**Web-based supporting materials**

**"Friendship Networks and Academic Success: The Impact of Class Size and Gender Composition"**

**by**

**Winfried Pohlmeier, Livia Shkoza, Derya Uysal**

April 21, 2026

# Invertibility of the Pooled Moment Matrices

In general, there is no simple closed-form expression for

$$(A_1 + A_2)^{-1}$$

in terms of $A_1^{-1}$ and $A_2^{-1}$, even if $A_1$ and $A_2$ are symmetric.

*(i) Under Invertibility:*

If $A_1$ (or alternatively $A_2$) is invertible, the matrix inversion lemma (also known as the Woodbury formula) applies. The general form is:

$$(A_1 + A_2)^{-1} = A_1^{-1} - A_1^{-1}(I + A_2 A_1^{-1})^{-1} A_2 A_1^{-1}. \tag{WA.1}$$

Alternatively, this can also be written symmetrically as:

$$(A_1 + A_2)^{-1} = A_2^{-1} - A_2^{-1}(I + A_1 A_2^{-1})^{-1} A_1 A_2^{-1}. \tag{WA.2}$$

Both equations are equivalent when $A_1$ and $A_2$ are invertible. However in our case $A_1$ and $A_2$ are both singular. The Woodbury formulas (WA.1) and (WA.2) do **not** apply for our case.

*(ii.) Under Non-Invertibility:*

In our case we have to take a closer look at the structure of $X_1'X_1 + X_2'X_2$. Let us take a look at the moment matrix of network $l$:

$$\frac{1}{n_l}X_l'X_l = \frac{1}{n_l}\begin{pmatrix} \sum_i x_{i,l}^2 & v_l \sum_i x_{i,l}^2 \\ v_l \sum_i x_{i,l}^2 & v_l^2 \sum_i x_{i,l}^2 \end{pmatrix} = \begin{pmatrix} a_l & v_l a_l \\ v_l a_l & v_l^2 a_l \end{pmatrix}$$

where $a_l = \frac{1}{n_l}\sum_i x_{i,l}^2 > 0$. For the sum of the moment matrices we get:

$$M = \frac{1}{n_1}X_1'X_1 + \frac{1}{n_2}X_2'X_2 = \begin{pmatrix} a_1 + a_2 & v_1 a_1 + v_2 a_2 \\ v_1 a_1 + v_2 a_2 & v_1^2 a_1 + v_2^2 a_2 \end{pmatrix} \tag{WA.3}$$

The determinant of the moment matrix $M$ is given by:

$$\begin{aligned} \det(M) =& (a_1 + a_2)(v_1^2 a_1 + v_2^2 a_2) - (v_1 a_1 + v_2 a_2)^2 \\ =& a_1 a_2 (v_1 - v_2)^2, \end{aligned}$$

which is positive for $v_1 \neq v_2$, i.e. if the network specific factors differ.

## *Minimum Distance IV Estimator*

In the following we show that our proposed MVIV estimator has a Minimum Distance (MD) representation where the single network parameters are estimated by IV-GMM in the first stage stage.

The first estimation stage of the Minimum Distance Instrumental Variable (MDIV) estimator is based on the IV regression:

$$\mathbf{Y}_l \quad = \quad \mathbf{W}_l \, \pi_l + \varepsilon_l \,, \qquad (l = 1, \ldots, L) \tag{WA.4}$$

where $\pi_l = \left(\gamma, \gamma_g, \beta_{g,l}\right)'$ is the first-stage ("reduced form") parameter vector. This equation is in fact same as (8). The $n_l \times k_w$ dimensional regressor matrix $\mathbf{W}_l = \left[x_l \; x_l^g \; y_l^g\right]$ has full column rank.

Assuming heteroskedasticity within-network the IV-GMM estimator for network $l$ is:

$$\hat{\pi}_l = \left(\mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l\right)^{-1} \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{Y}_l \tag{WA.5}$$

with $\hat{\mathrm{V}}\left[\hat{\pi}_l\right] = \left(\mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l\right)^{-1}$.

The restriction between the first-stage reduced form parameter vector $\pi_l$ and the structural form parameter vector $\theta$ is given by

$$\pi_l(\theta) = M_l \, \theta \,, \tag{WA.6}$$

with

$$M_l = \begin{bmatrix} I_{k_x} & 0 & 0 & 0 \\ 0 & I_{k_x} & 0 & 0 \\ 0 & 0 & 1 & v_l \end{bmatrix} .$$

where $k_x = 1$ for the simplified model with only one regressor. Stacking the restrictions between the $L$ reduced form parameter vectors $\pi_l$ and the structural form parameter $\theta$

given by (WA.6) into a $(k_w L \times 1)$-vector containing the complete set of restrictions gives:

$$\pi(\theta) = M\theta\,, \tag{WA.7}$$

with $\pi(\theta) = (\pi_1(\theta)', \pi_2(\theta)', \ldots, \pi_L(\theta)')'$ and $M = (M_1', M_2', \ldots, M_L')'$.

