TWO-SAMPLE INSTRUMENTAL VARIABLE REGRESSION WITH MANY INSTRUMENTS

YUKITOSHI MATSUSHITA, YUTA OKAMOTO, AND TAISUKE OTSU

Abstract. This paper develops estimation and inference methods for two-sample instrumental variables (TSIV) regression when the number of instruments k grows proportionally with sample size n. Under this asymptotic setting, we show that the conventional two-sample two-stage least squares (TS2SLS) estimator is generally inconsistent because it suffers from an asymptotic bias of order k/n. To address this issue, we propose a bias-corrected TS2SLS estimator and establish its consistency and asymptotic normality. We further derive heteroskedasticity-robust variance estimators and provide valid Wald and Lagrange multiplier tests for structural parameters. An empirical application to Angrist and Krueger's (1992) study shows that the conventional estimator substantially understates the effect of school-entry age on educational attainment—by more than 20% when many instruments are used. A simulation study confirms that our bias-corrected estimator exhibits negligible bias and accurate coverage, while the conventional TS2SLS estimator suffers from severe attenuation bias and size distortions. These results underscore the importance of addressing many-instrument bias in empirical TSIV applications.

1. Introduction

Instrumental variables (IV) methods are central to empirical economics. In many applications, the outcome, endogenous regressors, and instruments are not jointly observed in a single dataset. Early contributions—including Klevmarken (1982), Angrist and Krueger (1992), and Arellano and Meghir (1992)—extended IV methodology to settings with separate instrument–outcome and instrument–regressor samples, now known as the two-sample instrumental variables (TSIV) framework. Building on this foundation, Angrist and Krueger (1995) introduced the two-sample two-stage least squares (TS2SLS) estimator as a split-sample analogue of conventional 2SLS. Subsequent work has further developed the properties of TS2SLS: Inoue and Solon (2010) established asymptotic efficiency improvements over the original TSIV estimator; Pacini and Windmeijer (2016) proposed heteroskedasticity-robust inference procedures; Choi et al. (2018) developed weak-instrument robust methods; and Zhao et al. (2019) extended TSIV to heterogeneous samples.

Data combination in the spirit of TSIV is widely employed. For instance, Angrist and Krueger (1992) combined the 1960 and 1980 U.S. Censuses to study how age at school entry affects educational attainment, while Arellano and Meghir (1992) estimated a female labor supply model by linking the UK Labour Force Survey and the Family Expenditure Survey. See Ridder and Moffitt (2007) for a broader survey. More recently, TSIV-type approaches have been prominent in work on intergenerational mobility, where earnings are typically observed in separate samples across generations (see Jerrim et al., 2014 for a review). Related ideas are increasingly used in epidemiology under the umbrella of Mendelian randomization, which deploys genetic variants as instruments and often relies on two-sample designs (see Burgess et al., 2017).

A salient practical issue in TSIV applications is the potentially large number of instruments. The many-instrument asymptotic framework—where the number of instruments grows proportionally with the sample size—was developed by Kunitomo (1980), Morimune (1983), and Bekker (1994) to improve approximations for IV procedures; see also Chao and Swanson (2005) for weaker-instrument extensions and Han and Phillips (2006); Newey and Windmeijer (2009) for nonlinear GMM formulations. Mikusheva and Sun (2024) provides a recent survey. In two-sample settings, many instruments arise naturally and often unavoidably. Angrist and Krueger (1992), for example, used over 900 instruments from interactions among quarter-of-birth, year-of-birth, and state-of-birth indicators. Intergenerational mobility studies that use first names or surnames as instruments can also yield a growing number of instruments (e.g., Olivetti and Paserman, 2015). While the many-instrument problem has been studied in Mendelian randomization with GWAS summary statistics (e.g., Zhao et al., 2020), to the best of our knowledge there is no many-instrument asymptotic analysis tailored to TSIV with two individual-level samples in the economic context. This paper fills this gap.

The contributions of this paper are fourfold. First, the many-instrument asymptotic framework (with the number of instruments proportional to the sample size) is extended to TSIV. It is shown that the conventional TS2SLS estimator is generally inconsistent under this regime. This mirrors the well-known inconsistency of one-sample 2SLS with many instruments. Second, a bias-corrected TS2SLS estimator is proposed. The correction targets the term responsible for the first-order many-instrument bias and is constructed via a leave-one-out quadratic form that is unbiased for the offending component. Third, large-sample properties are established: the bias-corrected estimator is consistent and asymptotically normal under heteroskedasticity. A feasible variance estimator is provided, yielding valid Wald and score-type (LM) tests. Fourth, the method is illustrated using the education setting in Angrist and Krueger (1992). Conventional TS2SLS estimates are found to be more than 20% smaller in magnitude than their bias-corrected counterparts, underscoring the empirical relevance of the many-instrument bias.

This paper is organized as follows. Section 2 introduces the setup and main theoretical results. Section 3 presents the empirical illustration. Section 4 reports a simulation study. Proofs are collected in the Appendix.

2. Main results

2.1. Setup and asymptotic framework. Consider the following IV regression model

$$\begin{aligned} Y_1 &= X_1\beta + \epsilon_1, \\ X_1 &= Z_1\Pi + \eta_1, \end{aligned} \tag{2.1}$$

where Y_1 is an $n_1 \times 1$ vector of dependent variables, X_1 is an $n_1 \times p$ matrix of *unobserved* endogenous regressors, Z_1 is an $n_1 \times k$ matrix of instrumental variables, ϵ_1 is an $n_1 \times 1$ vector of disturbances, η_1 is a $n_1 \times p$ matrix of disturbances, β is a $p \times 1$ vector of structural parameters of interest, and Π is a $k \times p$ matrix of nuisance parameters. The disturbance term ϵ_1 may be contemporaneously correlated with regressors X_1. This paper is concerned with the situation where we only observe (Y_1, Z_1) but not X_1, and instead observe another sample (X_2, Z_2) on the endogenous regressors and instruments that obeys

$$X_2 = Z_2\Pi + \eta_2, \tag{2.2}$$

where X_2 is an $n_2 \times p$ matrix of (observable) endogenous regressors, Z_2 is an $n_2 \times k$ matrix of instrumental variables, and η_2 is an $n_2 \times p$ matrix of disturbances. Note that the coefficient matrix, Π, of Z_1 and Z_2 are identical. The model defined by (2.1)-(2.2) is called the two-sample instrumental variable (TSIV) regression model. The focus of this paper is estimation and inference on the structural parameters β when the researcher has moderately large number of instruments (but β is low-dimensional). In particular, our theoretical analysis is conducted under the asymptotic framework:

$$\frac{k}{n_1} \to \alpha \in [0, 1), \qquad \frac{n_2}{n_1} \to \rho \in (0, \infty), \tag{2.3}$$

for $k = k_n$, as $n_1, n_2 \to \infty$, while p is held fixed. The first condition ($k/n_1 \to \alpha$) says that the number of instruments k is increasing proportionately with the sample size n, and α may be viewed as the ratio of k to n_1 in a given sample. This condition is commonly employed in the literature of many instruments (e.g., Bekker, 1994). The second condition ($n_2/n_1 \to \rho$) is also standard in the literature of the TSIV methods (e.g., Inoue and Solon, 2010), and ρ may be viewed as the ratio of the two samples.

The most widely used estimator for β is the two-sample two-stage least squares (TS2SLS) estimator. Let $S_{ZX}^{(s)} = \frac{1}{n_s} Z_s' X_s$, $S_{ZZ}^{(s)} = \frac{1}{n_s} Z_s' Z_s$, and $S_{ZY}^{(s)} = \frac{1}{n_s} Z_s' Y_s$ for $s = 1, 2$. The TS2SLS estimator is expressed as

$$\begin{aligned}\hat{\beta}_{\mathrm{TS2SLS}} &= \left\{X_2' Z_2 (Z_2' Z_2)^{-1} Z_1' Z_1 (Z_2' Z_2)^{-1} Z_2' X_2\right\}^{-1} X_2' Z_2 (Z_2' Z_2)^{-1} Z_1' Y_1 \\ &= \{S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{ZX}^{(2)}\}^{-1} S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZY}^{(1)}. \end{aligned} \tag{2.4}$$

It is known that when the number of instruments k is fixed, the TS2SLS estimator is consistent, asymptotically normal, and also asymptotically more efficient than the TSIV estimator proposed by Angrist and Krueger (1992) under mild regularity conditions (Inoue and Solon, 2010).

2.2. **Inconsistency of TS2SLS under many instruments.** Our first main result shows that the TS2SLS estimator $\hat{\beta}_{\mathrm{TS2SLS}}$ is inconsistent for β under the many instruments asymptotics in (2.3). To elaborate, we insert the model (2.1)-(2.2) into the definition of $\hat{\beta}_{\mathrm{TS2SLS}}$ and express the estimation error as

$$\hat{\beta}_{\mathrm{TS2SLS}} - \beta = \{A_1 + A_2 + A_3 + A_3'\}^{-1}(B_1 + B_2 - A_2\beta), \tag{2.5}$$

where

$$\begin{aligned} A_1 &= \Pi' S_{ZZ}^{(1)} \Pi, \qquad A_2 = S_{Z\eta}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{Z\eta}^{(2)}, \qquad A_3 = S_{Z\eta}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} \Pi, \\ B_1 &= \Pi' \{ S_{Z\eta}^{(1)} \beta + S_{Z\epsilon}^{(1)} - S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{Z\eta}^{(2)} \beta \}, \qquad B_2 = S_{Z\eta}^{(2)\prime} S_{ZZ}^{(2)-1} (S_{Z\eta}^{(1)} \beta + S_{Z\epsilon}^{(1)}), \end{aligned}$$

with $S_{Z\eta}^{(s)} = \frac{1}{n_s} Z_s' \eta_s$ for $s = 1, 2$ and $S_{Z\epsilon}^{(1)} = \frac{1}{n_1} Z_1' \epsilon_1$. Under the conventional asymptotics with fixed k, the terms A_1 and B_1 dominate and the limiting distribution is obtained by $\hat{\beta}_{\text{TS2SLS}} - \beta = A_1^{-1} B_1 \{1 + o_p(1)\}$. On the other hand, under the many instruments asymptotics in (2.3), the term A_2 is of order $O_p(k/n_1)$ and causes the asymptotic bias for $\hat{\beta}_{\text{TS2SLS}}$.

To derive the asymptotic properties of $\hat{\beta}_{\text{TS2SLS}}$, we impose the following assumptions, which are standard in the TSIV and many-instrument literatures. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the minimum and maximum eigenvalues of a matrix A, respectively. For $s = 1, 2$, let $\epsilon_1 = (\epsilon_{11}, \ldots, \epsilon_{1n_1})'$, $\eta_s = (\eta_{s1}, \ldots, \eta_{sn_s})'$, $Z_s = (z_{s1}, \ldots, z_{sn_s})'$, and $P^{(s)} = Z_s (Z_s' Z_s)^{-1} Z_s'$ be the $n_s \times n_s$ projection matrix.

Assumption 1.

