

TESTING RANDOM ASSIGNMENT TO PEER GROUPS

ONLINE APPENDIX

KOEN JOCHMANS*

UNIVERSITY OF CAMBRIDGE

July 15, 2020

Normalization. Our procedures accommodate fixed effects at the urn level. Moreover, all sample statistics involved are functions of observations from which the within-urn mean has been subtracted and are, thus, invariant to the fixed effects. Therefore, it is without loss of generality to set all fixed effects equal to zero. This implies that

$$\mathbb{E}_0(x_{g,i} x_{g,j}) = \begin{cases} \sigma_g^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \quad (\text{A.1})$$

This normalization simplifies the derivations below and is maintained throughout the Appendix.

Proof of Equation (1.2). Under the null of random assignment the bias in the normal equation is

$$\mathbb{E}_0 \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{x}_{g,i} \right) = \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} x_{g,i} \right) - \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_g \right).$$

We first calculate each of the expectations on the right-hand side and then collect results to arrive at (1.2).

*Address: University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. E-mail: kj345@cam.ac.uk.

For the first term on the right-hand side, observe that

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} x_{g,i} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j} x_{g,i}}{m_g(i)} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,j} x_{g,i} | \mathbf{A}_g)}{m_g(i)} \right) \\
&= 0,
\end{aligned}$$

where the first equality uses the definition of $\bar{x}_{g,[i]}$, the second equality iterates expectations, and the final equality follows from the fact that $\mathbb{E}_0(x_{g,i} x_{g,j} | \mathbf{A}_g) = \mathbb{E}_0(x_{g,i} x_{g,j}) = 0$ for all $i \neq j$ and the accounting identity $\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} = m_g(i)$.

For the second term on the right-hand side, we have

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_g \right) &= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j} x_{g,j'}}{m_g(i)} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,j} x_{g,j'})}{m_g(i)} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,j}^2)}{m_g(i)} \right) \\
&= \sigma_g^2,
\end{aligned}$$

using the same arguments as for the first term.

Taking differences of the expectations just calculated and summing over the r urns now shows that

$$\mathbb{E}_0 \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{x}_{g,i} \right) = - \sum_{g=1}^r \sigma_g^2,$$

which is Equation (1.2). □

Proof of Equation (1.3). The within-group estimator is

$$\hat{\rho} := \frac{\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{x}_{g,i}}{\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{x}_{g,[i]}}.$$

The expectation (under the null) of the numerator has already been calculated in (1.2) so it remains only to calculate the expectation of the denominator, $\mathbb{E}_0(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{\tilde{x}}_{g,[i]})$. Using the definition of $\bar{x}_{g,[i]}$, it can be written as

$$\sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \left(\sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j}}{m_g(i)} \right)^2 \right) - \sum_{g=1}^r \mathbb{E}_0 \left(\frac{1}{n_g} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j}}{m_g(i)} \right)^2 \right).$$

As in the proof of (1.2) we again start by calculating each of the expectations involved and then collect results.

To calculate the expectation in the first term we expand the square and iterate expectations to write

$$\begin{aligned} \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \left(\sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j}}{m_g(i)} \right)^2 \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i,j'} \mathbb{E}_0(x_{g,j} x_{g,j'})}{m_g(i)^2} \right) \\ &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}^2 \mathbb{E}_0(x_{g,j}^2)}{m_g(i)^2} \right) \\ &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} \right) \sigma_g^2, \end{aligned}$$

where we have first exploited (A.1) and then used $\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} = m_g(i)$ together with the fact that $(\mathbf{A}_g)_{i,j}^2 = (\mathbf{A}_g)_{i,j}$.

To calculate the expectation in the second term, we proceed in the same way. Doing so reveals that

$$\begin{aligned} \mathbb{E}_0 \left(\frac{1}{n_g} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j}}{m_g(i)} \right)^2 \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j'} \mathbb{E}_0(x_{g,j} x_{g,j'})}{n_g m_g(i) m_g(i')} \right) \\ &= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j} \mathbb{E}_0(x_{g,j}^2)}{m_g(i) m_g(i')} \right) \\ &= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{m_g(i \cap i')}{m_g(i) m_g(i')} \right) \sigma_g^2, \end{aligned}$$

where we recall that $\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j} = m_g(i \cap i')$.

Taking differences of the two expectations just calculated and summing over the r urns yields

$$\mathbb{E}_0 \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{\tilde{x}}_{g,[i]} \right) = \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} \right) \sigma_g^2.$$

Combined with (1.2) this yields (1.3) on letting $r \rightarrow \infty$.

The comparative statics in n_g and $m_g(i \cap j)$ are immediate. When peer groups do not overlap,

$$\begin{aligned}
\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} &= \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{m_g(i \cap i) + \sum_{j \neq i} m_g(i \cap j)}{m_g(i)^2} \\
&= \frac{1}{n_g} \sum_{i=1}^{n_g} \left(\frac{1}{m_g(i)} + \frac{\sum_{j \neq i} (\mathbf{A}_g)_{i,j} (m_g(i) - 1)}{m_g(i)^2} \right) \\
&= \frac{1}{n_g} \sum_{i=1}^{n_g} \left(\frac{1}{m_g(i)} + \frac{m_g(i) - 1}{m_g(i)} \right) \\
&= 1,
\end{aligned} \tag{A.2}$$

implying that, in this case the expectation of the denominator simplifies to

$$\sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - 1 \right) \sigma_g^2,$$

which is indeed decreasing in the $m_g(i)$.

When all $n_g = n$ and $m_g(i) = m$ the above equation becomes

$$\sum_{g=1}^r \left(\frac{n}{m} - 1 \right) \sigma_g^2 = \left(\frac{n - m}{m} \right) \sum_{g=1}^r \sigma_g^2.$$

Combined with Equation (1.2) this then yields

$$\text{plim}_{r \rightarrow \infty} = \frac{m}{n - m},$$

as claimed. This, after simplification of their formula (and noting that, in their notation $L = n$ and $K - 1 = m$), corresponds to the expressions in [Caeyers and Fafchamps \(2020\)](#), as claimed. \square

Proof of Theorem 1. By independence of the urns the variance of q_r^{HO} is

$$v_r^{\text{HO}} := \sum_{g=1}^r \mathbb{E}_0 \left(\left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right)^2 \right).$$

We need to prove that

$$\text{(i) } \frac{q_r^{\text{HO}} - b_r}{\sqrt{v_r^{\text{HO}}}} \xrightarrow{d} N(0, 1), \quad \text{and (ii) } \left(s_r^{\text{HO}} - \sqrt{v_r^{\text{HO}}} \right) \xrightarrow{p} 0$$

We handle each of these in turn. As subtracting b_r amounts to a mere recentering of q_r^{HO} to make it zero mean it suffices to set $b_r = 0$.

