}^n P_{ij}^4 |x_i x_j| \right] \\
&\leq \frac{2C}{K^2} \sum_{i=1}^n \mathbb{E}[\{1 - \lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^2] = O(nK^{-2}).
\end{aligned}
$$

For C_3, notice that the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\psi_{ij}\psi_{il}|X]| \leq Cx_i^2,$$

and we also have

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} P_{ij}^2 P_{il}^2 x_i^2 \leq K \sum_{i=1}^{n} x_i^2.$$

Combining these results, C_3 is bounded as

$$\begin{aligned} C_3 &\leq \frac{2C}{K}\mathbb{E}\left[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,j} P_{ij}^2 P_{il}^2 x_i^2\right] \\ &\leq \frac{2C}{K}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^2] = O(nK^{-1}). \end{aligned}$$

For C_4, the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\psi_{ij}\psi_{jl}|X]| \leq C|x_i x_j|,$$

and Mikusheva and Sun (2022, Lemma S1.3 (b)) implies

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} P_{ij}^2 P_{jl}^2 |x_i||x_j| \leq \sum_{i=1}^{n} x_i^2.$$

Combining these results, C_4 is bounded as

$$\begin{aligned} C_4 &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,j} P_{ij}^2 P_{jl}^2 |x_i x_j|\right] \\ &\leq \frac{2C}{K^2}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^2] = O(nK^{-2}). \end{aligned}$$

For C_5, the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\psi_{ij}\psi_{ki}|X]| \leq C|x_i x_k|,$$

and Mikusheva and Sun (2022, Lemma S1.3 (e)) implies

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} P_{ij}^2 P_{ki}^2 |x_i||x_k| \leq \sqrt{K}\sum_{i=1}^{n} x_i^2.$$

Combining these results, C_5 is bounded as

$$\begin{aligned} C_5 &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j} P_{ij}^2 P_{ki}^2 |x_i x_k|\right] \\ &\leq \frac{2C}{K^2}\sqrt{K}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4} x_i^2] = O(nK^{-3/2}). \end{aligned}$$

For C_6, the Cauchy-Schwarz inequality implies

$$|\mathbb{E}[\psi_{ij}\psi_{kj}|X]| \leq (\mathbb{E}[\psi_{ij}^2|X])^{1/2}(\mathbb{E}[\psi_{kj}^2|X])^{1/2} \leq C|x_i x_k|,$$

and Mikusheva and Sun (2022, Lemma S1.3 (e)) implies

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} P_{ij}^2 P_{kj}^2 |x_i||x_k| \leq K \sum_{i=1}^{n} x_i^2.$$

Combining these results, C_6 is bounded as

$$\begin{aligned} C_6 &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} P_{ij}^2 P_{kj}^2 |x_i||x_k|\right] \\ &\leq \frac{2C}{K}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}x_i^2] = O(nK^{-1}). \end{aligned}$$

For C_7, note that

$$\mathbb{E}[\psi_{ij}\psi_{kl}|X] = \mathbb{E}[(\psi_{ij}^a + \psi_{ij}^b + \psi_{ij}^c)(\psi_{kl}^a + \psi_{kl}^b + \psi_{kl}^c)|X].$$

Since the indices in C_7 are distinct, $\mathbb{E}[\psi_{ij}^q \psi_{kl}^r|X] = 0$ for $q \neq r$. Also we have $\mathbb{E}[\psi_{ij}^s \psi_{kl}^s|X] = C|x_i x_k|$ for $s \in \{a, b, c\}$. Applying Mikusheva and Sun (2022, Lemma S1.3 (a)) twice yields

$$\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} P_{ij}^2 P_{kl}^2 |x_i x_k| \leq K \sum_{i=1}^{n} x_i^2.$$

Combining these results, we have

$$\begin{aligned} C_7 &= \frac{2}{K^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} \tilde{P}_{ij}^2 \tilde{P}_{kl}^2 \{\mathbb{E}[\psi_{ij}^a\psi_{kl}^a|X] + \mathbb{E}[\psi_{ij}^b\psi_{kl}^b|X] + \mathbb{E}[\psi_{ij}^c\psi_{kl}^c|X]\}\right] \\ &\leq \frac{2C}{K^2}\mathbb{E}\left[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} P_{ij}^2 P_{kl}^2 |x_i x_k|\right] \\ &\leq \frac{2C}{K}\sum_{i=1}^{n}\mathbb{E}[\{1-\lambda_{\min}^{-1}(n^{-1}B'B)\zeta_K^2 n^{-1}\}^{-4}x_i^2] = O(nK^{-1}). \end{aligned}$$

Therefore, the Markov inequality yields $T_2 = O_p(\max\{n^{1/2}K^{-3/2}, n^{1/2}K^{-1}, n^{1/2}K^{-1/2}\})$ and Assumption (ii) guarantees (A.2).