In the second step we estimate $\theta$ by minimizing the weighted quadratic distance between the estimated reduced form parameter vector $\hat{\pi}(\theta) = \left(\hat{\pi}_1', \hat{\pi}_2', \ldots, \hat{\pi}_L'\right)'$ and $M\theta$ with respect to the structural parameter vector $\theta$ based on the estimated optimal weight matrix $\hat{\Omega}^{-1} = \hat{\mathrm{V}}\left[\hat{\pi}\right]^{-1}$:

$$\hat{\theta}_{MDIV} \equiv \arg\min_{\theta} \left[\hat{\pi} - M\theta\right]' \hat{\Omega}^{-1} \left[\hat{\pi} - M\theta\right], \tag{WA.8}$$

Because of the independence of the networks, the estimated weighting matrix is block-diagonal of the form: $\hat{\Omega}^{-1} = \mathrm{diag}\left[\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1}\right] = \mathrm{diag}\left[\,\mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l\,\right]$. Moreover, due to the linearity of the restriction between $\pi$ and $\theta$, the MDIV can be represented as a generalized least square estimator of a regression of $\hat{\pi}$ on $M$:

$$\begin{aligned} \hat{\theta}_{MDIV} &= \left(M'\hat{\Omega}^{-1}M\right)^{-1} M'\hat{\Omega}^{-1}\hat{\pi} \\ &= \left(\sum_{l=1}^{L} M_l' \hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} M_l\right)^{-1} \left(\sum_{l=1}^{L} M_l' \hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} \hat{\pi}_l\right). \end{aligned} \tag{WA.9}$$

## Equivalence of MVIV and MDIV

**Proposition 1** (Equivalence of Estimators)

For the a heteroskedastic network design given by Assumption A.1 the Minimum Distance estimator $\hat{\theta}_{MDIV}$ given by (WA.9) and the the MVIV estimator $\hat{\theta}_{MVIV}$ given by (15) are numerically equivalent.

*Proof* 1 (Equivalence of MVIV and MDIV)

Decomposing the first stage estimate of the reduced form parameter vector and using the restriction $\pi_l = M_l\theta$:

$$\hat{\pi}_l = \pi_l + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l \tag{WA.10}$$

$$= M_l\theta + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l \tag{WA.11}$$

yields the first stage estimator in terms of $\theta$. Inserting (WA.11) into (WA.9) gives:

$$\begin{aligned}
\hat{\theta}_{MDIV} =& \Big(\sum_{l=1}^{L} M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\,M_l\Big)^{-1}\cdot\Big(\sum_{l=1}^{L} M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\,[M_l\theta + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l]\Big)\\
=& \,\theta + \Big(\sum_{l=1}^{L} M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\,M_l\Big)^{-1}\cdot\Big(\sum_{l=1}^{L} M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l\Big)\\
=& \,\theta + \Big(\sum_{l=1}^{L} \mathbf{X}'\,\mathbf{P}_l\,\mathbf{X}_l\Big)^{-1}\cdot\Big(\sum_{l=1}^{L} \mathbf{X}'\,\mathbf{P}_l\,\varepsilon_l\Big)
\end{aligned}$$

where we used to obtain the final equation the following relationships:

- $\mathbf{W}_l\,M_l = \mathbf{X}_l$
- $M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\,M_l = M_l[\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l]M_l = \mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{X}_l$
- $M_l'\hat{\mathbb{V}}\,[\hat{\pi}_l]^{-1}\left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l = \mathbf{X}_l'\,\mathbf{P}_l\,\varepsilon_l$

Inserting (12) into (15) yields the decomposed $\hat{\theta}_{GMM}$ and the desired result:

$$\hat{\theta}_{MVIV} = \theta + \big(\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{X}_l\big)^{-1}\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\varepsilon_l = \hat{\theta}_{MDIV}$$

□

**Practical Note on the Equivalence of MVIV and MDIV:** One can verify the equivalence numerically by constructing $P_l$ from a common set of classwise reduced-form residuals $Y_l = W_l\pi_l + \varepsilon_l$ and plugging the same $P_l$ into both expressions. However, the natural implementation of the two estimators leads to different first-stage choices, and therefore to different weighting matrices. In the MVIV formulation, the most natural first stage is a system-level regression that stacks all classes and estimates

$$\hat{\theta}_1 = \left( \sum_{l=1}^{L} X_l' P_{1,l} X_l \right)^{-1} \left( \sum_{l=1}^{L} X_l' P_{1,l} Y_l \right),$$

after which the class-specific residuals $e_l = Y_l - X_l \hat{\theta}_1$ are used to form $P_l$. In contrast, the MDIV approach naturally relies on classwise reduced-form first stages, estimating each $\hat{\pi}_l$ separately from $Y_l = W_l \pi_l + \varepsilon_l$ and constructing $P_l$ from the corresponding class-level residuals.

Because these two first-stage strategies differ, the resulting weighting matrices $P_l$ also differ in practice, and the two-step GMM estimators no longer coincide in finite samples even though they remain asymptotically equivalent. Thus, the equality $\hat{\theta}^{MVIV} = \hat{\theta}^{MDIV}$ in Proposition 1 applies only to a specific implementation that forces both procedures to use exactly the same weighting matrices; it is not a statement that the two estimators coincide for all reasonable two-step implementations.