To show (i) we verify that the conditions of Lyapunov's central limit theorem are met. Here, Lyapunov's condition is

$$\lim_{r \rightarrow \infty} \frac{\sum_{g=1}^r \mathbb{E} \left(\left| \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g-1} \right) \right|^{2+\delta} \right)}{\left(\sum_{g=1}^r \mathbb{E} \left(\left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g-1} \right) \right)^2 \right) \right)^{\frac{2+\delta}{2}}} = 0,$$

for some $\delta > 0$. To do so it is useful to introduce

$$\lambda_{i,j}^g := \begin{cases} 1/(n_g - 1) & \text{if } i = j \\ (\mathbf{A}_g)_{i,j}/m_g(i) & \text{if } i \neq j \end{cases}$$

Then we can write

$$\sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g-1} \right) = \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \lambda_{i,j}^g \tilde{x}_{g,i} x_{g,j} = \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g-1} \right) x_{g,i} x_{g,j},$$

where we use that $n_g^{-1} \sum_{j=1}^{n_g} \lambda_{i,j} = n_g/(n_g-1)$. Let $\delta > 0$ be fixed. Then, by an application of Hölder's inequality,

$$\begin{aligned} \mathbb{E} \left(\left| \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g-1} \right) x_{g,i} x_{g,j} \right|^{2+\delta} \right) &\leq \left(\mathbb{E} \left(\max_{i,j} \left| \lambda_{i,j}^g - \frac{n_g}{n_g-1} \right|^{\frac{(2+\delta)(1+\theta)}{\theta}} \right) \right)^{\frac{\theta}{1+\theta}} \\ &\quad \times \left(\mathbb{E} \left(\left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} |x_{g,i} x_{g,j}| \right)^{(2+\delta)(1+\theta)} \right) \right)^{\frac{1}{1+\theta}}, \end{aligned}$$

for some $\theta > 0$. The first right-hand side term is finite for any choice of θ because the (re-centered) weights $\lambda_{i,j}^g - n_g/(n_g-1)$ are bounded. For second right-hand side term, letting $\varepsilon := 2\theta + \delta + \delta\theta > 0$,

$$\mathbb{E} \left(\left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} |x_{g,i} x_{g,j}| \right)^{2+\varepsilon} \right) \leq \mathbb{E} \left(\left(\sum_{i=1}^{n_g} |x_{g,i}| \right)^{4+2\varepsilon} \right) \leq \mathbb{E} \left(\left(\sum_{i=1}^{n_g} |x_{g,i}|^{4+2\varepsilon} \right) n_g \right) = O(1)$$

because $\max_g \max_i \mathbb{E}(|x_{g,i}|^{4+\delta}) < \infty$ by assumption. Consequently,

$$\sum_{g=1}^r \mathbb{E} \left(\left| \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g-1} \right) \right|^{2+\delta} \right) = O(r),$$

and Lyapunov's condition will follow if we can show that v_r^{HO} grows at the rate r . We may again write

$$\mathbb{E} \left(\left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right)^2 \right) = \mathbb{E} \left(\left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g - 1} \right) x_{g,i} x_{g,j} \right)^2 \right).$$

Expanding the square inside the expectation, iteration expectations, and recalling that we have normalized the data to have mean zero yields

$$\begin{aligned} \mathbb{E} \left(\left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g - 1} \right) x_{g,i} x_{g,j} \right)^2 \right) &= \mathbb{E} \left(\sum_{i=1}^{n_g} \left(\lambda_{i,i}^g - \frac{n_g}{n_g - 1} \right)^2 \mathbb{E}(x_{g,i}^4) \right) \\ &\quad + \mathbb{E} \left(\sum_{i=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g - 1} \right)^2 \sigma_g^4 \right) \\ &\quad + \mathbb{E} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,i}^g - \frac{n_g}{n_g - 1} \right) \left(\lambda_{j,j}^g - \frac{n_g}{n_g - 1} \right) \sigma_g^4 \right) \\ &\quad + \mathbb{E} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g - 1} \right) \left(\lambda_{j,i}^g - \frac{n_g}{n_g - 1} \right) \sigma_g^4 \right). \end{aligned}$$

Each of these terms is non-zero and bounded. Hence, v_r^{HO} , which is the sum of all these terms over the r urns, grows at the rate r and Lyapunov's condition holds.

To show (ii) it suffices to confirm that

$$\mathbb{E} \left(\left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(\lambda_{i,j}^g - \frac{n_g}{n_g - 1} \right) x_{g,i} x_{g,j} \right)^4 \right)$$

is bounded by a constant independent of g . Expanding the square and iterating expectations shows that this is indeed the case, as the urn sizes are fixed, the weights are uniformly bounded, and $\mathbb{E}(x_{g,i}^8) < \infty$. The consistency result is then a consequence of Kolmogorov's strong law. This completed the proof. \square

Proof of Equation (1.4). We need to compute the variance of

$$q_r^{\text{HO}} = \sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right)$$

under the null of random assignment. The calculations here allow for overlap between peer groups and allow for $\mathbb{E}_0(x_{g,i}^4) =: \gamma_g^4$ to depend on g (but not on i). Recall that q_r^{HO} has mean zero (under the null) by construction. By independence of the urns its variance is

$$v_r^{\text{HO}} := \sum_{g=1}^r \mathbb{E}_0 \left(\left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right)^2 \right),$$

which on expanding the sum is equal to

$$\sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \left(\tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right) \left(\tilde{x}_{g,i'} \left(\bar{x}_{g,[i']} + \frac{x_{g,i'}}{n_g - 1} \right) \right) \right). \quad (\text{A.3})$$

We first calculate the expectation of each of the cross-terms in this last expression and then combine the results.

The first term that needs to be calculated is

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \tilde{x}_{g,i} \tilde{x}_{g,i'} \bar{x}_{g,[i]} \bar{x}_{g,[i']} \right) = \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} (x_{g,i} - \bar{x}_g) (x_{g,i'} - \bar{x}_g) \right),$$

and expands as

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} x_{g,i} x_{g,i'} \right) - 2 \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} x_{g,i} \bar{x}_g \right) + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} \bar{x}_g^2 \right).$$

We now calculate each of the three expectations involved and collect results.

The first expectation is

$$\begin{aligned} \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} x_{g,i} x_{g,i'} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j'} \mathbb{E}_0(x_{g,i} x_{g,i'} x_{g,j} x_{g,j'})}{m_g(i) m_g(i')} \right) \\ &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i}^2 x_{g,j}^2)}{m_g(i) m_g(i)} \right) \\ &\quad + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{j,i} \mathbb{E}_0(x_{g,i}^2 x_{g,j}^2)}{m_g(i) m_g(j)} \right) \\ &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} \right) \sigma_g^4 + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i) m_g(j)} \right) \sigma_g^4, \end{aligned}$$

where we have used the definition of $\bar{x}_{g,[i]}$ together with the observation that only summands for which (i) $i = i'$ and $j = j'$ or (ii) $i = j'$ and $j = i'$ deliver a non-zero contribution. This follows

from the fact that the $x_{g,i}$ within each urn are independent under the null (conditional on their common urn fixed effect).

The second expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} x_{g,i} \bar{x}_g \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j'} \mathbb{E}_0(x_{g,i} x_{g,j} x_{g,j'} x_{g,k})}{m_g(i) m_g(i') n_g} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j} \mathbb{E}_0(x_{g,i}^2 x_{g,j}^2)}{m_g(i) m_g(i') n_g} \right) \\
&\quad + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',i} \mathbb{E}_0(x_{g,i}^2 x_{g,j}^2)}{m_g(i) m_g(i') n_g} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j}}{m_g(i) m_g(i') n_g} \right) \sigma_g^4 + \sigma_g^4 \\
&= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{m_g(i \cap i')}{m_g(i) m_g(i')} \right) \sigma_g^4 + \sigma_g^4,
\end{aligned}$$

which follows by the same arguments, now only retaining summands for which (i) $j' = j$ and $k = i$ or (ii) $j' = i$ and $k = j$.

The third expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_{g,[i']} \bar{x}_g^2 \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \sum_{k=1}^{n_g} \sum_{k'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j'} \mathbb{E}_0(x_{g,j} x_{g,j'} x_{g,k} x_{g,k'})}{m_g(i) m_g(i') n_g^2} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j} \mathbb{E}_0(x_{g,j}^2 x_{g,k}^2)}{m_g(i) m_g(i') n_g^2} \right) \\
&\quad + 2 \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j'} \mathbb{E}_0(x_{g,j}^2 x_{g,j'}^2)}{m_g(i) m_g(i') n_g^2} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{i',j}}{m_g(i) m_g(i')} \right) \sigma_g^4 \\
&\quad + 2 \mathbb{E}_0 \left(\frac{1}{n_g^2} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \sum_{j'=1}^{n_g} (\mathbf{A}_g)_{i',j'}}{m_g(i) m_g(i')} \right) \sigma_g^4 \\
&= \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{m_g(i \cap i')}{m_g(i) m_g(i')} \right) \sigma_g^4 + 2 \sigma_g^4,
\end{aligned}$$

where, now, three types of summands contribute; they are those for which (i) $j' = j$ and $k' = k$ or (ii) $k = j$ and $k' = j'$ or (iii) $k = j'$ and $k' = j$.