- **(i):** *The samples $[Y_1 : Z_1]$ and $[X_2 : Z_2]$ are independent. The rows of $[\epsilon_1 : \eta_1]$ are mutually independent with $\mathbb{E}[\epsilon_{1i} | Z_1] = 0$ and $\mathbb{E}[\eta_{1i} | Z_1] = 0$. The rows of η_2 are mutually independent with $\mathbb{E}[\eta_{2i} | Z_2] = 0$.*
- **(ii):** *For each $s = 1, 2$ and n_s, $P^{(s)}$ is of rank k, and there exists a positive constant C_1 independent of n_1 and n_2 such that $\max_i P_{ii}^{(s)} \le C_1 < 1$.*
- **(iii):** *(2.3) holds true and $H_n := \Pi' S_{ZZ}^{(1)} \Pi \xrightarrow{p} H$ as $n_1, n_2 \to \infty$ for some positive definite matrix H. There exists a positive constant C_2 independent of n_1 and n_2 such that*

$$\begin{aligned} &\max_i z_{1i}' (Z_2' Z_2)^{-1} z_{1i} \le C_2, \qquad \max_i z_{2i}' (Z_1' Z_1)^{-1} z_{2i} \le C_2, \\ &\frac{1}{C_2} \le \lambda_{\min}(S_{ZZ}^{(s)}) \le \lambda_{\max}(S_{ZZ}^{(s)}) \le C_2 \text{ for } s = 1, 2. \end{aligned}$$

- **(iv):** *There exists a positive constant C_3 such that for all n large enough $\max_{1 \le i \le n_1} \mathbb{E}[\epsilon_{1i}^4 | Z_1] \le C_3 < \infty$ and $\max_{1 \le i \le n_s} \mathbb{E}[\|\eta_{si}\|^4 | Z_s] \le C_3 < \infty$ for $s = 1, 2$.*

Assumption 1 (i) is standard for the TSIV setup. Although we consider the case where Z is random, similar results can be established for non-random Z. Assumption 1 (ii) is standard conditions on the projection matrix $P^{(s)}$. Assumption 1 (iii) is a key condition for the many instruments asymptotics. This assumption is analogous to the one imposed in the one-sample many instruments setup (e.g., Bekker, 1994). Assumption 1 (iv) imposes mild moment requirements on the error terms.

Under these assumptions, the probability limit of the TS2SLS estimator is characterized as follows.

Theorem 1. *Consider the setup of this section and suppose that Assumption 1 holds true. Then*

$$\hat{\beta}_{\text{TS2SLS}} \xrightarrow{p} \beta - \{H + \Lambda\}^{-1} \Lambda \beta = \{H + \Lambda\}^{-1} H \beta,$$

where $\Lambda = \text{plim} \frac{1}{n_1} \sum_{i=1}^{n_2} \mathbb{E}[\eta_{2i} \eta_{2i}' | Z_2] z_{2i}' (Z_2' Z_2)^{-1} (Z_1' Z_1) (Z_2' Z_2)^{-1} z_{2i}$.

We can show that when $\alpha = 0$, we have $\Lambda = 0$ so that the TS2SLS estimator is consistent. However, in general we have $\Lambda \neq 0$ so that the TS2SLS estimator $\hat{\beta}_{\text{TS2SLS}}$ is inconsistent for β. In the one-sample setup, the two-stage least squares estimator is also inconsistent, and is typically biased toward the OLS estimator (Bekker, 1994). On the other hand, in the two-sample setup, the TS2SLS estimator is biased toward the origin; see Angrist and Krueger (1995) for a related result on the exact bias of their split sample estimator.

2.3. **Bias-corrected estimator and its asymptotic normality.** An inspection of the proof of this theorem suggests that the main sources of the bias in $\hat{\beta}_{\text{TS2SLS}}$ is the term A_2 in (2.5), which is of order $O_p(k/n_1)$. Based on this insight, we propose the following bias-corrected TS2SLS estimator

$$\hat{\beta}_{\text{bc}} = \{S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{ZX}^{(2)} - A^*\}^{-1} S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZY}^{(1)}, \tag{2.6}$$

where the correction term A^* is defined as

$$A^* = \frac{1}{n_2^2} \sum_{i=1}^{n_2} x_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1} \sum_{j=1}^{n_2} M_{ij}^{(2)} x_{2j}',$$

with $M^{(2)} = I_{n_2} - Z_2(Z_2'Z_2)^{-1}Z_2'$ and $M_{ij}^{(2)}$ is its (i,j)-th element.[1] To see the rationale of this bias correction, note that

$$\mathbb{E}[A^*|Z] = \frac{1}{n_1} \sum_{i=1}^{n_2} \mathbb{E}[\eta_{2i}\eta_{2i}'|Z_2] z_{2i}' (Z_2'Z_2)^{-1} (Z_1'Z_1)(Z_2'Z_2)^{-1} z_{2i} = \mathbb{E}[A_2|Z],$$

with $Z = (Z_1, Z_2)$, i.e., A^* is an unbiased estimator of A_2. Indeed the subtraction of A^* in (2.6) eliminates the influences of A_2 in (2.5) for $\hat{\beta}_{\text{TS2SLS}}$ at the same time.

Formally, the asymptotic properties of the bias-corrected TS2SLS estimator $\hat{\beta}_{\text{bc}}$ is obtained as follows.

Theorem 2. *Consider the setup of this section and suppose that Assumption 1 holds true. Then* $\hat{\beta}_{\text{bc}} \xrightarrow{p} \beta$ *and*

$$(H^{-1}\Psi_n H^{-1})^{-1/2} \sqrt{n_1}(\hat{\beta}_{\text{bc}} - \beta) \xrightarrow{d} N(0, I_p),$$

where

$$\begin{aligned}\Psi_n \quad = \quad & \frac{1}{n_1}\Pi' \left(\sum_{i=1}^{n_1} z_{1i} z_{1i}' \sigma_{vi}^{2(1)} \right) \Pi + \frac{1}{n_1} \sum_{i=1}^{n_2} \sum_{j \neq i}^{n_2} \sum_{k \neq i}^{n_2} \tilde{P}_{ij}^{(2)} \tilde{P}_{ik}^{(2)} \Pi' z_{2j} z_{2k}' \Pi \beta' \Sigma_{\eta i}^{(2)} \beta \\ & + \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\tilde{P}_{ij}^{(1)})^2 \Sigma_{\eta j}^{(2)} \sigma_{vi}^{2(1)} + \frac{1}{n_1} \sum_{i=1}^{n_2} \sum_{j \neq i}^{n_2} (\tilde{P}_{ij}^{(2)})^2 \Sigma_{\eta j}^{(2)} \beta' \Sigma_{\eta i}^{(2)} \beta, \end{aligned} \tag{2.7}$$

[1] An alternative expression of A^* is

$$A^* = \frac{1}{n_2^2} \sum_{i=1}^{n_2} x_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} (x_{2i} - \hat{\Pi}_{(i)}' z_{2i})',$$

where $\hat{\Pi}_{(i)} = \left(\sum_{j \neq i}^{n_2} z_{2j} z_{2j}' \right)^{-1} \sum_{j \neq i}^{n_2} z_{2j} x_{2j}'$ is the leave-i-out OLS estimator for Π, by the fact that $x_{2i} - \hat{\Pi}_{(i)}' z_{2i} = M_{ii}^{(2)-1} \sum_{j=1}^{n_2} M_{ij}^{(2)} \eta_{2j}'$.

$\Sigma^{(2)}_{\eta i} = \mathbb{E}[\eta_{2i}\eta'_{2i}|Z_2]$, $\sigma^{2(1)}_{vi} = \mathbb{E}[v^2_{1i}|Z_1]$ *with* $v_{1i} = \eta'_{1i}\beta + \epsilon_{1i}$, *and*

$$\tilde{P}^{(1)}_{ij} = z'_{1i}(Z'_2 Z_2)^{-1} z_{2j},$$

$$\tilde{P}^{(2)}_{ij} = z'_{2i}(Z'_2 Z_2)^{-1} Z'_1 Z_1 (Z'_2 Z_2)^{-1} z_{2j} + \frac{z'_{2j}(Z'_2 Z_2)^{-1} Z'_1 Z_1 (Z'_2 Z_2)^{-1} z_{2j}}{n_2} M^{(2)-1}_{jj} P^{(2)}_{ij}.$$

This theorem establishes consistency and asymptotic normality of the bias-corrected TS2SLS estimator $\hat{\beta}_{\mathrm{bc}}$. We note that this variance formula allows for heteroskedastic errors. For Ψ_n, the first two terms dominates under the conventional asymptotic framework with a fixed number of instruments. On the other hand, all terms may be of the same order under the many instruments setup as in this paper.

2.4. **Variance estimation and inference.** To conduct inference on β, we develop the asymptotic variance estimator of $\hat{\beta}_{\mathrm{bc}}$. Based on the expression in (2.7), we propose to estimate the variance component Ψ_n by

$$\hat{\Psi}_n = \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2} \tilde{P}^{(1)}_{ij}\tilde{P}^{(1)}_{ik} x_{2j}x'_{2k}\hat{\sigma}^{2(1)}_{vi} + \frac{1}{n_1}\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2} \tilde{P}^{(2)}_{ij}\tilde{P}^{(2)}_{ik} x_{2j}x'_{2k}\hat{\beta}'\hat{\Sigma}^{(2)}_{\eta i}\hat{\beta}, \tag{2.8}$$

where $\hat{\sigma}^{2(1)}_{vi} = \frac{1}{M^{(1)}_{Z,ii}} y_{1i} m^{(1)\prime}_{Z,i} Y_1$ and $\hat{\Sigma}^{(2)}_{\eta i} = \frac{1}{M^{(2)}_{Z,ii}} x_{2i} m^{(2)\prime}_{Z,i} X_2$. To see the validity of $\hat{\sigma}^{2(1)}_{vi}$ and $\hat{\Sigma}^{(2)}_{\eta i}$, note that

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^{2(1)}_{vi}|Z] &= \mathbb{E}\left[(z'_{1i}\Pi\beta + v_{1i})\sum_{l=1}^{n_1}\frac{M^{(1)}_{il}}{M^{(1)}_{ii}} v_{1l}\middle| Z\right] = \sigma^{2(1)}_{vi}, \\
\mathbb{E}[\hat{\Sigma}^{(2)}_{\eta i}|Z] &= \mathbb{E}\left[(z'_{2i}\Pi + \eta_{2i})\sum_{l=1}^{n_2}\frac{M^{(2)}_{il}}{M^{(2)}_{ii}} \eta'_{2l}\middle| Z\right] = \Sigma^{(2)}_{\eta,i}.
\end{aligned}$$

Also the component H can be estimated by $\hat{H}_n = S^{(2)\prime}_{ZX} S^{(2)-1}_{ZZ} S^{(1)}_{ZZ} S^{(2)-1}_{ZZ} S^{(2)}_{ZX} - A^*$. Combining these estimated components, the asymptotic variance $H^{-1}\Psi_n H^{-1}$ is estimated by $\hat{H}^{-1}_n\hat{\Psi}_n\hat{H}^{-1}_n$. The consistency of this variance estimator is established as follows.

