# TESTING RANDOM ASSIGNMENT TO PEER GROUPS
# SUPPLEMENT

KOEN JOCHMANS*
TOULOUSE SCHOOL OF ECONOMICS, UNIVERSITY OF TOULOUSE CAPITOLE

October 3, 2022

# 1 Power calculations

We consider three types of local alternatives, where the $x_{g,i}$ are correlated across peers. In the terminology of Manski (1993) these are (i) endogenous effects, (ii) contextual effects, and (iii) correlated effects. We begin by providing a closed-form expression for the variance of $q_r^{\mathrm{HO}}$ under the null. We then calculate $b_r$ under the alternatives (i)–(iii). Taken together, these results then yield the non-centrality parameter in the limit distribution of $t_r^{\mathrm{HO}}$. This is then used to assess power.

Throughout this section we focus attention on settings where peer groups do not overlap, which makes the final expressions more easily interpretable. We also enforce that $\mathbb{E}_0(x_{g,i}^4) = 3\sigma_g^4$, which yields a slightly shorter variance formula but is in no way essential to our findings. The underlying derivations, collected further down in this Appendix, do not make use of these restrictions.

**Variance expression.** Under these conditions the variance of $q_r^{\mathrm{HO}}$ under the null is equal to

$$v_r^{\mathrm{HO}} := \mathbb{E}_0(q_r^{\mathrm{HO}}\, q_r^{\mathrm{HO}}) = 2\sum_{g=1}^{r} \sigma_g^4\; \mathbb{E}_0\left(\sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{n_g}{n_g-1}\right). \tag{A.1}$$

We observe that $v_r^{\mathrm{HO}}$ is increasing in the size of the urns and decreasing in the size of the peer groups.

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.

**Endogenous effects.** In our first set of alternatives correlation among peers arises through

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

where $-1 < \rho < 1$ and the $\varepsilon_{g,i}$ are independent of the matrix $\boldsymbol{A}_g$. A drifting sequence of this model towards the null is obtained by setting $\rho = \varrho/\sqrt{r}$ for fixed values of $\varrho$. Such local alternatives imply that

$$b_r = 2\frac{\varrho}{\sqrt{r}} \sum_{g=1}^{r} \sigma_g^2\, \mathbb{E}\left( \sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{n_g}{n_g - 1} \right). \tag{A.2}$$

Note that this term depends on the design in the same way as does $v_r^{\mathrm{HO}}$ and so the same comparative statistics apply. Taken together, by an application of Theorem 1, $t_r^{\mathrm{HO}}$ will converge in distribution to a normal random variable with mean $\mu := \lim_{r\to\infty} b_r/\sqrt{v_r^{\mathrm{HO}}}$ and variance one. The larger $\mu$ (in magnitude) the smaller the probability of a type-II error. The non-centrality parameter $\mu$ is even simpler when errors are homoskedastic and the adjacency matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_r$ are drawn from a common distribution as, in that case,

$$\mu = \varrho \sqrt{2\,\mathbb{E}\left( \sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{n_g}{n_g - 1} \right)},$$

showing that power is monotone increasing in the (expected) size of the urns and decreasing in the size of the peer groups. When variances are urn specific the expression for $\mu$ is to be multiplied by

$$\lim_{r\to\infty} \frac{1}{\sqrt{r}} \frac{\sum_{g=1}^{r} \sigma_g^2}{\sqrt{\sum_{g=1}^{r} \sigma_g^4}} \leq 1,$$

where the bound follows from the Cauchy-Schwarz inequality. Hence, urn-specific variances are always power reducing. Nonetheless, note that $\mu > 0$, and so our test will detect endogenous-effect violations with probability approaching one for all possible configurations of urn sizes and peer groups.

**Contextual effects.** In our second class of alternatives correlation in peer characteristics comes from (latent) exogenous effects. Moreover,

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

where $\theta$ is a finite constant and, again, the $\varepsilon_{g,i}$ are independent of the matrix $\boldsymbol{A}_g$. For drifting sequences of the form $\theta = \vartheta/\sqrt{r}$,

$$b_r = 2\frac{\vartheta}{\sqrt{r}} \sum_{g=1}^{r} \sigma_g^2 \, \mathbb{E}\left( \sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{n_g}{n_g - 1} \right), \tag{A.3}$$

which is the identical to the bias under an endogenous-effect alternative where $\varrho = \vartheta$. Consequently, endogenous and exogenous effects are locally asymptotically equivalent. This finding is not surprising in light of the similar results on autoregressive and moving-average alternatives in classical testing problems in the time series literature (see, for example, Godfrey 1981).

**Correlated effects.** In our third class of alternatives peers are subject to a common additive shock drawn from a distribution with variance $\sigma_\eta^2$, independent of everything else. Thus (conditional on an urn fixed effect) the variance of $x_{g,i}$ is equal to $\sigma_\eta^2 + \sigma_g^2$ while the covariance between characteristics of peers is $\sigma_\eta^2$. In this case, the relevant drifting sequence has $\sigma_\eta^2 = \varsigma^2/\sqrt{r}$ and we find that the bias in $q_r^{\text{HO}}$ equals

$$b_r = \frac{\varsigma^2}{\sqrt{r}} \sum_{g=1}^{r} \mathbb{E}\left( (n_g - 1) - \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{m_g(i)}{n_g - 1} \right). \tag{A.4}$$