Putting everything together shows that $\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \tilde{x}_{g,i} \tilde{x}_{g,i'} \bar{x}_{g,[i]} \bar{x}_{g,[i']} \right)$ is equal to

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} + \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i) m_g(j)} \right) \sigma_g^4, \quad (\text{A.4})$$

which deals with the first term in (A.3).

The second term that needs to be calculated is

$$\begin{aligned} \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\tilde{x}_{g,i} \tilde{x}_{g,i'} \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{(x_{g,i} - \bar{x}_g)(x_{g,i'} - \bar{x}_g) \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) \\ &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} \bar{x}_{g,[i]} x_{g,i'}^2}{(n_g - 1)} \right) \\ &\quad - \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} \bar{x}_g \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) \\ &\quad - \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\bar{x}_g \bar{x}_{g,[i]} x_{g,i'}^2}{(n_g - 1)} \right) \\ &\quad + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\bar{x}_g^2 \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right). \end{aligned}$$

We evaluate each of these expectations in turn next.

The first expectation is

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} \bar{x}_{g,[i]} x_{g,i'}^2}{(n_g - 1)} \right) = \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i} x_{g,i'}^2 x_{g,j})}{m_g(i) (n_g - 1)} \right) = 0,$$

which follows from the fact that $(\mathbf{A}_g)_{i,i} = 0$ so that no combination of indices gives rise to a summand that has non-zero mean.

The second expectation is

$$\begin{aligned} \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} \bar{x}_g \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i} x_{g,i'} x_{g,j} x_{g,k})}{m_g(i) n_g (n_g - 1)} \right) \\ &= 2 \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i}^2 x_{g,j}^2)}{m_g(i) n_g (n_g - 1)} \right) \\ &= 2 \mathbb{E}_0 \left(\frac{1}{(n_g - 1)} \right) \sigma_g^4, \end{aligned}$$

where the summands that contribute are those for which (i) $i' = i$ and $k = j$ or (ii) $i' = j$ and $k = i$.

The third expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\bar{x}_g \bar{x}_{g,[i]} x_{g,i'}^2}{(n_g - 1)} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i'}^2 x_{g,j} x_{g,k})}{m_g(i) n_g (n_g - 1)} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i'}^2 x_{g,j}^2)}{m_g(i) n_g (n_g - 1)} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,i'} \mathbb{E}_0(x_{g,i'}^4) + \sum_{j \neq i'} (\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i'}^2 x_{g,j}^2)}{m_g(i) n_g (n_g - 1)} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{(n_g - 1)} \right) \gamma_g^4 + \mathbb{E}_0 \left(\frac{1}{n_g (n_g - 1)} \sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{m_g(i) - (\mathbf{A}_g)_{i,i'}}{m_g(i)} \right) \sigma_g^4 \\
&= \mathbb{E}_0 \left(\frac{1}{(n_g - 1)} \right) \gamma_g^4 + \sigma_g^4;
\end{aligned}$$

note that, here, only summands for which $k = j$ contribute, but their contribution depends on whether (i) $i' = j$ (which contributes a fourth-order moment) or whether (ii) $i' \neq j$ (which contributes a squared second moment).

The fourth expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\bar{x}_g^2 \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \sum_{k'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i'} x_{g,j} x_{g,k} x_{g,k'})}{m_g(i) n_g^2 (n_g - 1)} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{(\mathbf{A}_g)_{i,i'} \mathbb{E}_0(x_{g,i'}^4) + 3 \sum_{j \neq i'} (\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,i'}^2 x_{g,j}^2)}{m_g(i) n_g^2 (n_g - 1)} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{n_g (n_g - 1)} \right) \gamma_g^4 + 3 \mathbb{E}_0 \left(\frac{1}{n_g} \right) \sigma_g^4.
\end{aligned}$$

Here, again, a fourth-order term arises from the summands where $i' = j = k = k'$ while three different combinations of indices contribute terms involving σ_g^4 ; these are those where (i) $i' = j$ and $k = k'$ (but not both) or (ii) $i' = k$ and $k' = j$ (but not both) or (iii) $i' = k'$ and $k = j$ (but not both).

Combining results yields

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\tilde{x}_{g,i} \tilde{x}_{g,i'} \bar{x}_{g,[i]} x_{g,i'}}{(n_g - 1)} \right) = \mathbb{E}_0 \left(\frac{3}{n_g} - \frac{2}{(n_g - 1)} - 1 \right) \sigma_g^4 - \mathbb{E}_0 \left(\frac{1}{n_g} \right) \gamma_g^4, \quad (\text{A.5})$$

and gives (up to the factor 2) and expression for the second term in (A.3).

The third term and final term that needs to be calculated is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\tilde{x}_{g,i} \tilde{x}_{g,i'} x_{g,i} x_{g,i'}}{(n_g - 1)^2} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{(x_{g,i} - \bar{x}_g)(x_{g,i'} - \bar{x}_g) x_{g,i} x_{g,i'}}{(n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i}^2 x_{g,i'}^2}{(n_g - 1)^2} \right) \\
&\quad - 2 \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i}^2 x_{g,i'} \bar{x}_g}{(n_g - 1)^2} \right) \\
&\quad + \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} x_{g,i'} \bar{x}_g^2}{(n_g - 1)^2} \right).
\end{aligned}$$

This again requires calculating three distinct expectations, and we will again take on each in turn.

The first expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i}^2 x_{g,i'}^2}{(n_g - 1)^2} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{\mathbb{E}_0(x_{g,i}^4) + \sum_{i' \neq i} \mathbb{E}_0(x_{g,i}^2 x_{g,i'}^2)}{(n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\frac{n_g}{(n_g - 1)^2} \right) \gamma_g^4 + \mathbb{E}_0 \left(\frac{n_g}{(n_g - 1)} \right) \sigma_g^4.
\end{aligned}$$

The second expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i}^2 x_{g,i'} \bar{x}_g}{(n_g - 1)^2} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{k=1}^{n_g} \frac{\mathbb{E}_0(x_{g,i}^2 x_{g,i'} x_{g,k})}{n_g (n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{\mathbb{E}_0(x_{g,i}^4) + \sum_{i' \neq i} \mathbb{E}_0(x_{g,i}^2 x_{g,i'}^2)}{n_g (n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{(n_g - 1)^2} \right) \gamma_g^4 + \mathbb{E}_0 \left(\frac{1}{(n_g - 1)} \right) \sigma_g^4.
\end{aligned}$$

The third expectation is

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{x_{g,i} x_{g,i'} \bar{x}_g^2}{(n_g - 1)^2} \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \sum_{k=1}^{n_g} \sum_{k'=1}^{n_g} \frac{\mathbb{E}_0(x_{g,i} x_{g,i'} x_{g,k} x_{g,k'})}{n_g^2 (n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{\mathbb{E}_0(x_{g,i}^4) + 3 \sum_{i' \neq i} \mathbb{E}_0(x_{g,i}^2 x_{g,i'}^2)}{n_g^2 (n_g - 1)^2} \right) \\
&= \mathbb{E}_0 \left(\frac{1}{n_g (n_g - 1)^2} \right) \gamma_g^4 + 3 \mathbb{E}_0 \left(\frac{1}{n_g (n_g - 1)} \right) \sigma_g^4.
\end{aligned}$$

This then yields

$$\begin{aligned}
\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{i'=1}^{n_g} \frac{\tilde{x}_{g,i} \tilde{x}_{g,i'} x_{g,i} x_{g,i'}}{(n_g - 1)^2} \right) &= \mathbb{E}_0 \left(\frac{n_g}{(n_g - 1)} \left(1 - \frac{2}{n_g} + \frac{3}{n_g^2} \right) \right) \sigma_g^4 \\
&\quad + \mathbb{E}_0 \left(\frac{n_g}{(n_g - 1)^2} \left(1 - \frac{2}{n_g} + \frac{1}{n_g^2} \right) \right) \gamma_g^4
\end{aligned} \tag{A.6}$$

for the third and final term in (A.3).