Theorem 3. *Consider the setup of this section and suppose that Assumption 1 holds true. Furthermore, there exists a positive constant C independent of n_1 and n_2 such that $\max_i |z'_{1i}\Pi| \leq C$ a.s. Then*

$$\hat{H}^{-1}_n\hat{\Psi}_n\hat{H}^{-1}_n - H^{-1}\Psi_n H^{-1} \xrightarrow{p} 0.$$

Based on this theorem, statistical inference on the structural parameters β can be conducted by the Wald statistic. For example, if we want to test the null hypothesis $\mathbb{H}_0 : \beta = \beta_0$ for some known β_0, the Wald test is given by $\mathbb{I}\{W(\beta_0) \geq \chi^2_{p,\alpha}\}$, where $W(\beta) = n_1(\hat{\beta}_{\mathrm{bc}} - \beta)'[\hat{H}^{-1}_n\hat{\Psi}_n\hat{H}^{-1}_n]^{-1}(\hat{\beta}_{\mathrm{bc}} - \beta)$ and $\chi^2_{p,\alpha}$ is the $(1-\alpha)$-th quantile of the χ^2_p distribution.

Remark 1 (Lagrange multiplier test)**.** Alternatively, one can construct a Lagrange multiplier test for $\mathbb{H}_0 : \beta = \beta_0$ based on the score-type statistic

$$\begin{aligned}\mathcal{S}(\beta_0) &= S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZY}^{(1)} - (S_{ZX}^{(2)\prime} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{ZX}^{(2)} - A^*)\beta_0 \\ &= B_1 + B_2 - (A_2 - A^*)\beta_0.\end{aligned}$$

By inspection of the proofs of the above theorems, we can see that

$$LM(\beta_0) := \mathcal{S}(\beta_0)' \hat{\Psi}_n^{-1} \mathcal{S}(\beta_0) \xrightarrow{d} \chi_p^2,$$

under $\mathbb{H}_0 : \beta = \beta_0$. Thus, the Lagrange multiplier test is obtained as $\mathbb{I}\{LM(\beta_0) \geq \chi^2_{p,\alpha}\}$.

2.5. **Extension: Many weak instruments.** Our approach can be extended to accommodate many and weak instruments. Let $r_n = \min\{\mu_1^2, \ldots, \mu_p^2\}$ be the concentration parameter, where $(\mu_1^2, \ldots, \mu_p^2)$ are defined in (A.2) in the Appendix. For the case of $p = 1$, it is written as $r_n = \Pi' Z_1' Z_1 \Pi$. Let also $S_n = diag(\mu_1, \ldots, \mu_p) Q_n'$, where Q_n is defined in (A.1) in the Appendix.

Theorem 4. *Consider the setup of this section and suppose that Assumption 1 holds true, with the exception that* $\Pi' S_{ZZ}^{(1)} \Pi \xrightarrow{p} H$ *as* $n_1, n_2 \to \infty$*. If* $\sqrt{k}/r_n \to 0$*, then it holds*

$$\left(\frac{1}{n_1} H_n^{-1} \Psi_n H_n^{-1} \right)^{-1/2} \left(\hat{\beta}_{\text{bc}} - \beta \right) \xrightarrow{d} N(0, I_p).$$

This theorem establishes asymptotic normality of the bias-corrected TS2SLS estimator $\hat{\beta}_{\text{bc}}$ under the many weak instruments asymptotics in the sense that $\frac{k}{n_1} \to [0, 1)$ and $\sqrt{k}/r_n \to 0$. The additional condition $\sqrt{k}/r_n \to 0$ is on the concentration parameter, which requires that the instruments cannot be arbitrarily weak.

Theorem 5. *Consider the setup of this section and suppose that Assumption 1 holds true, with the exception that* $\Pi' S_{ZZ}^{(1)} \Pi \xrightarrow{p} H$ *as* $n_1, n_2 \to \infty$*, and* $\sqrt{k}/r_n \to 0$*. Furthermore, if there exists a positive constant* C *independent of* n_1 *and* n_2 *such that* $\max_i |z_{1i}' \Pi| \leq C$ *a.s., then the followings hold true.*

(i) When k/r_n *is bounded for all* n_1 *and* n_2 *large enough,*

$$S_n \left(\frac{1}{n_1} \hat{H}_n^{-1} \hat{\Psi}_n \hat{H}_n^{-1} - \frac{1}{n_1} H_n^{-1} \Psi_n H_n^{-1} \right) S_n' \xrightarrow{p} 0.$$

(ii) When $k/r_n \to \infty$*,*

$$\frac{r_n}{k} S_n \left(\frac{1}{n_1} \hat{H}_n^{-1} \hat{\Psi}_n \hat{H}_n^{-1} - \frac{1}{n_1} H_n^{-1} \Psi_n H_n^{-1} \right) S_n' \xrightarrow{p} 0.$$

This theorem implies that the Wald test defined in Section 2 is asymptotically valid under the many weak instruments asymptotics. See Chao et al. (2023).

3. Empirical illustration

We revisit Angrist and Krueger's (1992) study, which investigates the effect of school-entry age on years of educational attainment using quarter of birth as an instrument. It is rare to observe such a long-term outcome variable jointly with the endogenous variable of interest and an instrument. To address this limitation, Angrist and Krueger (1992) use two data sources: the 1960 Census and the 1980 Census. We accessed these Census data through ICPSR.[2] We largely follow the sample selection procedure described in Appendix B of Angrist and Krueger (1992) and focus on Black and White boys born in the United States between 1946 and 1952, although exact replication appears difficult because the replication code for the original study is not available. The resulting sample sizes in our illustration are $n_1 = 442{,}095$ and $n_2 = 110{,}210$.

Following Angrist and Krueger (1992), we consider eight specifications in total (see Table 1). The first four differ in how they control for cohort trends. The first specification, MA(± 2), assumes that years of education are determined by school-entry age and a moving-average-type trend term, defined as $(m_{-2} + m_{-1} + m_1 + m_2)/4$, where m_s denotes the mean years of education for the cohort born s quarters before or after the current quarter. The second specification includes six year-of-birth dummies. The third and fourth specifications include, respectively, linear and quadratic trend terms. The remaining four specifications augment these first four specifications by additionally including state-of-birth (SOB) dummies. See Section 3.1 of Angrist and Krueger (1992) for details.

To deal with the potential endogeneity of the school-entry age, Angrist and Krueger (1992) consider two types of instruments. The first consists of the levels and interactions of three quarter-of-birth (QOB) dummies and six year-of-birth (YOB) dummies, yielding 27 instruments. The second consists of the levels and interactions of QOB dummies, YOB dummies, and 50 state-of-birth (SOB) dummies, yielding more than 900 instruments. The latter specification, in particular, raises concerns about many-instrument bias in the conventional TS2SLS estimator. We therefore estimate the effect using our proposed bias-corrected estimator and, for comparison, also report estimates based on the conventional estimator.

The estimation results are summarized in Table 1. Several noteworthy patterns emerge. First, when the number of instruments is relatively small compared to the sample size (i.e., rows 1–4), the conventional estimator $\hat{\beta}_{\text{TS2SLS}}$ and our proposed bias-corrected estimator $\hat{\beta}_{\text{bc}}$ are very similar across all specifications. This pattern is in line with the theory: when the number of instruments is small, the bias term is second-order and therefore negligible in samples of this size.

Second, in contrast, in the many-instrument cases (i.e., rows 5–8), the conventional estimates are uniformly smaller in absolute value than the bias-corrected estimates. This finding is in line with the theory suggesting that $\hat{\beta}_{\text{TS2SLS}}$ is generally biased toward the origin. The magnitudes of the estimated biases are also nonnegligible: under the assumption that the bias-corrected estimates are accurate, the bias of the conventional estimator amounts to more than 20% of the effect size. This highlights the importance of accounting for many-instrument bias in empirical applications.

[2]The 1960 Census is available at `https://doi.org/10.3886/ICPSR07756.v1` and the 1980 Census is available at `https://doi.org/10.3886/ICPSR08101.v2`.

Instruments	Covariates	$\hat{\beta}_{\text{TS2SLS}}$	$\hat{\beta}_{\text{bc}}$
QOB∗YOB	MA(±2)	−0.150 (0.036)	−0.151 (0.036)
QOB∗YOB	YOB	−0.046 (0.029)	−0.046 (0.030)
QOB∗YOB	YOBQ	−0.074 (0.031)	−0.074 (0.032)
QOB∗YOB	YOBQ, YOBQ^2	−0.152 (0.032)	−0.153 (0.033)
QOB∗YOB∗SOB	MA(±2), SOB	−0.161 (0.027)	−0.205 (0.034)
QOB∗YOB∗SOB	YOB, SOB	−0.118 (0.022)	−0.148 (0.029)
QOB∗YOB∗SOB	YOBQ, SOB	−0.144 (0.023)	−0.185 (0.031)
QOB∗YOB∗SOB	YOBQ, YOBQ^2, SOB	−0.186 (0.024)	−0.239 (0.032)

TABLE 1. Empirical estimates of the effects of age at school entry

Note: Analytical standard errors are reported in parentheses. Although our sample and estimation procedure differ from those in the original study, the qualitative implications are in line with those in Angrist and Krueger (1992).

4. NUMERICAL EXPERIMENT

We conclude the paper with a small simulation study assessing the finite-sample properties of our proposed procedure. Building on the simulation design of Donald and Newey (2001), we consider the following data-generating process:

$$Y_i = \beta X_i + \epsilon_i, \quad X_i = Z_i'\pi + \eta_i,$$

$$Z_i \sim N(0, I_k), \quad (\epsilon_i, \eta_i)' \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right).$$

The coefficients are set as follows: $\beta = 0.1$ and $\pi = (\pi_1, \ldots, \pi_k)'$, where

$$\pi_j = c_k \left(1 - \frac{j}{K+1} \right)^4, \quad j = 1, \ldots, k,$$

and c_k is chosen so that $\sum_{j=1}^{k} \pi_j^2 = R^2/(1-R^2)$, where R^2 denotes the theoretical first-stage R-squared, which is set to 0.1.

We observe n_1 pairs of (Y, Z) and n_2 pairs of (X, Z). For simplicity, the simulations set $n_1 = n_2 = n$ and $n \in \{500, 1000, 2000\}$. The number of instruments is set to $k = k_n \in \{n/100, n/20, n/10\}$. Based on 2000 simulation replications, we compare the average point estimates and coverage probabilities of 95% confidence intervals for the conventional and our proposed bias-corrected estimators. For the conventional estimator, we employ the heteroskedasticity-robust inference procedure of Pacini and Windmeijer (2016).

(n,k)	Point Estimate		Coverage Probability	
	$\hat{\beta}_{\mathrm{TS2SLS}}$	$\hat{\beta}_{\mathrm{bc}}$	$\hat{\beta}_{\mathrm{TS2SLS}}$	$\hat{\beta}_{\mathrm{bc}}$
$(500,5)$	0.086 (0.142)	0.097 (0.161)	94.8	95.2
$(500,25)$	0.067 (0.122)	0.107 (0.201)	93.1	95.7
$(500,50)$	0.048 (0.103)	0.107 (0.390)	90.1	96.7
$(1000,10)$	0.093 (0.100)	0.103 (0.111)	94.6	94.7
$(1000,50)$	0.065 (0.085)	0.099 (0.132)	92.2	95.4
$(1000,100)$	0.054 (0.071)	0.116 (0.158)	87.6	96.0
$(2000,20)$	0.091 (0.067)	0.100 (0.073)	95.7	95.7
$(2000,100)$	0.068 (0.057)	0.103 (0.086)	91.2	95.6
$(2000,200)$	0.051 (0.049)	0.105 (0.103)	81.8	96.4

TABLE 2. Simulation results

Note: Standard deviations, computed from 2,000 simulation replications, are reported in parentheses.