Since (A.1) and (A.2) are verified, we obtain the conclusion.

Appendix B. Proof of Theorem 2

The following lemma, an adaptation of Mikusheva and Sun (2022, Lemma 2), is used in our proof.

Lemma 1. *Suppose that* $\xi_i = (\xi_i^{(1)}, \xi_i^{(2)}, \xi_i^{(3)})'$ *for* $i = 1, \ldots, n$ *are independent mean zero random vectors with* $\max_i \mathbb{E}|\xi_i|^6 \leq C$. *Let* $\eta_{ij} = \{\xi_i^{(1)}(m_i'\xi^{(2)})\}\{\xi_j^{(1)}(m_j'\xi^{(3)})\}$, *where* m_i *is the* i*-th column of* M. *Then*

$$\max_{i,j}\mathbb{E}[\eta_{ij}^2] \leq C_1, \qquad \max_{i,j,k}|\mathbb{E}[\eta_{ij}\eta_{ik}]| \leq C_2,$$

for some finite constants C_1 *and* C_2.

Proof of Lemma 1: Notice that η_{ij} can be decomposed into $\eta_{ij} = A_{1,ij} + A_{2,ij} + A_{3,ij}$, where

$$\begin{aligned}
A_{1,ij} &= M_{ii}M_{jj}\xi_i^{(1)}\xi_i^{(2)}\xi_j^{(1)}\xi_j^{(3)} + M_{ii}M_{ij}\xi_i^{(1)}\xi_i^{(2)}\xi_i^{(3)}\xi_j^{(1)} + M_{ij}M_{jj}\xi_i^{(1)}\xi_j^{(1)}\xi_j^{(2)}\xi_j^{(3)} + M_{ij}^2\xi_i^{(1)}\xi_i^{(3)}\xi_j^{(1)}\xi_j^{(2)}, \\
A_{2,ij} &= \xi_i^{(1)}\xi_j^{(1)}\sum_{a\neq i,j}\{M_{ii}M_{ja}\xi_i^{(2)}\xi_a^{(3)} + M_{ij}M_{ja}\xi_j^{(2)}\xi_a^{(3)} + M_{jj}M_{ia}\xi_j^{(3)}\xi_a^{(2)} + M_{ij}M_{ia}\xi_i^{(3)}\xi_a^{(2)}\}, \\
A_{3,ij} &= \xi_i^{(1)}\xi_j^{(1)}\sum_{a\neq i,j}\sum_{b\neq i,j} M_{ia}M_{jb}\xi_a^{(2)}\xi_b^{(3)}.
\end{aligned}$$

For the first statement, it suffices to show that $\max_{i,j}\mathbb{E}[A_{s,ij}^2]$ is bounded for each $s = 1, 2, 3$. The moment condition $\max_i \mathbb{E}|\xi_i|^6 \leq C$ implies that $\max_{i,j}\mathbb{E}[A_{1,ij}^2]$ is bounded. From the proof of Mikusheva and Sun (2022, Lemma 2), we have boundedness of $\max_{i,j}\mathbb{E}[A_{2,ij}^2]$ and $\max_{i,j}\mathbb{E}[A_{3,ij}^2]$. Therefore, the first statement follows. The second statement follows from the Cauchy-Schwarz inequality.

Proof of Theorem 2. For simplicity, we present the proof for the case of $G = 1$. Let $\tilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$, $u = (u_1, \ldots, u_n)'$, and $v = (v_1, \ldots, v_n)'$. Since $\Phi = \frac{1}{K}\sum_{i=1}^n\sum_{j\neq i}\tilde{P}_{ij}^2\mathbb{E}[u_i(m_i u)' u_j(m_j' u)]$ by Mikusheva and Sun (2022), $\hat{\Phi}_{\text{cf}} - \Phi$ can be decomposed as