Now, collecting results by combining (A.3) with (A.4)–(A.6) yields

$$\begin{aligned} v_r^{\text{HO}} &= \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} + \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i) m_g(j)} - \frac{(n_g - 1)^2 + 2}{n_g(n_g - 1)} \right) \sigma_g^4 \\ &\quad - \sum_{g=1}^r \mathbb{E}_0 \left(\frac{1}{n_g} \right) \gamma_g^4. \end{aligned}$$

When $\gamma_g^4 = 3\sigma_g^4$ and peer groups do not overlap we obtain

$$v_r^{\text{HO}} = \sum_{g=1}^r \mathbb{E}_0 \left(2 \sum_{i=1}^{n_g} \frac{1}{m_g(i)} - 1 - \frac{n_g + 1}{n_g - 1} \right) \sigma_g^4$$

by (A.2). This is the variance formula stated in Equation (1.4) in the main text. \square

Proof of Equation (1.5). Suppose that

$$x_{g,i} = \rho \bar{x}_{g,[i]} + \varepsilon_{g,i}, \quad \varepsilon_{g,i} \sim \text{independent}(\alpha_g, \sigma_g^2),$$

for some $|\rho| < 1$. Maintaining our normalization, we set $\alpha_g = 0$ for all urns. Recall that, by definition,

$$\bar{x}_{g,[i]} = \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} x_{g,j}.$$

Consequently, collecting variables for urn g in $\mathbf{x}_g := (x_{g,1}, \dots, x_{g,n_g})'$ and $\boldsymbol{\varepsilon}_g := (\varepsilon_{g,1}, \dots, \varepsilon_{g,n_g})'$ and letting

$$\mathbf{G}_g := \mathbf{D}_g^{-1} \mathbf{A}_g, \quad \mathbf{D}_g := \text{diag}(m_g(1), \dots, m_g(n_g)),$$

we have the linear system

$$\mathbf{x}_g = \rho \mathbf{G}_g \mathbf{x}_g + \boldsymbol{\varepsilon}_g,$$

which has reduced form

$$\mathbf{x}_g = (\mathbf{I}_{n_g} - \rho \mathbf{G}_g)^{-1} \boldsymbol{\varepsilon}_g = \sum_{o=0}^{\infty} (\rho^o \mathbf{G}_g^o) \boldsymbol{\varepsilon}_g.$$

Moreover,

$$x_{g,i} = \varepsilon_{g,i} + \rho \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \varepsilon_{g,j} + \rho^2 \frac{1}{m_g(i)} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{j,k} \varepsilon_{g,k} + \dots \quad (\text{A.7})$$

We then obtain

$$\begin{aligned}
\mathbb{E}_{\frac{\rho}{\sqrt{r}}} (x_{g,i} x_{g,i'} | \mathbf{A}_g) &= \mathbb{E}_{\frac{\rho}{\sqrt{r}}} (\varepsilon_{g,i} \varepsilon_{g,i'} | \mathbf{A}_g) \\
&+ \mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\varepsilon_{g,i} \left(\frac{\rho}{\sqrt{r}} \frac{1}{m_g(i')} \sum_{j'=1}^{n_g} (\mathbf{A}_g)_{i',j'} \varepsilon_{g,j'} \right) \middle| \mathbf{A}_g \right) \\
&+ \mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\varepsilon_{g,i'} \left(\frac{\rho}{\sqrt{r}} \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \varepsilon_{g,j} \right) \middle| \mathbf{A}_g \right) \\
&+ o(r^{-1/2}).
\end{aligned}$$

The expectations on the right-hand side can be worked out. First we have

$$\mathbb{E}_{\frac{\rho}{\sqrt{r}}} (\varepsilon_{g,i} \varepsilon_{g,i'} | \mathbf{A}_g) = \begin{cases} \sigma_g^2 & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases},$$

which mimics (A.1). Next we calculate

$$\begin{aligned}
\mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\varepsilon_{g,i} \left(\frac{\rho}{\sqrt{r}} \frac{1}{m_g(i')} \sum_{j'=1}^{n_g} (\mathbf{A}_g)_{i',j'} \varepsilon_{g,j'} \right) \middle| \mathbf{A}_g \right) &= \frac{\rho}{\sqrt{r}} \sum_{j'=1}^{n_g} \frac{(\mathbf{A}_g)_{i',j'} \mathbb{E}_{\frac{\rho}{\sqrt{r}}} (\varepsilon_{g,i} \varepsilon_{g,j'} | \mathbf{A}_g)}{m_g(i')} \\
&= \frac{\rho}{\sqrt{r}} \frac{(\mathbf{A}_g)_{i',i} \mathbb{E}_{\frac{\rho}{\sqrt{r}}} (\varepsilon_{g,i}^2 | \mathbf{A}_g)}{m_g(i')} \\
&= \frac{\rho}{\sqrt{r}} \frac{(\mathbf{A}_g)_{i',i}}{m_g(i')} \sigma_g^2
\end{aligned}$$

and, finally,

$$\mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\varepsilon_{g,i'} \left(\frac{\rho}{\sqrt{r}} \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \varepsilon_{g,j} \right) \middle| \mathbf{A}_g \right) = \frac{\rho}{\sqrt{r}} \frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)} \sigma_g^2,$$

follows in the same way. Putting everything together then reveals that, up to terms that are $o(r^{-1/2})$,

$$\mathbb{E}_{\frac{\rho}{\sqrt{r}}} (x_{g,i} x_{g,i'} | \mathbf{A}_g) = \begin{cases} \sigma_g^2 & \text{if } i = i' \\ \frac{\rho}{\sqrt{r}} \left(\frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)} + \frac{(\mathbf{A}_g)_{i',i}}{m_g(i')} \right) \sigma_g^2 & \text{if } i \neq i' \end{cases}. \quad (\text{A.8})$$

This expression is key in deriving the asymptotic bias, which we turn to next.

The bias is

$$\mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right) = \sum_{g=1}^r \mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \bar{x}_{g,[i]} \right) + \sum_{g=1}^r \mathbb{E}_{\frac{\rho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{\tilde{x}_{g,i} x_{g,i}}{n_g - 1} \right).$$

We calculate each of the expectations in turn.