The simulation results are summarized in Table 2. The results indicate that although the conventional estimator $\hat{\beta}_{\mathrm{TS2SLS}}$ performs reasonably well when the number of instruments is small (i.e., $k = n/100$), it becomes severely biased toward zero as the number of instruments increases. The magnitude of the bias is substantial when $k = n/10$, amounting to roughly half of the true effect size. In contrast, across all scenarios we consider, the bias-corrected estimator is nearly unbiased. In line with this finding, the coverage probabilities of our proposed estimator are excellent, whereas the conventional estimator exhibits substantial size distortions, especially when the number of instruments is large. In particular, the empirical coverage probabilities of the bias-corrected confidence intervals are close to 95% in every case, whereas those of the conventional confidence intervals deteriorate to levels below 90% in some cases.

Appendix A. Mathematical appendix

Notation: Hereafter, let $||A|| = \sqrt{\mathrm{tr}[A'A]}$ for a vector or matrix A, $||X||_{2,Z} = \sqrt{\mathbb{E}[X^2|Z]}$ for a random variable X, and $A_N \lesssim B_N$ means that there exists a positive constant C independent of N such that $A_N \leq CB_N$ for all N a.s., $A_N \gtrsim B_N$ means $B_N \lesssim A_N$, and $A_N \asymp B_N$ means $A_N \lesssim B_N$ and $B_N \lesssim A_N$. Let $A^{1/2}$ denote a squared root matrix for a positive semidefinite matrix A, satisfying $A^{1/2}A^{1/2\prime} = A$, and $A^{-1/2} = (A^{1/2})^{-1}$ for nonsingular matrix A.

A.1. **Proof of Theorem 1.** From (2.5), it is sufficient for the conclusion to characterize the probability limits of the terms $(A_1, A_2, A_3, B_1, B_2)$. For A_1, Assumption says $A_1 \xrightarrow{p} H$. For A_3, it holds $\mathbb{E}[A_3|Z] = \mathbb{E}[S_{Z\eta}^{(2)\prime}|Z_2]S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}\Pi = 0$, and the (l,m)-th element $A_3^{(l,m)}$ of A_3 satisfies

$$\begin{aligned}\|A_3^{(l,m)}\|_{2,Z}^2 &= \frac{1}{n_2^2}\sum_{i=1}^{n_2}\mathbb{E}[(\eta_{2i}^{(l)})^2|Z]\Pi^{(m)\prime}S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}z_{2i}z_{2i}'S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}\Pi^{(m)} \\ &\lesssim \frac{1}{n_2}\Pi^{(m)\prime}S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}\Pi^{(m)} \lesssim n_1^{-1},\end{aligned}$$

where the inequality follows from Assumption (iv) and the second equality follow from Assumption (iii). Thus Chebyshev's inequality implies $A_3 \xrightarrow{p} 0$. For B_1, observe that $\mathbb{E}[B_1|Z] = \Pi'\{\mathbb{E}[S_{Z\eta}^{(1)}|Z_1]\beta + \mathbb{E}[S_{Z\epsilon}^{(1)}|Z_1] - S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}\mathbb{E}[S_{Z\eta}^{(2)}|Z_2]\beta\} = 0$, and the l-th element $B_1^{(l)}$ of B_1 satisfies

$$\begin{aligned}\|B_1^{(l)}\|_{2,Z}^2 &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\mathbb{E}[(\eta_{1i}'\beta + \epsilon_{1i})^2|Z_1]\Pi^{(l)\prime}z_{1i}z_{1i}'\Pi^{(l)} \\ &\quad + \frac{1}{n_2^2}\sum_{i=1}^{n_2}\mathbb{E}[(\eta_{2i}'\beta)^2|Z_2]\Pi^{(l)\prime}S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}z_{2i}z_{2i}'S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}\Pi^{(l)} \\ &\lesssim \frac{1}{n_1}\Pi^{(l)\prime}S_{ZZ}^{(1)}\Pi^{(l)} + \frac{1}{n_2}\Pi^{(l)\prime}S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}\Pi^{(l)} = O(n_1^{-1}).\end{aligned}$$

Thus we have $B_1 \xrightarrow{p} 0$. For B_2, observe that $\mathbb{E}[B_2|Z] = \mathbb{E}[S_{Z\eta}^{(2)\prime}|Z_1]S_{ZZ}^{(2)-1}\mathbb{E}[S_{Z\eta}^{(1)}\beta + S_{Z\epsilon}^{(1)}|Z_1] = 0$ and the l-th element $B_2^{(l)}$ of B_2 satisfies

$$\begin{aligned}\|B_2^{(l)}\|_{2,Z}^2 &= \frac{1}{n_1^2n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbb{E}[(\eta_{1i}'\beta + \epsilon_{1i})^2|Z_1]\mathbb{E}[(\eta_{2j}^{(l)})^2|Z_2](z_{1i}'S_{ZZ}^{(2)-1}z_{2j})^2 \\ &\lesssim \frac{1}{n_1^2n_2}\sum_{i=1}^{n_1}z_{1i}'S_{ZZ}^{(2)-1}z_{1i} \lesssim \frac{1}{n_1^2n_2}\sum_{i=1}^{n_1}z_{1i}'S_{ZZ}^{(1)-1}z_{1i} = \frac{k}{n_1n_2} \lesssim n_1^{-1},\end{aligned}$$

where the first inequality follows from Assumption (iv), the second inequality follows from Assumption (iii), and the second equality follows form $\frac{1}{n_1}\sum_{i=1}^{n_1}z_{1i}'S_{ZZ}^{(1)-1}z_{1i} = \mathrm{tr}[S_{ZZ}^{(1)-1}S_{ZZ}^{(1)}] = k$. Thus we have $B_2 \xrightarrow{p} 0$.

Finally, for A_2,

$$\mathbb{E}[A_2|Z] = \frac{1}{n_2^2}\sum_{i=1}^{n_2}\mathbb{E}[\eta_{2i}\eta_{2i}'|Z_2]z_{2i}'S_{ZZ}^{(2)-1}S_{ZZ}^{(1)}S_{ZZ}^{(2)-1}z_{2i},$$

and the (l,m)-th element $A_2^{(l,m)}$ of A_2 satisfies

$$
\begin{aligned}
&\|A_2^{(l,m)} - \mathbb{E}[A_2^{(l,m)}|Z]\|_{2,Z}^2 \\
=&\frac{1}{n_2^4}\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\{\mathbb{E}[(\eta_{2i}^{(l)})^2|Z_2]\mathbb{E}[(\eta_{2j}^{(m)})^2|Z_2] + \mathbb{E}[(\eta_{2i}^{(l)}\eta_{2i}^{(m)})|Z_2]\mathbb{E}[(\eta_{2j}^{(l)}\eta_{2j}^{(m)})|Z_2]\}(z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j})^2 \\
\lesssim&\frac{1}{n_2^3}\sum_{i=1}^{n_2} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} \lesssim n_1^{-1},
\end{aligned}
$$

where the first inequality follows from Assumption (iv) and the last equality follows from Assumption (iii). Therefore we have $A_2 \xrightarrow{p} \Lambda$. Combining these results, the conclusion follows. $\square$

A.2. **Proof of Theorem 2.** By using the notation in (2.5), the estimation error of $\hat{\beta}_{\text{bc}}$ can be written as

$$
\sqrt{n_1}(\hat{\beta}_{\text{bc}} - \beta) = \{A_1 + (A_2 - A^*) + A_3 + A_3'\}^{-1}\sqrt{n_1}(B_1 + B_2 - (A_2 - A^*)\beta).
$$

We first look at the denominator. Assumption (iii) guarantees $A_1 \xrightarrow{p} H$, and the proof of Theorem 1 implies $A_3 \xrightarrow{p} 0$. For $A_2 - A^*$, note that

$$
\begin{aligned}
A_2 - A^* \quad = \quad & \frac{1}{n_2^2}\sum_{i=1}^{n_2}\eta_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} \sum_{j\neq i}^{n_2} z_{2j}\eta_{2j}' \\
& -\frac{1}{n_2^2}\sum_{i=1}^{n_2}\Pi' z_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1}\sum_{j=1}^{n_2} M_{ij}^{(2)}\eta_{2j}' \\
& -\frac{1}{n_2^2}\sum_{i=1}^{n_2}\eta_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1}\sum_{j\neq i}^{n_2} M_{ij}^{(2)}\eta_{2j}' \\
=: \quad & D_1 - D_2 - D_3.
\end{aligned}
$$

The proof of Theorem 1 implies $D_1 \xrightarrow{p} 0$. For D_2, its (l,m)-th element satisfies

$$
\begin{aligned}
& \mathbb{E}[(D_2^{(l,m)})^2|Z] \\
\lesssim \quad & \frac{1}{n_2^4}\sum_{i=1}^{n_2}\sum_{i_1=1}^{n_2}\left[\begin{array}{c} \Pi^{(l)\prime} z_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1}\left(\sum_{j=1}^{n_2} M_{ij}^{(2)} M_{i_1 j}^{(2)}\right) \\ \times M_{i_1 i_1}^{(2)-1} z_{2i_1}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i_1} z_{2i_1}' \Pi^{(l)} \end{array}\right] \\
= \quad & \frac{1}{n_2^4}\sum_{i=1}^{n_2}\sum_{i_1=1}^{n_2}\left[\begin{array}{c} \Pi^{(l)\prime} z_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1} M_{ii_1}^{(2)} \\ \times M_{i_1 i_1}^{(2)-1} z_{2i_1}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i_1} z_{2i_1}' \Pi^{(l)} \end{array}\right] \\
\leq \quad & \frac{1}{n_2^4}\sum_{i=1}^{n_2}\left[\begin{array}{c} \Pi^{(l)\prime} z_{2i} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-2} \\ \times z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} z_{2i}' \Pi^{(l)} \end{array}\right] \\
\leq \quad & \frac{1}{n_2^3}\frac{\max_i (z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i})^2}{\min_i (M_{ii}^{(2)})^2}\Pi^{(l)\prime} S_{ZZ}^{(2)}\Pi^{(l)} \\
\lesssim \quad & n_1^{-1},
\end{aligned}
$$

where the first wave inequality follows from Assumption (iv), the first equality follows from $\sum_{j=1}^{n_2} M_{ij}^{(2)} M_{i_1 j}^{(2)} = M_{ii_1}^{(2)}$, the first inequality follows from the fact that $\sum_{i=1}^{n_2}\sum_{i_1=1}^{n_2} a_i' M_{ii_1}^{(2)} a_{i_1} \leq \sum_{i=1}^{n_2} a_i' a_i$ for any $(a_1, \ldots, a_n)'$, and the last wave inequality follows from Assumption (iii) and the fact that $\frac{1}{\min_i (M_{ii}^{(2)})^2} \lesssim_{as} 1$ under Assumption (ii) (Mikusheva and Sun, 2022, Lemma S1.1). Similarly, for D_3, its (l,m)-th element satisfies

$$
\begin{aligned}
& \mathbb{E}[(D_3^{(l,m)})^2 | Z] \\
= \ & \frac{1}{n_2^4} \sum_{i=1}^{n_2} \sum_{j \neq i}^{n_2} \sum_{i_1=1}^{n_2} \sum_{j_1 \neq i_1}^{n_2} \eta_{2i}^{(l)} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} M_{ii}^{(2)-1} M_{ij}^{(2)} \eta_{2j}^{(m)} \eta_{2i_1}^{(l)} z_{2i_1}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i_1} M_{i_1 i_1}^{(2)-1} M_{i_1 j_1}^{(2)} \eta_{2j_1}^{(m)} \\
\lesssim \ & \frac{1}{n_2^4} \frac{\max_i (z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i})^2}{\min_i (M_{ii}^{(2)})^2} \sum_{i=1}^{n_2} \sum_{j \neq i}^{n_2} M_{ij}^{(2)2} \\
\lesssim \ & \frac{1}{n_2^4} \frac{\max_i (z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i})^2}{\min_i (M_{ii}^{(2)})^2} \sum_{i=1}^{n_2} P_{ii}^{(2)} \\
\lesssim \ & n_1^{-1}.
\end{aligned}
$$