$$\begin{aligned}
\hat{\Phi}_{\text{cf}} - \Phi &= \frac{1}{K}\sum_{i=1}^n\sum_{j\neq i}\tilde{P}_{ij}^2\Big[\{u_i(m_i'u)u_j(m_j'u) - \mathbb{E}[u_i(m_i'u)u_j(m_j'u)]\} \\
&\quad +(\hat{\beta}-\beta)^4 x_i(m_i'v)x_j(m_j'v) \\
&\quad +(\hat{\beta}-\beta)^3\{x_i(m_i'v)x_j(m_j'u) + x_i(m_i'v)u_j(m_j'v) + x_i(m_i'u)x_j(m_j'v) + u_i(m_i'v)x_j(m_j'v)\} \\
&\quad +(\hat{\beta}-\beta)^2\left\{\begin{array}{c} x_i(m_i'v)u_j(m_j'u) + x_i(m_i'u)x_j(m_j'u) + x_i(m_i'u)u_j(m_j'v) + u_i(m_i'v)x_j(m_j'u) \\ +u_i(m_i'v)u_j(m_j'v) + u_i(m_i'v)u_j(m_j'u) + u_i(m_i'u)x_j(m_j'v) \end{array}\right\} \\
&\quad +(\hat{\beta}-\beta)\{x_i(m_i'u)u_j(m_j'u) + u_i(m_i'u)x_j(m_j'u) + u_i(m_i'u)u_j(m_j'v)\}\Big]. \qquad \text{(B.1)}
\end{aligned}$$

By Mikusheva and Sun (2022, Lemma 2), the first term of (B.1) satisfies

$$\frac{1}{K}\sum_{i=1}^n\sum_{j\neq i}\tilde{P}_{ij}^2\{u_i(m_i'u)u_j(m_j'u) - \mathbb{E}[u_i(m_i'u)u_j(m_j'u)]\} = o_p(1).$$

It remains to show that the other terms of (B.1) are $o_p(1)$. Hereafter, we present the proof for the second term. The other terms are handled in the same manner.

Since $\hat{\beta} \xrightarrow{p} \beta$, it is sufficient to show

$$\mathbb{E}\left[\left(\frac{1}{K}\sum_{i=1}^n\sum_{j\neq i}\tilde{P}_{ij}^2\phi_{ij}\right)^2\right] = O(1), \qquad \text{(B.2)}$$

where $\phi_{ij} = x_i x_j (m_i' v)(m_j' v)$. From $\phi_{ij} = \phi_{ji}$, this can be decomposed as

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{K}\sum_{i=1}^n \sum_{j\neq i} \tilde{P}_{ij}^2 \phi_{ij}\right)^2\right] &= \frac{2}{K^2}\sum_{i=1}^n \sum_{j\neq i} \tilde{P}_{ij}^4 \mathbb{E}[\phi_{ij}^2] + \frac{2}{K^2}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\neq i,j} \tilde{P}_{ij}^2 \tilde{P}_{ki}^2 \mathbb{E}[\phi_{ij}\phi_{ki}] \\
&\quad + \frac{2}{K^2}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\neq i,j} \tilde{P}_{ij}^2 \tilde{P}_{kj}^2 \mathbb{E}[\phi_{ij}\phi_{kj}] + \frac{1}{K^2}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} \tilde{P}_{ij}^2 \tilde{P}_{kl}^2 \mathbb{E}[\phi_{ij}\phi_{kl}] \\
&=: A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

Hereafter, C means a generic constant. For A_1, the leading term of $\mathbb{E}[\phi_{ij}^2]$ is $\mathbb{E}[\{v_i(m_i'v)v_j(m_j'v)\}^2]$, and Lemma 1 with $\eta_{ij} = v_i(m_i'v)v_j(m_j'v)$ implies $\max_{i,j} \mathbb{E}[\{v_i(m_i'v)v_j(m_j'v)\}^2] \leq C$. Thus, we obtain $\max_{i,j} \mathbb{E}[\phi_{ij}^2] \leq C$. From Assumption (ii), we have

$$\tilde{P}_{ij}^2 \leq \frac{P_{ij}^2}{(1-P_{ii})(1-P_{jj})} \leq \frac{P_{ij}^2}{(1-\delta)^2}, \tag{B.3}$$

Therefore, the order of A_1 is

$$A_1 \leq \frac{2C}{K^2(1-\delta)^2}\sum_{i=1}^n \sum_{j\neq i} P_{ij}^4 \leq \frac{2C}{K(1-\delta)^2} = O(K^{-1}),$$

where the first inequality follows from (B.3), and the second inequality follows from Chao *et al.* (2012, Lemma B1 (i)).

Similarly, for A_2, the leading term of $\mathbb{E}[\phi_{ij}\phi_{ik}]$ is $\mathbb{E}[\{v_i(m_i'v)v_j(m_j'v)\}\{v_i(m_i'v)v_k(m_k'v)\}]$ is bounded by Lemma 1 with $\eta_{ij} = v_i(m_i'v)v_j(m_j'v)$ and $\eta_{ik} = v_i(m_i'v)v_k(m_k'v)$, which implies $\max_{i,j,k} |\mathbb{E}[\phi_{ij}\phi_{ik}]| \leq C$. Therefore, the order of A_2 is

$$A_2 \leq \frac{2C}{K^2(1-\delta)^4}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\neq i,j} P_{ij}^2 P_{ki}^2 \leq \frac{2C}{K(1-\delta)^4} = O(K^{-1}),$$

where the first inequality follows from (B.3), and the second inequality follows from Chao *et al.* (2012, Lemma B1 (i)). We can show that $A_3 = O(K^{-1})$ in the same manner.