First, up to terms that are $o(r^{-1/2})$,

$$\begin{aligned}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\tilde{x}_{g,i}\bar{x}_{g,[i]}\right) &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}x_{g,i}\bar{x}_{g,[i]}\right) - \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\bar{x}_g\bar{x}_{g,[i]}\right) \\ &= \left(\sum_{i=1}^{n_g}\frac{1}{m_g(i)} - \frac{2}{n_g}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\frac{m_g(i\cap j)}{m_g(i)m_g(j)} + \sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\frac{(\mathbf{A}_g)_{i,j}}{m_g(i)m_g(j)}\right)\sigma_g^2\frac{\varrho}{\sqrt{r}} - \sigma_g^2,\end{aligned}$$

because,

$$\begin{aligned}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}x_{g,i}\bar{x}_{g,[i]}\right) &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}(x_{g,i}x_{g,i'}|\mathbf{A}_g)}{m_g(i)}\right) \\ &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)^2} + \sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)m_g(i')}\right)\sigma_g^2\frac{\varrho}{\sqrt{r}} + o(r^{-1/2}) \\ &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\frac{1}{m_g(i)} + \sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)m_g(i')}\right)\sigma_g^2\frac{\varrho}{\sqrt{r}} + o(r^{-1/2}),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\bar{x}_g\bar{x}_{g,[i]}\right) &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}\sum_{k=1}^{n_g}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}(x_{g,k}x_{g,i'}|\mathbf{A}_g)}{m_g(i)n_g}\right) \\ &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{(\mathbf{A}_g)_{i,i'}\left(\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}(x_{g,i'}^2|\mathbf{A}_g) + \sum_{k\neq i'}\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}(x_{g,k}x_{g,i'}|\mathbf{A}_g)\right)}{m_g(i)n_g}\right) \\ &= \sigma_g^2 + \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\frac{2}{n_g}\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{\sum_{k\neq i'}(\mathbf{A}_g)_{i,i'}(\mathbf{A}_g)_{i',k}}{m_g(i)m_g(i')}\right)\sigma_g^2\frac{\varrho}{\sqrt{r}} + o(r^{-1/2}) \\ &= \sigma_g^2 + \mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(\frac{2}{n_g}\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{m_g(i\cap i')}{m_g(i)m_g(i')}\right)\sigma_g^2\frac{\varrho}{\sqrt{r}} + o(r^{-1/2}),\end{aligned}$$

where we have made extensive use of (A.8).

Second, again up to terms that are $o(r^{-1/2})$,

$$\begin{aligned}
\mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{\tilde{x}_{g,i} x_{g,i}}{n_g - 1} \right) &= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{x_{g,i}^2}{n_g - 1} \right) - \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{x_{g,i} \bar{x}_g}{n_g - 1} \right) \\
&= \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\frac{n_g}{n_g - 1} \right) \sigma_g^2 \\
&\quad - \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\frac{1}{n_g - 1} \right) \sigma_g^2 \\
&\quad - \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\frac{1}{n_g(n_g - 1)} \sum_{i=1}^{n_g} \sum_{i' \neq i} \left(\frac{(\mathbf{A}_g)_{i,i'}}{m_g(i)} + \frac{(\mathbf{A}_g)_{i,i'}}{m_g(i')} \right) \right) \frac{\varrho}{\sqrt{r}} \sigma_g^2 \\
&= \sigma_g^2 - \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\frac{2}{n_g - 1} \right) \frac{\varrho}{\sqrt{r}} \sigma_g^2,
\end{aligned}$$

again exploiting (A.8).

Hence, up to terms that are $o(\sqrt{r})$, the bias in the normal equation is

$$\frac{\varrho}{\sqrt{r}} \sum_{g=1}^p \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{2}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} + \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i) m_g(j)} - \frac{2}{n_g - 1} \right) \sigma_g^2.$$

When peer groups do not overlap we can exploit the fact that

$$\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} = 1, \quad \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i) m_g(j)} = \sum_{i=1}^{n_g} \frac{1}{m_g(i)},$$

to reduce the bias expression to

$$2 \frac{\varrho}{\sqrt{r}} \sum_{g=1}^p \mathbb{E}_{\frac{\varrho}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - 1 - \frac{1}{n_g - 1} \right) \sigma_g^2 + o(\sqrt{r}).$$

Because the adjacency matrices do not vary with the alternative the subscript on the expectations operator in this expression can be dropped. This delivers Equation (1.5). \square

Proof of Equation (1.6). To see that endogenous effects and contextual effects are locally asymptotically equivalent note that the latter violation of the null is of the form

$$x_{g,i} = \varepsilon_{g,i} + \frac{\theta}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \varepsilon_{g,j}, \quad \varepsilon_{g,i} \sim \text{independent } (\alpha_g, \sigma_g^2),$$

for drifting sequences $\theta = \vartheta/\sqrt{r}$. Our normalization again sets $\alpha_g = 0$ for all urns. Clearly, on setting $\varrho = \vartheta$, this data generating process co-incides with the reduced form in (A.7), up to first-order. Consequently, it is immediate that $\mathbb{E}_{\frac{\vartheta}{\sqrt{r}}} (x_{g,i} x_{g,i'} | \mathbf{A}_g)$ satisfies the expansions in (A.8). This, then, implies that the asymptotic bias induced by contextual effects, too, is identical. \square

Proof of Equation (1.7). Consider non-overlapping peer groups. Peers are subject to a common random effect drawn from a distribution with zero mean and variance σ_η^2 . Consequently,

$$\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,i'} | \mathbf{A}_g) = \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,i'} | (\mathbf{A}_g)_{i,i'}) .$$

For drifting sequences of the form $\sigma_\eta^2 = \varsigma^2 / \sqrt{r}$ this implies that

$$\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,i'} | (\mathbf{A}_g)_{i,i'} = 0) \begin{cases} \frac{\varsigma^2}{\sqrt{r}} + \sigma_g^2 & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases} , \quad \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,i'} | (\mathbf{A}_g)_{i,j} = 1) = \frac{\varsigma^2}{\sqrt{r}},$$

where, recall, the urn fixed effects have been normalized to zero.

We are now ready to tackle the calculation of

$$\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{g=1}^p \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \right) .$$

We again proceed by first calculating all expectations involved separately and then combining all results.

First, we have

$$\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \tilde{x}_{g,i} \bar{x}_{g,[i]} \right) = \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} x_{g,i} \bar{x}_{g,[i]} \right) - \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \bar{x}_g \bar{x}_{g,[i]} \right) = \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (n_g - 2) \frac{\varsigma^2}{\sqrt{r}} - \sigma^2 .$$

This follows from the observations that

$$\begin{aligned} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} x_{g,i} \bar{x}_{g,[i]} \right) &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j \neq i} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,j} | \mathbf{A}_g)}{m_g(i)} \right) \\ &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j \neq i} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,i} x_{g,j} | (\mathbf{A}_g)_{i,j} = 1)}{m_g(i)} \right) \\ &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j \neq i} \frac{(\mathbf{A}_g)_{i,j}}{m_g(i)} \right) \frac{\varsigma^2}{\sqrt{r}} \\ &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (n_g) \frac{\varsigma^2}{\sqrt{r}}, \end{aligned}$$

and that

$$\begin{aligned}
\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \bar{x}_g \bar{x}_{g,[i]} \right) &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,j} x_{g,k} | \mathbf{A}_g)}{m_g(i) n_g} \right) \\
&= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,j}^2 | (\mathbf{A}_g)_{j,j} = 0)}{m_g(i) n_g} \right) \\
&\quad + \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k \neq j} \frac{(\mathbf{A}_g)_{i,j} (\mathbf{A}_g)_{j,k} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,j} x_{g,k} | (\mathbf{A}_g)_{j,k} = 1)}{m_g(i) n_g} \right) \\
&= \left(\frac{\varsigma^2}{\sqrt{r}} + \sigma_g^2 \right) + \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i)} \right) \frac{\varsigma^2}{\sqrt{r}} \\
&= 2 \frac{\varsigma^2}{\sqrt{r}} + \sigma_g^2;
\end{aligned}$$

here, the last equality exploits the fact that peer groups do not overlap by appealing to (A.2).