Combining these results, we obtain $A_2 - A^* \xrightarrow{p} 0$ and so we can write

$$
\sqrt{n_1}(\hat{\beta}_{\text{bc}} - \beta) = H^{-1} \sqrt{n_1} \{B_1 + B_2 - (A_2 - A^*)\beta\} + o_p(1).
$$

We now look at the numerator. Decompose

$$
\begin{aligned}
&\sqrt{n_1}\{B_1 + B_2 - (A_2 - A^*)\beta\} \\
&= \underbrace{\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \Pi' z_{1i}(\eta_{1i}'\beta + \epsilon_{1i})}_{=\sqrt{n_1}(B_1 + S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{Z\eta}^{(2)}\beta)} + \underbrace{\frac{1}{\sqrt{n_1} n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \eta_{1j} z_{2j}' S_{ZZ}^{(2)-1} z_{1i}(\eta_{1i}'\beta + \epsilon_{1i})}_{=\sqrt{n_1} B_2} - \sqrt{n_1}\Pi' S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{Z\eta}^{(2)}\beta \\
&\quad - \underbrace{\left\{ \begin{array}{c} \frac{\sqrt{n_1}}{n_2^2} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \eta_{2j} z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} \eta_{2i}'\beta \\ -\frac{\sqrt{n_1}}{n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} \Pi' z_{2j} z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j} M_{jj}^{(2)-1} M_{ji}^{(2)} \eta_{2i}'\beta \\ -\frac{\sqrt{n_1}}{n_2^2} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \eta_{2j} z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j} M_{jj}^{(2)-1} M_{ji}^{(2)} \eta_{2i}'\beta \end{array} \right\}}_{=\sqrt{n_1}(A_2 - A^*)\beta} \\
&= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \Pi' z_{1i}(\eta_{1i}'\beta + \epsilon_{1i}) \\
&\quad - \frac{\sqrt{n_1}}{n_2^2} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \Pi' z_{2j} \left(z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} + \frac{z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j}}{n_2} M_{jj}^{(2)-1} z_{2j}' S_{ZZ}^{(2)-1} \right) z_{2i} \eta_{2i}'\beta \\
&\quad + \frac{1}{\sqrt{n_1} n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \eta_{2j} z_{2j}' S_{ZZ}^{(2)-1} z_{1i}(\eta_{1i}'\beta + \epsilon_{1i}) \\
&\quad - \frac{\sqrt{n_1}}{n_2^2} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \eta_{2j} \left(z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} + \frac{z_{2j}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j}}{n_2} M_{jj}^{(2)-1} z_{2j}' S_{ZZ}^{(2)-1} z_{2i} \right) \eta_{2i}'\beta \\
&= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \Pi' z_{1i}(\eta_{1i}'\beta + \epsilon_{1i}) - \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \Pi' z_{2j} \tilde{P}_{ij}^{(2)} \eta_{2i}'\beta \\
&\quad + \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \eta_{2j} \tilde{P}_{ij}^{(1)} (\eta_{1i}'\beta + \epsilon_{1i}) - \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_2} \sum_{j\neq i}^{n_2} \eta_{2j} \tilde{P}_{ij}^{(2)} \eta_{2i}'\beta \\
&=: N_1 + N_2 + N_3 + N_4,
\end{aligned}
$$

where the second equality follows from $M_{ij}^{(2)} = -P_{ij}^{(2)}$ for $i \neq j$ and the third equality follows from the definition of $\tilde{P}_{ij}^{(1)}$ and $\tilde{P}_{ij}^{(2)}$, Noting that N_1, N_2, N_3 and N_4 are uncorrelated each other, by applying Chao et al. (2012, Lemma A2) and Matsushita and Otsu (2025, Lemma 4), the conclusion follows. □

A.3. **Proof of Theorem 3.** Lemmas 7 and 8 imply $\hat{\Psi} - \Psi_n \xrightarrow{p} 0$, and the proof of Theorem 2 implies $\hat{H} - H \xrightarrow{p} 0$. Thus the conclusion follows from the continuous mapping theorem. □

A.4. **Proof of Theorem 4.** By the spectral decomposition, there exists an orthogonal matrix $Q_n = (q_1, \ldots, q_p)$ such that $Q_n Q_n' = I$ and

$$n_1 Q_n' \Psi_n Q_n = \Lambda_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_p). \tag{A.1}$$

Based on $Q_n = (q_1, \ldots, q_p)$, denote

$$\mu_t^2 = n_1 q_t' H_n q_t = q_t' \Pi' Z_1' Z_1 \Pi q_t, \tag{A.2}$$

for $t = 1, \ldots, p$.

Recall that $r_n = \min\{\mu_1^2, \ldots, \mu_p^2\}$ and $S_n = \mathrm{diag}(\mu_1, \ldots, \mu_p) Q_n'$. First, if $\sqrt{k}/r_n \to 0$, we can show that

$$\begin{aligned} n_1 S_n'^{-1} \hat{H}_n S_n^{-1} &= n_1 S_n'^{-1} \left\{ A_1 + (A_2 - A^*) + A_3 + A_3' \right\} S_n^{-1} \\ &= n_1 S_n'^{-1} \left\{ A_1 + (D_1 - D_2 - D_3) + A_3 + A_3' \right\} S_n^{-1} \\ &= n_1 S_n'^{-1} H_n S_n^{-1} + o_p(1) \\ &= I + o_p(1), \end{aligned} \tag{A.3}$$

where the third equality follows from similar arguments to the proofs of Theorems 1 and 2. For example, the (l, m)-th element of $n_1 S_n^{-1} D_1 S_n^{-1}$ satisfies

$$\begin{aligned} & \left\| \left(n_1 S_n'^{-1} D_1 S_n^{-1} \right)^{(l,m)} \right\|_{2,Z}^2 \\ = & \frac{n_1^2}{n_2^4} \frac{1}{\mu_l^2 \mu_m^2} \sum_{i=1}^{n_2} \sum_{j \neq i}^{n_2} q_l' \Sigma_{\eta i}^{(2)} q_l q_m' \Sigma_{\eta j}^{(2)} q_m \left(z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2j} \right)^2 \\ \lesssim & \frac{n_1^2}{n_2^3} \frac{1}{\mu_l^2 \mu_m^2} \sum_{i=1}^{n_2} z_{2i}' S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} S_{ZZ}^{(1)} S_{ZZ}^{(2)-1} z_{2i} \\ \lesssim & \frac{k}{r_n^2}, \end{aligned}$$

where the first inequality follows from Assumption (iv) and the second inequality follows from Assumption (iii). Thus Markov's inequality implies that $n_1 S_n'^{-1} D_1 S_n^{-1} = o_p(1)$ under the condition $\sqrt{k}/r_n \to 0$.

Now letting $F_n = (n_1 \Psi_n)^{1/2\prime} S_n^{-1}$, we have

$$\begin{aligned} & \left(\frac{1}{n_1} H_n^{-1} \Psi_n H_n^{-1} \right)^{-1/2} \left(\hat{\beta} - \beta \right) = Q_n \left(n_1 S_n'^{-1} \Psi_n S_n^{-1} \right)^{-1/2} S_n \left(\hat{\beta} - \beta \right) \\ = & Q_n \left(n_1 S_n'^{-1} \Psi_n S_n^{-1} \right)^{-1/2} \left(n_1 S_n'^{-1} \hat{H}_n S_n^{-1} \right)^{-1} S_n'^{-1} \left\{ \sqrt{n_1} \left(N_1 + N_2 + N_3 + N_4 \right) \right\} \\ = & Q_n \left(n_1 S_n'^{-1} \Psi_n S_n^{-1} \right)^{-1/2} S_n'^{-1} \left(n_1 \Psi_n \right)^{1/2} \left\{ \left(n_1 \Psi_n \right)^{-1/2} \sqrt{n_1} \left(N_1 + N_2 + N_3 + N_4 \right) \right\} + o_p(1) \\ = & Q_n (F_n' F_n)^{-1/2} F_n' \left\{ \left(n_1 \Psi_n \right)^{-1/2} \sqrt{n_1} \left(N_1 + N_2 + N_3 + N_4 \right) \right\} + o_p(1), \end{aligned}$$

where the first equality follows from $H_n = \frac{1}{n_1} Q_n \mathrm{diag}(\mu_1^2, \ldots, \mu_p^2) Q_n'$ and $\Psi_n = \frac{1}{n_1} Q_n \Lambda_n Q_n'$, and the third equality follows from (A.3). Therefore, the same argument as the proof of Theorem 2 combined with the continuous mapping theorem and the fact that Q_n is an orthogonal matrix yields the conclusion. □

A.5. **Proof of Theorem 5.** By (A.3), we first note that

$$S_n\hat{H}_n^{-1}\hat{\Psi}_n\hat{N}_n^{-1}S_n' = (S_n'^{-1}\hat{H}_nS_n^{-1})^{-1}(S_n'^{-1}\hat{\Psi}_nS_n^{-1})(S_n'^{-1}\hat{H}_nS_n^{-1})^{-1}$$
$$= (S_n'^{-1}H_nS_n^{-1})^{-1}(S_n'^{-1}\hat{\Psi}_nS_n^{-1})(S_n'^{-1}H_nS_n^{-1})^{-1} + o_p(1).$$

For Part (i) (i.e., the case where k/r_n is bounded), we have

$$S_n'^{-1}(\hat{\Psi}_n - \Psi_n)S_n^{-1} = \mathrm{diag}(\mu_1^{-1}, \dots, \mu_p^{-1})Q_n'(\hat{\Psi}_n - \Psi_n)Q_n\mathrm{diag}(\mu_1^{-1}, \dots, \mu_p^{-1}) \xrightarrow{p} 0,$$

where the convergence follows from similar arguments to Lemmas 7 and 8.