Finally, for A_4, the leading term of $\mathbb{E}[\phi_{ij}\phi_{kl}]$ is $\mathbb{E}[\{v_i(m_i'v)v_j(m_j'v)\}\{v_k(m_k'v)v_l(m_l'v)\}]$. Let us decompose $\{v_i(m_i'v)v_j(m_j'v)\} = B_{1,ij} + B_{2,ij} + B_{3,ij}$ as in Lemma 1, where

$$\begin{aligned}
B_{1,ij} &= M_{ii}M_{jj}v_i^2v_j^2 + M_{ii}M_{ij}v_i^3v_j + M_{ij}M_{jj}v_jv_j^3 + M_{ij}^2v_i^2v_j^2, \\
B_{2,ij} &= v_iv_j\sum_{a\neq i,j}(M_{ii}M_{ja}v_iv_a + M_{ij}M_{ja}v_jv_a + M_{jj}M_{ia}v_jv_a + M_{ij}M_{ia}v_iv_a), \\
B_{3,ij} &= v_iv_j\sum_{a\neq i,j}\sum_{b\neq i,j} M_{ia}M_{jb}v_av_b.
\end{aligned}$$

Notice that $B_{2,ij}$ and $B_{3.ij}$ are exactly the same as the terms appeared in the proof of Mikusheva and Sun (2022, Lemma 2) (correspond to "$A_{2,ij}$" and "$A_{3,ij}$", respectively, in their notation). Since the indices in A_4 are distinct, $\mathbb{E}[B_{q,ij}B_{r,kl}] = 0$ for $q \neq r$. Also $|\mathbb{E}[B_{1,ij}B_{1,kl}]| \leq C$ by the moment condition. Combining these results, we have

$$A_4 = \frac{1}{K^2}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} \tilde{P}_{ij}^2 \tilde{P}_{kl}^2 (\mathbb{E}[A_{1,ij}A_{1,kl}] + \mathbb{E}[A_{2,ij}A_{2,kl}] + \mathbb{E}[A_{3,ij}A_{3,kl}]) = O(K^{-1}),$$

where the orders of the last two terms are shown by Mikusheva and Sun (2022, Lemma 2).

Appendix C. Proof of unbiasedness of $\hat{\Psi}_{\text{cf},1}$ and $\hat{\Psi}_{\text{cf},2}$ in (3.2)

Unbiasedness of $\hat{\Psi}_{\text{cf},1}$ follows from that of $\hat{\Psi}_{\text{cf},2}$. For the first term of $\hat{\Psi}_{\text{cf},2}$, observe that

$$\begin{aligned}
&\sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}\left[x_i P_{ik} \frac{\tilde{u}_{0k} u_{0k}}{M_{kk}} P_{kj} x_j' \middle| Z \right] \\
=& \sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}\left[x_i P_{ik} \frac{M_{kk} u_{0k}^2}{M_{kk}} P_{kj} x_j' \middle| Z \right] + \underbrace{\sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}\left[x_i P_{ik} \frac{\sum_{l\neq k} M_{kl} u_{0l} u_{0k}}{M_{kk}} P_{kj} x_j' \middle| Z \right]}_{=0 \text{ because } \mathbb{E}[u_{0l}u_{0k}|Z]=0} \\
=& \sum_{i,j,k,i\neq k,j\neq k}^{n} \mathbb{E}[x_i P_{ik} u_{0k}^2 P_{kj} x_j' | Z],
\end{aligned}$$

where the first equality follows from the definition $\tilde{u}_{0k} = \sum_{l=1}^{n} M_{kl} u_{0l}$. Thus, the (conditional) expectation of the first term of $\hat{\Psi}_{\text{cf},2}$ coincides with the one of Ψ.