Second, we have

$$\begin{aligned}
\mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{\tilde{x}_{g,i} x_{g,i}}{(n_g - 1)} \right) &= \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{x_{g,i}^2}{(n_g - 1)} \right) - \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{k=1}^{n_g} \frac{x_{g,k} x_{g,i}}{n_g (n_g - 1)} \right) \\
&= \frac{n_g}{(n_g - 1)} \left(\frac{\varsigma^2}{\sqrt{r}} + \sigma^2 \right) \\
&\quad - \frac{1}{(n_g - 1)} \left(\frac{\varsigma^2}{\sqrt{r}} + \sigma^2 \right) \\
&\quad - \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \sum_{k \neq i} \frac{(\mathbf{A}_g)_{i,k} \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} (x_{g,k} x_{g,i} | (\mathbf{A}_g)_{i,k} = 1)}{n_g (n_g - 1)} \right) \\
&= \left(\frac{\varsigma^2}{\sqrt{r}} + \sigma^2 \right) \\
&\quad - \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\sum_{i=1}^{n_g} \frac{m_g(i)}{n_g (n_g - 1)} \right) \frac{\varsigma^2}{\sqrt{r}}.
\end{aligned}$$

Combining results then shows that the bias equals

$$\sum_{g=1}^r \mathbb{E}_{\frac{\varsigma^2}{\sqrt{r}}} \left(\frac{n_g (n_g - 1)^2 - \sum_{i=1}^{n_g} m_g(i)}{n_g (n_g - 1)} \right) \frac{\varsigma^2}{\sqrt{r}},$$

as claimed. \square

Proof of Equation (2.8). By the Frisch-Waugh-Lovell theorem the estimated slope on $\bar{x}_{g,[i]}$ from a within-group regression of $x_{g,i}$ on $\bar{x}_{g,[i]}$ and $x_{g,i}/(n_g - 1)$ can be written as the ratio of the sum

$$\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \left(\tilde{x}_{g,i} - \hat{\delta}_1 \frac{\tilde{x}_{g,i}}{n_g - 1} \right), \quad \hat{\delta}_1 := \frac{\sum_{g=1}^r \frac{1}{(n_g - 1)} \sum_{i=1}^{n_g} x_{g,i} \tilde{x}_{g,i}}{\sum_{g=1}^r \frac{1}{(n_g - 1)^2} \sum_{i=1}^{n_g} x_{g,i} \tilde{x}_{g,i}},$$

to

$$\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \left(\tilde{\bar{x}}_{g,[i]} - \hat{\delta}_2 \frac{\tilde{x}_{g,i}}{n_g - 1} \right), \quad \hat{\delta}_2 := \frac{\sum_{g=1}^r \frac{1}{(n_g - 1)} \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{x}_{g,i}}{\sum_{g=1}^r \frac{1}{(n_g - 1)^2} \sum_{i=1}^{n_g} x_{g,i} \tilde{x}_{g,i}}.$$

Although our main interest lies in the numerator, it is easily established that for the denominator we have that, under the null,

$$\text{plim}_{r \rightarrow \infty} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \left(\tilde{x}_{g,i} - \hat{\delta}_2 \frac{\tilde{x}_{g,i}}{n_g - 1} \right) \right)$$

is equal to

$$\lim_{r \rightarrow \infty} \frac{1}{r} \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{m_g(i \cap j)}{m_g(i) m_g(j)} - \frac{1}{n_g - 1} \right) \sigma_g^2.$$

Here we have used that

$$\text{plim}_{p \rightarrow \infty} \hat{\delta}_2 = - \frac{\lim_{p \rightarrow \infty} \frac{1}{p} \sum_{g=1}^p \mathbb{E}_0 \left(\frac{\sigma_g^2}{n_g - 1} \right)}{\lim_{p \rightarrow \infty} \frac{1}{p} \sum_{g=1}^p \mathbb{E}_0 \left(\frac{\sigma_g^2}{n_g - 1} \right)} = -1,$$

under random assignment. Note that this probability limit is strictly smaller than its counterpart in the denominator of (1.3).

We now turn to the behavior of the numerator under the null. It is easy to show that

$$\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \left(\tilde{x}_{g,i} - \hat{\delta}_1 \frac{\tilde{x}_{g,i}}{n_g - 1} \right) = \sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) \left(1 - \frac{\delta}{n_g - 1} \right) + o_p(\sqrt{r}),$$

where

$$\delta := \text{plim}_{r \rightarrow \infty} \hat{\delta}_1 = \frac{\lim_{r \rightarrow \infty} \frac{1}{r} \sum_{g=1}^r \sigma_g^2}{\lim_{r \rightarrow \infty} \frac{1}{r} \sum_{g=1}^r \mathbb{E}_0 \left(\frac{\sigma_g^2}{n_g - 1} \right)}.$$

This is Equation (2.8) in the main text. The leading term is equal to our re-centered normal equation, up to the factor $1 - \delta/(n_g - 1)$ in the summand. This result shows that (in large samples) the multiple-regression strategy of [Guryan, Kroft and Notowidigdo \(2009\)](#) implements our bias correction to the (numerator of) the simple within-group estimator by leveraging on variation in urn size. \square

Proof of (2.9). A short calculation shows that, in (2.8), the large-size urns get assigned the weight

$$w_n := 1 - \frac{\delta}{\bar{n}_2 - 1} = \frac{(\bar{n}_2 - \bar{n}_1)(1 - p_n)}{(\bar{n}_1 - 1) + (\bar{n}_2 - \bar{n}_1)(1 - p_n)} \in (0, 1)$$

while the small-size urns get assigned the weight

$$-\frac{p_n}{1 - p_n} \omega_n < 0.$$

Then the variance of the modified score equation is

$$v(\bar{n}_1) \left(\frac{p_n}{1 - p_n} \omega_n \right)^2 (1 - p_n) + v(\bar{n}_2) \omega_n^2 p_n = \frac{p_n}{1 - p_n} \omega_n^2 (v(\bar{n}_1) p_n + v(\bar{n}_2) (1 - p_n))$$

and, in the same way, its bias is

$$p_n \omega_n (b(\bar{n}_2) - b(\bar{n}_1)).$$

Hence, combining results, the non-centrality parameter in the limit distribution takes the form

$$\mu^* = \sqrt{p_n(1 - p_n)} \frac{b(\bar{n}_2) - b(\bar{n}_1)}{\sqrt{v(\bar{n}_1) p_n + v(\bar{n}_2) (1 - p_n)}} = \frac{b(\bar{n}_2) - b(\bar{n}_1)}{\sqrt{\frac{v(\bar{n}_1)}{1 - p_n} + \frac{v(\bar{n}_2)}{p_n}}},$$

as stated in the main text. Note that the denominator in the expression after the first equality is not equal to the standard deviation s_r^{HO} . Moreover, because $p_n \in (0, 1)$,

$$v(\bar{n}_2) p_n < v(\bar{n}_2) < \frac{v(\bar{n}_2)}{p_n}, \quad v(\bar{n}_1) (1 - p_n) < v(\bar{n}_1) < \frac{v(\bar{n}_1)}{1 - p_n},$$

so that

$$v_r^{\text{HO}} = v(\bar{n}_1) (1 - p_n) + v(\bar{n}_2) p_n < \frac{v(\bar{n}_1)}{1 - p_n} + \frac{v(\bar{n}_2)}{p_n},$$

implying that the denominator of μ^* is always larger than the variance of q_r^{HO} . The numerators of μ and μ^* cannot be ranked at this level of generality. \square

Proof of Equation (3.10). The proof mimics the proof of Equation (1.2). The only difference arises in the second term where, now,

$$\mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} \bar{x}_g \right) = \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \mathbb{E}_0(x_{g,j}^2)}{m_g(i)} \right) = \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \sigma_{g,j}^2 \right),$$

from which the result follows. \square

Proof of Equation (3.11). For urn g write the $n_g \times n_g$ matrix that transforms observations into deviations from the within-urn mean as

$$(\mathbf{M}_g)_{i,j} := \begin{cases} 1 - \frac{1}{n_g} & \text{if } i = j \\ -\frac{1}{n_g} & \text{if } i \neq j \end{cases}.$$