For Part (ii) (i.e., the case where $k/r_n \to \infty$), we have

$$\begin{aligned}\frac{r_n}{k}S_n'^{-1}(\hat{\Psi}_n - \Psi_n)S_n^{-1} &= \frac{r_n}{k}\mathrm{diag}(\mu_1^{-1}, \dots, \mu_p^{-1})Q_n'(\hat{\Psi}_n - \Psi_n)Q_n\mathrm{diag}(\mu_1^{-1}, \dots, \mu_p^{-1}) \\ &\leq \frac{1}{k}Q_n'(\hat{\Psi}_n - \Psi_n)Q_n \xrightarrow{p} 0,\end{aligned}$$

where the convergence follows from similar arguments to Lemmas 7 and 8.

Combining these results, the conclusion follows. □

Appendix B. Lemmas

Lemma 6. *Suppose Assumption 1 holds true. Then*

- **(i):** $\max_{i,j}(\tilde{P}_{ij}^{(s)})^2 \lesssim 1$ *for* $s = 1, 2$;
- **(ii):** $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2 \lesssim k$, *and* $\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(2)})^2 \lesssim k$;
- **(iii):** $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^4 \lesssim k$, *and* $\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(2)})^4 \lesssim k$;
- **(iv):** $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2(\tilde{P}_{ik}^{(1)})^2 \lesssim k$, *and* $\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}(\tilde{P}_{ij}^{(2)})^2(\tilde{P}_{ik}^{(2)})^2 \lesssim k$;
- **(v):** $\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_2}|\tilde{P}_{ik}^{(1)}\tilde{P}_{jk}^{(1)}\tilde{P}_{i\ell}^{(1)}\tilde{P}_{j\ell}^{(1)}| \lesssim k$ *and* $\sum_{i<j<k<\ell}|\tilde{P}_{ik}^{(2)}\tilde{P}_{jk}^{(2)}\tilde{P}_{i\ell}^{(2)}\tilde{P}_{j\ell}^{(2)}| \lesssim k$.

Proof: For $\tilde{P}^{(1)}$, the proof is given by Matsushita and Otsu (2025, Lemma 1). For $\tilde{P}^{(2)}$, the conclusion follows from the same argument as Chao et al. (2012, Lemmas B1 and B2). □

Lemma 7. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$\frac{1}{n_1}\left\{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}x_{2j}x_{2k}'\hat{\sigma}_{vi}^{2(1)} - \Pi'\left(\sum_{i=1}^{n_1}z_{1i}z_{1i}'\sigma_{vi}^{2(1)}\right)\Pi - \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\Sigma_{\eta,j}^{(2)}\sigma_{vi}^{2(1)}\right\} \xrightarrow{p} 0.$$

Proof: Decompose

$$
\begin{aligned}
& \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}x_{2j}x_{2k}'\hat{\sigma}_{vi}^{2(1)} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}x_{2j}x_{2k}'\frac{1}{M_{Z,ii}^{(1)}}y_{1i}m_{Z,i}^{(1)\prime}Y_1 \\
= \quad & \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}x_{2j}x_{2k}'\left(z_{1i}'\Pi\beta + v_{1i}\right)\sum_{l=1}^{n_1}\frac{M_{il}^{(1)}}{M_{ii}^{(1)}}v_{1l} \\
= \quad & \sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)' v_{1i}^2 \\
& +\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)' v_{1i}\left(\sum_{l\neq i}^{n_1}\frac{M_{il}^{(1)}}{M_{ii}^{(1)}}v_{1l}\right) \\
& +\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)' (z_{1i}'\Pi\beta)\left(\sum_{l=1}^{n_1}\frac{M_{il}^{(1)}}{M_{ii}^{(1)}}v_{1l}\right) \\
=: \quad & A_1 + A_2 + A_3.
\end{aligned}
$$

For A_1, Lemma 9 implies

$$
\frac{1}{n_1}\left\{A_1 - \Pi'\left(\sum_{i=1}^{n_1}z_{1i}z_{1i}'\sigma_{vi}^{2(1)}\right)\Pi - \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\Sigma_{\eta,j}^{(2)}\sigma_{vi}^{2(1)}\right\} \xrightarrow{p} 0.
$$

For A_2 and A_3, Lemma 10 implies $\frac{1}{n_1}(A_2 + A_3) \xrightarrow{p} 0$. Therefore, the conclusion follows. □

Lemma 8. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$
\begin{aligned}
& \frac{1}{n_1}\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}x_{2j}x_{2k}'\hat{\beta}'\hat{\Sigma}_{\eta i}^{(2)}\hat{\beta} - \frac{1}{n_1}\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}\Pi' z_{2j}z_{2k}'\Pi\beta'\Sigma_{\eta i}^{(2)}\beta \\
& -\frac{1}{n_1}\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}(\tilde{P}_{ij}^{(2)})^2\Sigma_{\eta,j}^{(2)}\beta'\Sigma_{\eta i}^{(2)}\beta \\
\xrightarrow{p} \quad & 0.
\end{aligned}
$$

Proof: Decompose

$$\begin{aligned}
&\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}x_{2j}x_{2k}'\hat{\beta}'\hat{\Sigma}_{\eta i}^{(2)}\hat{\beta} \\
= \;& \sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}x_{2j}x_{2k}'\hat{\beta}'\left(x_{2i}m_{Z,i}^{(2)\prime}X_2\right)\hat{\beta} \\
= \;& \sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}x_{2j}x_{2k}'\hat{\beta}'\left\{\left(z_{2i}'\Pi+\eta_{2i}\right)\sum_{l=1}^{n_2}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}\eta_{2l}'\right\}\hat{\beta} \\
= \;& \sum_{i=1}^{n_2}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(\eta_{2i}'\beta)^2 \\
&+\sum_{i=1}^{n_1}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(\eta_{2i}'\beta)\left\{\sum_{l\neq i}^{n_1}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}(\eta_{2i}'\beta)\right\} \\
&+\sum_{i=1}^{n_1}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(z_{2i}'\Pi)\left\{\sum_{l=1}^{n_1}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}(\eta_{2i}'\beta)\right\} \\
&+\sum_{i=1}^{n_1}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)' \\
&\times\left[2\beta'\left\{\left(z_{2i}'\Pi+\eta_{2i}\right)\sum_{l=1}^{n_2}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}\eta_{2l}'\right\}(\hat{\beta}-\beta)+(\hat{\beta}-\beta)'\left\{\left(z_{2i}'\Pi+\eta_{2i}\right)\sum_{l=1}^{n_2}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}\eta_{2l}'\right\}(\hat{\beta}-\beta)\right] \\
=: \;& A_1+A_2+A_3+A_4.
\end{aligned}$$

For A_1, Lemma 11 implies

$$\frac{1}{n_1}\left\{A_1-\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}\Pi'z_{2j}z_{2k}'\Pi\beta'\Sigma_{\eta i}^{(2)}\beta-\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}(\tilde{P}_{ij}^{(2)})^2\Sigma_{\eta,j}^{(2)}\beta'\Sigma_{\eta i}^{(2)}\beta\right\}\xrightarrow{p}0.$$

For A_2 and A_3, Lemma 12 implies $\frac{1}{n_1}(A_2+A_3)\xrightarrow{p}0$. A similar argument with the fact that $\hat{\beta}-\beta=o_p(1)$ implies $\frac{1}{n_1}A_4\xrightarrow{p}0$. Therefore, the conclusion follows. □

Lemma 9. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$\frac{1}{n_1}\left\{\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)'v_{1i}^2-\Pi'\left(\sum_{i=1}^{n_1}z_{1i}z_{1i}'\sigma_{vi}^{2(1)}\right)\Pi-\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\Sigma_{\eta,j}^{(2)}\sigma_{vi}^{2(1)}\right\}\xrightarrow{p}0.$$

Proof: To simplify the presentation, we provide a proof for the case of $G = 1$. An extension to the cases of $G \geq 2$ is straightforward. Decompose

$$\begin{aligned}
& \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} x_{2j} \right) \left(\sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} x_{2j} \right)' v_{1i}^2 \\
= \;& \Pi' \left(\sum_{i=1}^{n_1} z_{1i} z_{1i}' v_{1i}^2 \right) \Pi + 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} \eta_{2j} (z_{1i}'\Pi) v_{1i}^2 + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_2} \tilde{P}_{ij}^{(1)} \tilde{P}_{ik}^{(1)} \eta_{2j} \eta_{2k} v_{1i}^2 \\
=: \;& B_1 + B_2 + B_3,
\end{aligned}$$

where the first equality follows from $\sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} z_{2j}'\Pi = z_{1i}'\Pi$. Let $V_{1i} = v_{1i}^2 - \mathbb{E}[v_{1i}^2|Z_1]$.

For B_1, observe that

$$\begin{aligned}
& \left\| B_1 - \Pi' \left(\sum_{i=1}^{n_1} z_{1i} z_{1i}' \mathbb{E}[v_{1i}^2|Z_1] \right) \Pi \right\|_{2,Z}^2 = \mathbb{E}\left[\left(\sum_{i=1}^{n_1} (\Pi' z_{1i})^2 V_{1i} \right)^2 \middle| Z \right] \\
= \;& \sum_{i=1}^{n_1} (\Pi' z_{1i})^4 \mathbb{E}[V_{1i}^2|Z_1] \lesssim \max_{1\leq i\leq n_1} (\Pi' z_{1i})^2 \max_{1\leq i\leq n_1} \mathbb{E}[V_{1i}^2|Z] \left(n_1 \Pi' S_{ZZ}^{(1)} \Pi \right) \lesssim n_1,
\end{aligned}$$

where the second inequality follows from Assumptions (ii) and (iii).

For B_2, we further decompose

$$\begin{aligned}
B_2 \;&=\; 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} \eta_{2j} (z_{1i}'\Pi) V_{1i} + 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} \eta_{2j} (z_{1i}'\Pi) \mathbb{E}[v_{1i}^2|Z_1] \\
&=: \; B_{21} + B_{22}.
\end{aligned}$$

For B_{21}, observe that

$$\begin{aligned}
||B_{21}||_{2,Z}^2 \;&=\; 4\mathbb{E}\left[\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \tilde{P}_{ij}^{(1)} \eta_{2j} (z_{1i}'\Pi) V_{1i} \sum_{i_1=1}^{n_1} \sum_{j_1=1}^{n_2} \tilde{P}_{i_1 j_1}^{(1)} \eta_{2j_1} (z_{1i_1}'\Pi) V_{1i_1} \middle| Z \right] \\
&=\; 4 \sum_{i=1}^{n} \sum_{j=1}^{n} (\tilde{P}_{ij}^{(1)})^2 \mathbb{E}[v_{2j}^2|Z_2] (z_{1i}'\Pi)^2 \mathbb{E}[V_{1i}^2|Z_1] \\
&\lesssim\; \sum_{i=1}^{n} \sum_{j=1}^{n} (\tilde{P}_{ij}^{(1)})^2 \lesssim n_1,
\end{aligned}$$

where the second equality follows from independence of the two sample and the last inequality follows from Lemma 6. For B_{22}, observe that

$$
\begin{aligned}
||B_{22}||_{2,Z}^2 &= 4\sum_{j=1}^{n_2}\mathbb{E}[\eta_{2j}^2|Z_2]\left\{\sum_{i=1}^{n_1}(\Pi' z_{1i})\tilde{P}_{ij}^{(1)}\mathbb{E}[v_{1i}^2|Z_1]\right\}^2 \\
&\lesssim \sum_{j=1}^{n_2}\sum_{i=1}^{n_1}\sum_{i_1=1}^{n_1}(\Pi' z_{1i})(\Pi' z_{1i_1})\tilde{P}_{ij}^{(1)}\tilde{P}_{i_1 j}^{(1)} \\
&= \Pi'(Z_2'Z_2)^{-1}\Pi \\
&= \Pi' Z_2' P^{(2)} Z_2 \Pi \lesssim n_1,
\end{aligned}
$$

where the last inequality follows from $\sum_{i=1}^n\sum_{j=1}^n x_i A_{ij} x_j \leq \lambda_{\max}(A)\sum_{i=1}^n x_j^2$ for any symmetric matrix A and $(x_1,\ldots,x_n)$ and Assumptions (i) and (ii).