For the second term of $\hat{\Psi}_{\text{cf},2}$, decompose

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} x_i \tilde{x}_j' \frac{\bar{u}_{0i} u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
=& \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} (\Pi' z_i + v_i) \left(\sum_{k=1}^{n} M_{jk} (z_k' \Pi + v_k') \right) \frac{(\sum_{m\neq j} M_{im} u_{0m}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
=& \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} (\Pi' z_i + v_i) \left(\sum_{k=1}^{n} M_{jk} v_k' \right) \frac{(\sum_{m\neq j} M_{im} u_{0m}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
=& \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} v_i \left(\sum_{k=1}^{n} M_{jk} v_k' \right) \frac{(\sum_{m\neq j} M_{im} u_{0m}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
=& \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} v_i (M_{jj} v_j') \frac{(M_{ii} u_{0i}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
&+ \underbrace{\mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} v_i (M_{jj} v_j') \frac{(\sum_{m\neq j,i} M_{im} u_{0m}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right]}_{=0 \text{ because } u_{0m} \text{can be isolated}} \\
&+ \underbrace{\mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} v_i \left(\sum_{k\neq j} M_{jk} v_k' \right) \frac{(M_{ii} u_{0i}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right]}_{=0 \text{ because } u_{0j} \text{can be isolated}} \\
&+ \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j\neq i} v_i \left(\sum_{k\neq j} M_{jk} v_k' \right) \frac{(\sum_{m\neq j,i} M_{im} u_{0m}) u_{0j}}{M_{ii} M_{jj}} P_{ij}^2 \middle| Z \right] \\
=:& \sum_{i=1}^{n} \sum_{j\neq i} \mathbb{E}[v_i v_j' u_{0i} u_{0j} P_{ij}^2 | Z] + C.
\end{aligned}$$

where $C = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i} v_i \left(\sum_{k\neq j} M_{jk}v_k'\right) \frac{(\sum_{m\neq j,i} M_{im}u_{0m})u_{0j}}{M_{ii}M_{jj}} P_{ij}^2 \middle| Z\right]$, the first equality follows from the definitions of $\bar{u}_{0i}$, x_i, and $\tilde{x}_j$, and the second and third equalities follow from $\mathbb{E}[u_{0m}u_{0j} \mid Z] = 0$ for $m \neq j$. Since the first term of the last line is identical to the second term of Ψ, it remains to show $C = 0$. Finally, we have

$$\begin{aligned} C &= \underbrace{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i} v_i \left(M_{ji}v_i\right) \frac{\left(\sum_{m\neq j,i} M_{im}u_{0m}\right)u_{0j}}{M_{ii}M_{jj}} P_{ij}^2 \middle| Z\right]}_{=0 \text{ because } u_{0j} \text{ can be isolated}} \\ &\quad + \underbrace{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i} v_i \left(\sum_{k\neq i,j} M_{jk}v_k'\right) \frac{\left(\sum_{m\neq j,i} M_{im}u_{0m}\right)u_{0j}}{M_{ii}M_{jj}} P_{ij}^2 \middle| Z\right]}_{=0 \text{ because } v_i \text{ and } u_{0j} \text{ can be isolated}} \\ &= 0, \end{aligned}$$

and we obtain (3.2).

References

[1] Angrist, J. D. and A. B. Krueger (1991) Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979-1014.

[2] Angrist, J. D. and B. Frandsen (2022) Machine labor, *Journal of Labor Economics*, 40, S97-S140.

[3] Chao, J., Hausman, J. A., Newey, W. K., Swanson, N. R. and T. Woutersen (2014) Testing overidentifying restrictions with many instruments and heteroskedasticity, *Journal of Econometrics*, 178, 15-21.

[4] Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. and N. R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments, *Quantitative Economics*, 3, 211-255.

[5] Hong, Y. and H. White (1995) Consistent specification testing via nonparametric series regression, *Econometrica*, 63, 1133-1159.

[6] Kline, P., Saggio, R. and M. Sølvsten (2020) Leave-out estimation of variance components, *Econometrica*, 88, 1859-1898.

[7] Matsushita, Y. and T. Otsu (2022) A jackknife Lagrange multiplier test with many weak instruments, forthcoming in *Econometric Theory*.

[8] Mikusheva, A. and L. Sun (2022) Inference with many weak instruments, *Review of Economic Studies*, 89, 2663-2686.

[9] Newey, W. K. and J. R. Robins (2018) Cross-fitting and fast remainder rates for semiparametric estimation, *arXiv: 1801.09138*.

[10] Sun, Y. and Q. Li (2006) An alternative series based consistent model specification test, *Economics Letters*, 93, 37-44.

[11] Tripathi, G. and Y. Kitamura (2003) Testing conditional moment restrictions, *Annals of Statistics*, 31, 2059-2095.

Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan.

Email address: matsushita.y@r.hit-u.ac.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK, and Keio Economic Observatory (KEO), 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: t.otsu@lse.ac.uk

Department of Economics, University of Rochester, Rochester, NY 14627, USA.

Email address: ksunada@ur.rochester.edu