Then $\tilde{x}_{g,i} = \sum_{j=1}^{n_g} (\mathbf{M}_g)_{i,j} x_{g,j}$ and so

$$\mathbb{E}_0(x_{g,i} \tilde{x}_{g,i}) = \mathbb{E}_0(\tilde{x}_{g,i}^2) = \mathbb{E}_0 \left(\sum_{j=1}^{n_g} \sum_{j'=1}^{n_g} (\mathbf{M}_g)_{i,j} (\mathbf{M}_g)_{i,j'} \mathbb{E}_0(x_{g,j} x_{g,j'}) \right) = \mathbb{E}_0 \left(\sum_{j=1}^{n_g} (\mathbf{M}_g)_{i,j}^2 \sigma_{g,j}^2 \right).$$

Let $\mathbf{x}_g := (x_{g,1}, \dots, x_{g,n_g})'$, $\boldsymbol{\sigma}_g^2 := (\sigma_{g,1}^2, \dots, \sigma_{g,n_g}^2)'$, and let $*$ denote the elementwise product between two matrices of conformable dimension. Then the above equation can be written in vector form as $\mathbb{E}_0(\mathbf{x}_g * \tilde{\mathbf{x}}_g) = \mathbb{E}_0((\mathbf{M}_g * \mathbf{M}_g) \boldsymbol{\sigma}_g^2)$. Consequently, $\mathbb{E}_0((\mathbf{M}_g * \mathbf{M}_g)^{-1} (\mathbf{x}_g * \tilde{\mathbf{x}}_g)) = \boldsymbol{\sigma}_g^2$ and

$$\sum_{j=1}^{n_g} ((\mathbf{M}_g * \mathbf{M}_g)^{-1})_{i,j} x_{g,j} \tilde{x}_{g,j}$$

is an unbiased estimator of $\sigma_{g,i}^2$ provided that the matrix $(\mathbf{M}_g * \mathbf{M}_g)$ is invertible. A calculation shows that the inverse is well-defined when $n_g > 2$ and that

$$((\mathbf{M}_g * \mathbf{M}_g)^{-1})_{i,j} =: (\mathbf{I}_g)_{i,j} = \begin{cases} \frac{n_g}{n_g-2} \left(1 - \frac{1}{n_g(n_g-1)} \right) & \text{if } i = j \\ -\frac{n_g}{n_g-2} \frac{1}{n_g(n_g-1)} & \text{if } i \neq j \end{cases}.$$

An unbiased plug-in estimator of the bias in (3.10) thus is

$$-\sum_{g=1}^r \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} \sum_{j'=1}^{n_g} (\mathbf{I}_g)_{j,j'} x_{g,j'} \tilde{x}_{g,j'} = -\sum_{g=1}^r \sum_{i=1}^{n_g} \sum_{j'=1}^{n_g} \omega_{g,j'} x_{g,j'} \tilde{x}_{g,j'},$$

on verifying that, indeed,

$$\omega_{g,j'} = \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{\sum_{j=1}^{n_g} (\mathbf{A}_g)_{i,j} (\mathbf{I}_g)_{j,j'}}{m_g(i)}.$$

This then immediately also implies the unbiasedness of q_r^{HC} as defined in (3.11). The proof is complete. \square

Proof of Theorem 2. The proof is the same as the proof of Theorem 1. It suffices to redefine

$$\lambda_{i,j}^g := \begin{cases} \omega_{g,i} & \text{if } i = j \\ (\mathbf{A}_g)_{i,j}/m_g(i) & \text{if } i \neq j \end{cases},$$

and note that these weights are again uniformly bounded. \square

Proof of Theorem 3. By the Frisch-Waugh-Lovell theorem the estimated slope on $\bar{x}_{g,[i]}$ from a within-group regression of $x_{g,i}$ on $\bar{x}_{g,[i]}$ and covariate vector $\mathbf{w}_{g,i}$ equals

$$\frac{\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \dot{x}_{g,i}}{\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \dot{\bar{x}}_{g,[i]}},$$

where $\dot{x}_{g,i} = \tilde{x}_{g,i} - \tilde{\mathbf{w}}'_{g,i} \hat{\beta}_1$, for

$$\hat{\beta}_1 := \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} x_{g,i} \right),$$

and $\dot{\bar{x}}_{g,[i]} = \tilde{\bar{x}}_{g,[i]} - \tilde{\mathbf{w}}'_{g,i} \hat{\beta}_2$, with

$$\hat{\beta}_2 := \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \bar{x}_{g,[i]} \right).$$

are residuals from auxiliary within-group regressions. We are again only concerned with the numerator.

Let $\ddot{x}_{g,i} := \tilde{x}_{g,i} - \tilde{\mathbf{w}}'_{g,i} \beta_1$ and $\ddot{\bar{x}}_{g,[i]} := \tilde{\bar{x}}_{g,[i]} - \tilde{\mathbf{w}}'_{g,i} \beta_2$ be the deviations of $\tilde{x}_{g,i}$ and $\tilde{\bar{x}}_{g,[i]}$ from their respective population linear projections on urn-specific intercepts and the covariate vector $\mathbf{w}_{g,i}$. Then

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \ddot{x}_{g,i} \right), \\ \hat{\beta}_2 &= \beta_2 + \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \ddot{\bar{x}}_{g,[i]} \right), \end{aligned}$$

are the conventional sample-error representations that follow from re-arrangement. Hence, some

elementary re-arrangement gives

$$\begin{aligned}
\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \dot{x}_{g,i} &= \sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} (\ddot{x}_{g,i} - \tilde{\mathbf{w}}'_{g,i}(\hat{\beta}_1 - \beta_1)) \\
&= \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \ddot{x}_{g,i} \right) \\
&\quad - \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \tilde{\mathbf{w}}'_{g,i} \right) \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}'_{g,i} \ddot{x}_{g,i} \right) \\
&= \sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} (\tilde{x}_{g,[i]} - \tilde{\mathbf{w}}'_{g,i} \hat{\beta}_2) \\
&= \sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} (\ddot{x}_{g,[i]} - \tilde{\mathbf{w}}'_{g,i} (\hat{\beta}_2 - \beta_2)) \\
&= \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} \ddot{x}_{g,[i]} \right) \\
&\quad - \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} \tilde{\mathbf{w}}'_{g,i} \right) \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right)^{-1} \left(\sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}'_{g,i} \ddot{x}_{g,[i]} \right) \\
&= \sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} \ddot{x}_{g,[i]} + o_p(\sqrt{r}).
\end{aligned}$$

Here, the last equality is a consequence of

$$\frac{1}{r} \sum_{g=1}^r \sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \xrightarrow{p} \lim_{r \rightarrow \infty} \frac{1}{r} \sum_{g=1}^r \mathbb{E} \left(\sum_{i=1}^{n_g} \tilde{\mathbf{w}}_{g,i} \mathbf{w}'_{g,i} \right),$$

where the probability limit is a well-defined and invertible matrix, together with the observation that

$$\frac{1}{r} \sum_{g=1}^r \sum_{i=1}^{n_g} (\tilde{\mathbf{w}}_{g,i} \ddot{x}_{g,i}) + O_p(r^{-1/2}), \quad \frac{1}{r} \sum_{g=1}^r \sum_{i=1}^{n_g} (\tilde{\mathbf{w}}_{g,i} \ddot{x}_{g,[i]}) + O_p(r^{-1/2}),$$

where we have used $\mathbb{E}(\tilde{\mathbf{w}}_{g,i} \ddot{x}_{g,i}) = 0$ and $\mathbb{E}(\tilde{\mathbf{w}}_{g,i} \ddot{x}_{g,[i]}) = 0$, which hold by basic properties of linear projection.