For B_3, we further decompose

$$
\begin{aligned}
B_3 &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\eta_{2j}\eta_{2k}V_{1i} + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k\neq j}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\eta_{2j}\eta_{2k}\mathbb{E}[v_{1i}^2|Z_1] + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\eta_{2j}^2\mathbb{E}[v_{1i}^2|Z_1] \\
&=: B_{31}+B_{32}+B_{33}.
\end{aligned}
$$

For B_{31}, observe that

$$
\begin{aligned}
||B_{31}||_{2,Z}^2 &= \mathbb{E}\left[\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\eta_{2j}\eta_{2k}V_{1i}\sum_{i_1=1}^{n_1}\sum_{j_1=1}^{n_2}\sum_{k_1=1}^{n_2}\tilde{P}_{i_1j_1}^{(1)}\tilde{P}_{i_1k_1}^{(1)}\eta_{2j_1}\eta_{2k_1}V_{1i_1}\middle| Z\right] \\
&= 3\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2(\tilde{P}_{ik}^{(1)})^2\mathbb{E}[\eta_{2j}^2|Z_2]\mathbb{E}[\eta_{2k}^2|Z]\mathbb{E}[V_{1i}^2|Z_1] \\
&\lesssim \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2(\tilde{P}_{ik}^{(1)})^2 \lesssim n_1,
\end{aligned}
$$

where the second equality follows from independence of the two sample and the last inequality follows from Lemma 6 (iv). For B_{32}, observe that

$$
\begin{aligned}
||B_{32}||_{2,Z}^2 &= \mathbb{E}\left[\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k\neq j}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\eta_{2j}\eta_{2k}\mathbb{E}[v_{1i}^2|Z_1]\sum_{i_1=1}^{n_1}\sum_{j_1=1}^{n_2}\sum_{k_1\neq j_1}^{n_2}\tilde{P}_{i_1j_1}^{(1)}\tilde{P}_{i_1k_1}^{(1)}\eta_{2j_1}\eta_{2k_1}\mathbb{E}[v_{1i_1}^2|Z_2]\middle| Z\right] \\
&= 2\sum_{i=1}^{n_1}\sum_{i_1=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k\neq j}^{n_2}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\tilde{P}_{i_1j}^{(1)}\tilde{P}_{i_1k}^{(1)}\mathbb{E}[\eta_{2j}^2|Z_2]\mathbb{E}[\eta_{2k}^2|Z_2]\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1i_1}^2|Z_1] \\
&\lesssim n_1,
\end{aligned}
$$

where the inequality follows from Lemma 6 (v). For B_{33}, observe that

$$\begin{aligned}
& \mathbb{E}\left[\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\{\eta_{2j}^2-\mathbb{E}[\eta_{2j}^2|Z_2]\}\mathbb{E}[v_{1i}^2|Z_1]\right)^2\middle| Z\right] \\
= \ & \mathbb{E}\left[\sum_{i=1}^{n_1}\sum_{i_1=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2(\tilde{P}_{i_1j}^{(1)})^2\{\eta_{2j}^2-\mathbb{E}[\eta_{2j}^2|Z_2]\}^2\mathbb{E}[v_{1i}^2|Z_2]\mathbb{E}[v_{1i_1}^2|Z_1]\middle| Z\right] \\
\lesssim \ & n_1,
\end{aligned}$$

where the inequality follows from Lemma 6 (iv). Thus we have

$$\left\| B_3-\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2\mathbb{E}[\eta_{2j}^2|Z_2]\mathbb{E}[v_{1i}^2|Z_1]\right\|_{2,Z}^2 \lesssim n_1$$

The conclusion follows by Markov's inequality and the triangle inequality. □

Lemma 10. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)' v_{1i}\left(\sum_{l\neq i}^{n_1}\frac{M_{il}^{(1)}}{M_{ii}^{(1)}}v_{1l}\right)\xrightarrow{p}0, \tag{B.1}$$

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}x_{2j}\right)' (z_{1i}'\Pi\beta)\left(\sum_{l=1}^{n_1}\frac{M_{il}^{(1)}}{M_{ii}^{(1)}}v_{1l}\right)\xrightarrow{p}0. \tag{B.2}$$

Proof of (B.1): To simplify the presentation, we provide a proof for the case of $G=1$. An extension to the cases of $G\geq 2$ is straightforward. Decompose

$$\begin{aligned}
& \sum_{i=1}^{n_1}\left(\Pi' z_{1i}+\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}\eta_{2j}\right)^2 v_{1i}\sum_{\ell\neq i}^{n_2}\frac{M_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1\ell} \\
= \ & -\sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_2}\frac{(\Pi' z_{1i})^2P_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1i}v_{1\ell}-\sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_1}\frac{P_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1i}v_{1l}\left(\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}\eta_{2j}\right)^2 \\
& -2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{\ell\neq i}^{n_1}\frac{(\Pi_2' z_{1i})\tilde{P}_{ij}^{(1)}P_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1i}v_{1\ell}\eta_{2j} \\
=: \ & -T_1-T_2-2T_3.
\end{aligned}$$

For T_1, note that $\mathbb{E}[T_1|Z]=0$ and

$$\begin{aligned}
||T_1||_{2,Z}^2 \ &= \ \sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_2}\left\{\frac{(\Pi' z_{1i})^4(P_{i\ell}^{(1)})^2}{(M_{ii}^{(1)})^2}+\frac{(\Pi' z_{1i})^2P_{i\ell}^{(1)}}{M_{ii}^{(1)}}\frac{(\Pi' z_{1\ell})^2P_{\ell i}^{(1)}}{M_{\ell\ell}^{(1)}}\right\}\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1\ell}^2|Z_1] \\
&\lesssim \ \sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_2}(P_{i\ell}^{(1)})^2\leq\sum_{i=1}^{n_1}P_{ii}^{(1)}=k,
\end{aligned}$$

where the first inequality follows from Assumptions (i) and (iii). For T_2, note that $\mathbb{E}[T_2|Z]=0$ and

$$\begin{aligned}
||T_2||_{2,Z}^2 &= 3\sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\left(\frac{P_{i\ell}^{(1)}\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}}{M_{ii}^{(1)}}\right)^2\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1\ell}^2|Z_1]\mathbb{E}[\eta_{2j}^2|Z_2]\mathbb{E}[v_{2k}^2|Z_2] \\
&\quad +3\sum_{i=1}^{n_1}\sum_{\ell\neq i}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\frac{\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\tilde{P}_{\ell j}^{(1)}\tilde{P}_{\ell k}^{(1)}(P_{i\ell}^{(1)})^2}{M_{ii}^{(1)}M_{\ell\ell}^{(1)}}\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1\ell}^2|Z_1]\mathbb{E}[\eta_{2j}^2|Z_2]\mathbb{E}[v_{2k}^2|Z_2] \\
&\lesssim \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}(\tilde{P}_{ij}^{(1)})^2(\tilde{P}_{ik}^{(1)})^2\left(\sum_{\ell\neq i}^{n}(P_{i\ell}^{(1)})^2\right)+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_1}|\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}\tilde{P}_{\ell j}^{(1)}\tilde{P}_{\ell k}^{(1)}|(P_{i\ell}^{(1)})^2 \\
&\lesssim n_1,
\end{aligned}$$

where the first inequality follows from Assumptions and (i) and (iii), and the second inequality follows from Assumption (i) and Lemma 6 (iv) and (v). For T_3, note that $\mathbb{E}[T_3|Z]=0$ and

$$\begin{aligned}
||T_3||_{2,Z}^2 &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{\ell\neq i}^{n_1}\left(\frac{(\Pi' z_{1i})\tilde{P}_{ij}^{(1)}P_{i\ell}^{(1)}}{M_{ii}^{(1)}}\right)^2\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1\ell}^2|Z_1]\mathbb{E}[\eta_{2j}^2|Z_2] \\
&\quad +\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{\ell\neq i}^{n_1}\frac{(\Pi' z_{1i})(\Pi' z_{1\ell})\tilde{P}_{ij}^{(1)}\tilde{P}_{\ell j}^{(1)}(P_{i\ell}^{(1)})^2}{M_{ii}^{(1)}M_{\ell\ell}^{(1)}}\mathbb{E}[v_{1i}^2|Z_1]\mathbb{E}[v_{1\ell}^2|Z_1]\mathbb{E}[\eta_{2j}^2|Z_2] \\
&\lesssim \sum_{i=1}^{n_1}(\Pi' z_{1i})^2(P_{ii}^{(1)})+\sum_{i=1}^{n_1}(\Pi_2' z_{1i})^2 \\
&\lesssim n_1,
\end{aligned}$$

where the first inequality follows from Assumptions (i) and (iii), and $\sum_{i=1}^{n}\sum_{j=1}^{n}x_iA_{ij}x_j\leq\lambda_{\max}(A)\sum_{i=1}^{n}x_j^2$ for any symmetric matrix A and $(x_1,\dots x_n)$, and the last inequality follows from tAssumptions (ii). The conclusion follows by Markov's inequality and the triangle inequality.

Proof of (B.2): To simplify the presentation, we provide a proof for the case of $G=1$. An extension to the cases of $G\geq 2$ is straightforward. Decompose

$$\begin{aligned}
&\sum_{i=1}^{n_1}\left(\Pi' z_{1i}+\sum_{j=1}^{n_2}\tilde{P}_{ij}^{(1)}\eta_{2j}\right)^2(z_{1i}'\Pi\beta)\sum_{\ell=1}^{n_2}\frac{M_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1\ell} \\
&= \sum_{i=1}^{n_1}\sum_{\ell=1}^{n_1}\frac{(\Pi' z_{1i})^2(z_{1i}'\Pi\beta)M_{i\ell}^{(1)}}{M_{ii}^{(1)}}v_{1\ell}+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_1}\frac{(z_{1i}'\Pi\beta)\tilde{P}_{ij}^{(1)}\tilde{P}_{ik}^{(1)}M_{i\ell}^{(1)}}{M_{ii}^{(1)}}\eta_{2j}\eta_{2k}v_{1\ell} \\
&\quad +2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_1}\frac{(\Pi' z_{1i})(z_{1i}'\Pi\beta)\tilde{P}_{ij}^{(1)}M_{i\ell}^{(1)}}{M_{ii}^{(1)}}\eta_{2j}v_{1\ell} \\
&=: T_1+T_2+2T_3.
\end{aligned}$$

For T_1, note that $\mathbb{E}[T_1|Z] = 0$ and

$$\begin{aligned}
||T_1||_{2,Z}^2 &= \sum_{\ell=1}^{n_1} \mathbb{E}[v_{1\ell}^2|Z_1] \left(\sum_{i=1}^{n_1} \frac{(\Pi' z_{1i})^2 (z_{1i}'\Pi\beta) M_{i\ell}^{(1)}}{M_{ii}^{(1)}} \right)^2 \\
&\lesssim \sum_{\ell=1}^{n_1} \left(\sum_{i=1}^{n_1} \frac{(\Pi' \tilde{z}_{1i})^2 (z_{1i}'\Pi\beta) M_{i\ell}^{(1)}}{M_{ii}^{(1)}} \right)^2 \\
&\lesssim \sum_{i=1}^{n_1} \sum_{i_1=1}^{n_1} \frac{(\Pi' z_{1i})^2 (z_{1i}'\Pi\beta)}{M_{ii}^{(1)}} \frac{(\Pi' z_{1i_1})^2 (z_{1i_1}'\Pi\beta)}{M_{i_1 i_1}^{(1)}} M_{ii_1}^{(1)} \\
&\lesssim \sum_{i=1}^{n_1} \left\{ \frac{(\Pi' z_{1i})^2 (z_{1i}'\Pi\beta)}{M_{ii}^{(1)}} \right\}^2 \lesssim \max_{1\le i\le n_1} \{(\Pi' z_{1i})^2 (z_{1i}'\Pi\beta)\} \sum_{i=1}^{n} (\Pi' z_{1i})^2 \lesssim n_1,
\end{aligned}$$

where the third inequality follows from $\sum_{i=1}^n \sum_{j=1}^n x_i A_{ij} x_j \le \lambda_{\max}(A) \sum_{i=1}^n x_j^2$ for any symmetric matrix A and $(x_1, \ldots x_n)$, and the fifth inequality follows from Assumption (ii) and the assumption $\max_{1\le i\le n_1} |z_{1i}'\Pi| \le C$. For T_2, note that $\mathbb{E}[T_2|Z] = 0$ and

$$\begin{aligned}
||T_2||_{2,Z}^2 &= 3 \sum_{i=1}^{n_1} \sum_{i_1=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_2} \sum_{\ell \ne i, i_1}^{n_2} \frac{(z_{1i}'\Pi\beta) \tilde{P}_{ij}^{(1)} \tilde{P}_{ik}^{(1)} P_{i\ell}^{(1)}}{M_{ii}^{(1)}} \frac{(z_{1i_1}'\Pi\beta) \tilde{P}_{i_1 j}^{(1)} \tilde{P}_{i_1 k}^{(1)} P_{i_1 \ell}^{(1)}}{M_{i_1 i_1}^{(1)}} \mathbb{E}[\eta_{2j}^2|Z_2] \mathbb{E}[\eta_{2k}^2|Z_2] \mathbb{E}[v_{1\ell}^2|Z_1] \\
&\lesssim \sum_{i=1}^{n_1} \sum_{i_1=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_2} |\tilde{P}_{ij}^{(1)} \tilde{P}_{ik}^{(1)} \tilde{P}_{i_1 j}^{(1)} \tilde{P}_{i_1 k}^{(1)}| \left(\sum_{\ell \ne i, i_1}^{n_2} |P_{i\ell}^{(1)} P_{i_1 \ell}^{(1)}| \right) \lesssim n_1,
\end{aligned}$$

where the first inequality follows from Assumptions (i) and (ii), and the assumption $\max_{1\le i\le N} |z_{1i}'\Pi| \le C$, and the second inequality follows from Lemma 6 (v) and Assumption (i). For T_3, note that $\mathbb{E}[T_3|Z] = 0$ and

$$\begin{aligned}
||T_3||_{2,Z}^2 &= \sum_{j=1}^{n_2} \sum_{\ell=1}^{n_1} \mathbb{E}[\eta_{2j}^2|Z_2] \mathbb{E}[v_{1\ell}^2|Z_1] \left(\sum_{i=1}^{n_1} \frac{(\Pi' z_{1i})(z_{1i}'\Pi\beta) \tilde{P}_{ij}^{(1)} M_{i\ell}^{(1)}}{M_{ii}^{(1)}} \right)^2 \\
&\lesssim \sum_{j=1}^{n_2} \sum_{\ell=1}^{n_1} \left(\sum_{i=1}^{n_1} \frac{(\Pi' z_{1i})(z_{1i}'\Pi\beta) \tilde{P}_{ij}^{(1)} M_{i\ell}^{(1)}}{M_{ii}^{(1)}} \right)^2 \\
&\lesssim \sum_{i=1}^{n_1} \sum_{i_1=1}^{n_1} \frac{(\Pi' z_{1i})(z_{1i}'\Pi\beta)}{M_{ii}^{(1)}} \frac{(\Pi' z_{1i_1})(z_{1i_1}'\Pi\beta)}{M_{i_1 i_1}^{(1)}} \tilde{P}_{ii_1}^{(1)} M_{ii_1}^{(1)} \\
&\lesssim \sum_{i=1}^{n_1} \left\{ \frac{(\Pi' z_{1i})(z_{1i}'\Pi\beta)}{M_{ii}^{(1)}} \tilde{P}_{ii}^{(1)} \right\}^2 \lesssim \max_{1\le i\le n_1} \{(z_{1i}'\Pi\beta)^2\} \sum_{i=1}^{n_1} (\Pi' z_{1i})^2 \lesssim n_1,
\end{aligned}$$

where the third inequality follows from $\sum_{i=1}^n \sum_{i_1=1}^n x_i A_{ij} x_j \le \lambda_{\max}(A) \sum_{i=1}^n x_j^2$ for any symmetric matrix A and $(x_1, \ldots x_n)$, and the fifth inequality follows from Assumptions (i) and (ii), and the assumption $\max_{1\le i\le n_1} |z_{1i}'\Pi| \le C$. The conclusion follows by Markov's inequality and the triangle inequality. $\square$

Lemma 11. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$\frac{1}{n_1}\left\{\begin{array}{c}\sum_{i=1}^{n_2}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(\eta_{2i}'\beta)^2\\ -\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}\sum_{k\neq i}^{n_2}\tilde{P}_{ij}^{(2)}\tilde{P}_{ik}^{(2)}\Pi' z_{2j}z_{2k}'\Pi\beta'\Sigma_{\eta i}^{(2)}\beta-\sum_{i=1}^{n_2}\sum_{j\neq i}^{n_2}(\tilde{P}_{ij}^{(2)})^2\Sigma_{\eta,j}^{(2)}\beta'\Sigma_{\eta i}^{(2)}\beta\end{array}\right\}\xrightarrow{p}0.$$

Proof: Similar to the proof of Lemma 9. □

Lemma 12. *Suppose that the assumptions of Theorem 3 hold true. Then it holds*

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(\eta_{2i}'\beta)\left\{\sum_{l\neq i}^{n_1}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}(\eta_{2l}'\beta)\right\}\xrightarrow{p}0, \tag{B.3}$$

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\sum_{j\neq i}^{n_2}\tilde{P}_{ij}^{(2)}x_{2j}\right)\left(\sum_{k\neq i}^{n_2}\tilde{P}_{ik}^{(2)}x_{2k}\right)'(z_{2i}'\Pi)\left\{\sum_{l=1}^{n_1}\frac{M_{il}^{(2)}}{M_{ii}^{(2)}}(\eta_{2l}'\beta)\right\}\xrightarrow{p}0.. \tag{B.4}$$

Proof: Similar to the proof of Lemma 10. □

References

Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87:328–336.

Angrist, J. D. and Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, 13:225–235.

Arellano, M. and Meghir, C. (1992). Female labour supply and on-the-job search: an empirical model estimated using complementary datasets. *Review of Economic Studies*, 59:537–559.

Bekker, P. A. (1994). Alternative approximation to the distribution of instrumental variable estimators. *Econometrica*, 62:657–681.

Burgess, S., Small, D. S., and Thompson, S. G. (2017). A review of instrumental variable estimators for mendelian randomization. *Statistical Methods in Medical Research*, 26:2333–2355.

Chao, J. and Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, 73:1673–1692.

Chao, J., Swanson, N. R., Hausman, J. A., Newey, W. K., and Woutersen, T. (2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory*, 28:42–86.

Chao, J., Swanson, N. R., and Woutersen, T. (2023). Jackknife estimation of a cluster-sample IV regression model with many weak instruments. *Journal of Econometrics*, 235:1747–69.

Choi, J., Gu, J., and Shen, A. (2018). Weak-instrument robust inference for two-sample instrumental variables regression. *Journal of Applied Econometrics*, 33:109–125.

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

Han, C. and Phillips, P. C. B. (2006). GMM with many moment conditions. *Econometrica*, 74:147–192.

Inoue, A. and Solon, G. (2010). Two sample instrumental variables estimators. *Review of Economics and Statistics*, 92:557–561.

Jerrim, J., Choi, A., and Rodríguez, R. S. (2014). Two-sample two-stage least squares (TSTSLS) estimates of earnings mobility: how consistent are they? Technical Report 14-17, Institute of Education, University of London.

Klevmarken, N. A. (1982). Missing variables and two-stage least squares estimation from more than one dataset. Technical Report 62, Research Institute of Industrial Economics, Stockholm, Sweden.

Kunitomo, N. (1980). Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association*, 75:693–700.

Matsushita, Y. and Otsu, T. (2025). Cross-fitted instrumental variable estimation with many instruments and many included exogenous regressors. Technical report, Working paper.

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

Mikusheva, A. and Sun, L. (2024). Weak identification with many instruments. *Econometrics Journal*, 27:C1–C28.

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

Newey, W. and Windmeijer, F. (2009). GMM with many weak moment conditions. *Econometrica*, 77:687–719.

Olivetti, C. and Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. *American Economic Review*, 105(8):2695–2724.

Pacini, D. and Windmeijer, F. (2016). Robust inference for the two-sample 2SLS estimator. *Economics Letters*, 146:50–54.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. In Heckman, J. J. and Leamer, H. E., editors, *Handbook of Econometrics*, volume 6B, chapter 75, pages 5469–5547.

Zhao, Q., Wang, J., Hemani, G., Bowden, J., and Small, D. S. (2020). Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *The Annals of Statistics*, 48(3):1742–1769.

Zhao, Q., Wang, J., Spiller, W., Bowden, J., and Small, D. S. (2019). Two-sample instrumental variable analyses using heterogeneous samples. *Statistical Science*, 34:317–333.

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

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

(Y. Okamoto) Graduate School of Economics, Kyoto University, Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan.

Email address: okamoto.yuuta.57w@st.kyoto-u.ac.jp

(T. Otsu) Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

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