With the representation

$$\sum_{g=1}^r \sum_{i=1}^{n_g} \bar{x}_{g,[i]} \dot{x}_{g,i} = \sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} \bar{x}_{g,[i]} + o_p(\sqrt{r}) \tag{A.9}$$

we turn to calculating the expectation of the sum on the right-hand side, under the null. It is useful to observe that, in the presence of covariates, the null of (conditional) random assignment implies that

$$\check{x}_{g,i} | \mathbf{w}_{g,1}, \dots, \mathbf{w}_{g,n_g} \sim \text{independent } (\alpha_g, \sigma_g^2), \quad \check{x}_{g,i} := x_{g,i} - \mathbf{w}'_{g,i} \boldsymbol{\beta}.$$

Further, we remark that, under the null, the projection coefficient $\hat{\beta}_1$ is a consistent estimator of $\boldsymbol{\beta}$. Indeed, it has the interpretation of an estimator of $\boldsymbol{\beta}$ that enforces the null. We therefore have

$$\ddot{x}_{g,i} = (x_{g,i} - \bar{x}_g) - (\mathbf{w}_{g,i} - \bar{\mathbf{w}}_g)' \boldsymbol{\beta} = (\check{x}_{g,i} - \bar{\check{x}}_g).$$

We may now proceed essentially as in the proof of (1.2) to calculate the bias. We find, again using our normalization that $\alpha_g = 0$, that

$$\begin{aligned} \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} \ddot{x}_{g,i} \right) &= \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} (\check{x}_{g,i} - \bar{\check{x}}_g) \right) \\ &= \sum_{g=1}^r \mathbb{E}_0 \left(\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} (\check{x}_{g,i} - \bar{\check{x}}_g) x_{g,j}}{m_g(i)} \right) \\ &= - \sum_{g=1}^r \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} x_{g,j} \check{x}_{g,k}}{m_g(i)} \right) \\ &= - \sum_{g=1}^r \mathbb{E}_0 \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} \frac{(\mathbf{A}_g)_{i,j} \check{x}_{g,j} \check{x}_{g,k}}{m_g(i)} \right) \\ &= - \sum_{g=1}^r \sigma_g^2. \end{aligned}$$

The transitions in this display have made use of

$$\mathbb{E}_0 (\check{x}_{g,i} x_{g,j}) = \mathbb{E}_0 (\check{x}_{g,i} \check{x}_{g,j}) + \mathbb{E}_0 (\check{x}_{g,i} \tilde{\mathbf{w}}'_{g,j}) \boldsymbol{\beta},$$

where

$$\mathbb{E}_0 (\check{x}_{g,i} \check{x}_{g,j}) \begin{cases} \sigma_g^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

and $\mathbb{E}_0 (\check{x}_{g,i} \tilde{\mathbf{w}}'_{g,j}) = \mathbb{E}_0 (\mathbb{E}_0 (\check{x}_{g,i} | \mathbf{w}_{g,1}, \dots, \mathbf{w}_{g,n_g}) \tilde{\mathbf{w}}'_{g,j}) = 0$.

The urn-specific variances are estimated by

$$\frac{1}{n_g - 1} \sum_{i=1}^{n_g} \dot{x}_{g,i} x_{g,i},$$

and so the corrected covariance estimator is

$$\sum_{g=1}^r \left(\sum_{i=1}^{n_g} \bar{x}_{g,[i]} \dot{x}_{g,i} + \frac{1}{n_g - 1} \sum_{i=1}^{n_g} \dot{x}_{g,i} x_{g,i} \right) = \sum_{g=1}^r \sum_{i=1}^{n_g} \dot{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right),$$

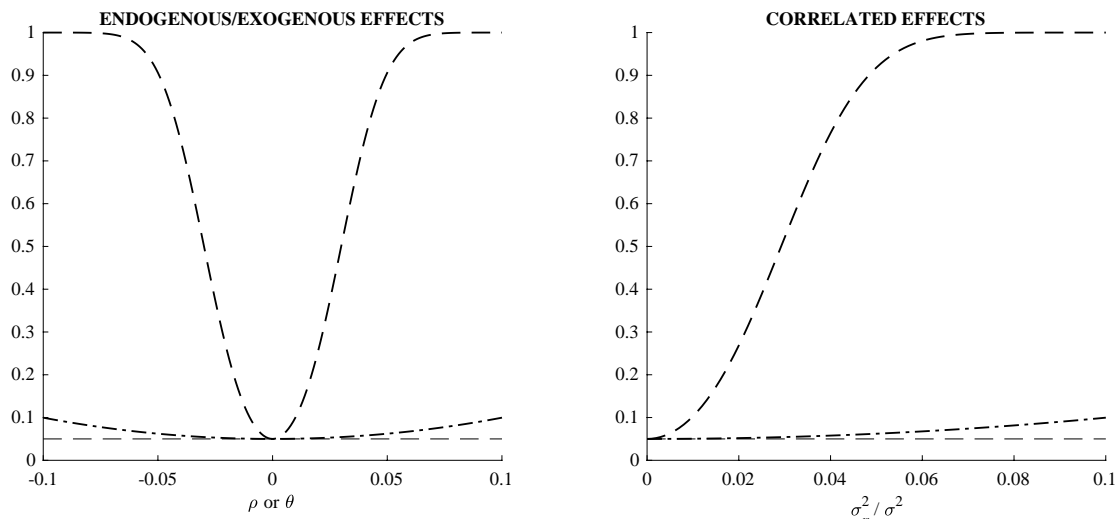
which is \hat{q}_r^{HO} .

From the argument above,

$$\hat{q}_r^{\text{HO}} = \sum_{g=1}^r \sum_{i=1}^{n_g} \ddot{x}_{g,i} \left(\bar{x}_{g,[i]} + \frac{x_{g,i}}{n_g - 1} \right) + o_p(\sqrt{r}).$$

The remained of the proof then parallels the proof of Theorem 1. □

Figure 1: Power in the design of [Stevenson \(2015\)](#)



Power against endogenous/exogenous effects (left plot) and correlated effects (right plot) in the design of [Guryan, Kroft and Notowidigdo \(2009\)](#) for our test (dashed line) and for the test of [Guryan, Kroft and Notowidigdo \(2009\)](#) (dashed-dotted line). A horizontal dashed-dotted line indicates the size of the test. Plots are based on theoretical calculations and are for 100 urns.

Additional asymptotic-power comparisons. [Stevenson \(2015\)](#) simulated power against the fixed correlated-effect alternative of $\sigma_\eta^2 / \sigma^2 = .03$ in a Monte Carlo setup similar to the one

used by [Guryan, Kroft and Notowidigdo \(2009\)](#). First 100 urns are randomly allocated a size $n_g \in \{39, 42, 45, 48, 51\}$ and then individuals are randomly allocated to a peer group of size three. Here we extend this exercise to all alternatives of the form $\sigma_\eta^2/\sigma^2 > 0$ and also provide power against endogenous/exogenous-effect alternatives. Figure 1 contains power curves for our test (dashed line) and the test of [Guryan, Kroft and Notowidigdo \(2009\)](#) (dashed-dotted line) based on our power calculations in the main text. Both tests are two-sided and are performed at the 5% significance level. The results confirm and generalize the low-power findings observed by [Stevenson \(2015\)](#). In addition, she also reported simulated power for the permutation test and for her split-sample test to be .502 and .455 when the alternative is $\sigma_\eta^2/\sigma^2 = .03$. In this case, the theoretical power of our test is .521.

References

- Caeyers, B. and M. Fafchamps (2020). Exclusion bias in the estimation of peer effects. NBER Working Paper No. 22565.
- Guryan, J., D. Kroft, and N. J. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics* 44, 289–302.
- Stevenson, M. (2015). Tests of random assignment to peers in the face of mechanical negative correlation: An evaluation of four techniques. Mimeo.