Because $\sum_{i=1}^{n_g} m_g(i) \leq n_g(n_g - 1)$, with equality if and only if all individuals in urn $g$ are each others peers we again have that $b_r > 0$ and so our test will be consistent against all correlated-effect alternatives. When $\sigma_g^2 = \sigma^2$ and the matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_r$ are drawn from a common distribution, the non-centrality parameter in the limit distribution of our test statistic is

$$\mu = \frac{\varsigma^2}{\sigma^2} \frac{\mathbb{E}\left( (n_g - 1) - \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{m_g(i)}{n_g - 1} \right)}{\sqrt{2\, \mathbb{E}\left( \sum_{i=1}^{n_g} \frac{1}{m_g(i)} - \frac{n_g}{n_g - 1} \right)}}.$$

Power is again increasing in $n_1, \ldots, n_r$. The impact of the size of the peers groups on power is less clear cut, however. On the one hand, larger peer groups reduce the variance and increase $\mu$. On the other hand, they also reduce the bias in $q_r^{\text{HO}}$, resulting in a loss of power.

# 2 Proofs

**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

$$\mathbb{E}_0(x_{g,i}\, x_{g,j}|\boldsymbol{A}_g) = \mathbb{E}_0(x_{g,i}\, x_{g,j}) = \begin{cases} \sigma_g^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \tag{A.5}$$

This normalization shortens the derivations to follow and is maintained throughout the Appendix.

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

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

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

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

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

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 $\mathbb{E}_0(x_{g,i}\, x_{g,j}|\boldsymbol{A}_g) = \mathbb{E}_0(x_{g,i}\, x_{g,j}) = 0$ for all $i \neq j$ and that $(\boldsymbol{A})_{i,i} = 0$.

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

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

using the same arguments as for the first term and the accounting identity $\sum_{j=1}^{n_g}(\boldsymbol{A}_g)_{i,j} = m_g(i)$.

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

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

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, $\mathbb{E}_0(\sum_{g=1}^{r}\sum_{i=1}^{n_g}\bar{x}_{g,[i]}\,\tilde{\bar{x}}_{g,[i]})$. Using the definition of $\bar{x}_{g,[i]}$, it can be written as

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

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

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

where we have first exploited (A.5) and then used $\sum_{j=1}^{n_g}(\boldsymbol{A}_g)_{i,j} = m_g(i)$ together with the fact that $(\boldsymbol{A}_g)_{i,j}^2 = (\boldsymbol{A}_g)_{i,j}$.

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

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

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

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 $r\to\infty$.

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

$$\begin{aligned}
\frac{1}{n_g}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\frac{m_g(i\cap j)}{m_g(i)\,m_g(j)} &= \frac{1}{n_g}\sum_{i=1}^{n_g}\frac{m_g(i\cap i)+\sum_{j\neq i}m_g(i\cap j)}{m_g(i)^2}\\
&= \frac{1}{n_g}\sum_{i=1}^{n_g}\left(\frac{1}{m_g(i)}+\frac{\sum_{j\neq i}(\boldsymbol{A}_g)_{i,j}\,(m_g(i)-1)}{m_g(i)^2}\right)\\
&= \frac{1}{n_g}\sum_{i=1}^{n_g}\left(\frac{1}{m_g(i)}+\frac{m_g(i)-1}{m_g(i)}\right)\\
&= 1,
\end{aligned} \tag{A.6}$$

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

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

as claimed. This, after simplication 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. □

**Proof of Theorem 1.** By independence of the urns the variance of $q_r^{\text{HO}}$ 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

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

We handle each of these in turn. As subtracting $b_r$ amounts to a mere recentering of $q_r^{\text{HO}}$ 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

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

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

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

Then we can write

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

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

Hölder's inequality,

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

for some $\theta>0$. The first right-hand side term is finite for any choice of $\theta$ 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 $v_r^{\mathrm{HO}}$ grows at the rate $r$. We may again write

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

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

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

Each of these terms is non-zero and bounded. Hence, $v_r^{\text{HO}}$, 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

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

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. □

**Proof of Equation (A.1).** 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 $q_r^{\text{HO}}$ 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

$$\sum_{g=1}^{r}\mathbb{E}_0\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\left(\tilde{x}_{g,i}\left(\bar{x}_{g,[i]}+\frac{x_{g,i}}{n_g-1}\right)\right)\left(\tilde{x}_{g,i'}\left(\bar{x}_{g,[i']}+\frac{x_{g,i'}}{n_g-1}\right)\right)\right). \tag{A.7}$$

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

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

and expands as

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

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

The first expectation is

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

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

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

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

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

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

$$\mathbb{E}_0\left(\sum_{i=1}^{n_g}\frac{1}{m_g(i)} - \frac{1}{n_g}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\frac{m_g(i\cap j)}{m_g(i)\,m_g(j)} + \sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\frac{(\boldsymbol{A}_g)_{i,j}}{m_g(i)\,m_g(j)}\right)\sigma_g^4, \tag{A.8}$$

which deals with the first term in (A.7).

The second term that needs to be calculated is

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

We evaluate each of these expectations in turn next.

The first expectation is

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

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

The second expectation is

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

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

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

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

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

Here, again, a fourth-order term arises from the summands where $i' = j = k = k'$ while three different combinations of indices contribute terms involving $\sigma_g^4$; 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

$$\mathbb{E}_0\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{\tilde{x}_{g,i}\,\tilde{x}_{g,i'}\,\bar{x}_{g,[i]}\,x_{g,i'}}{(n_g-1)}\right) = \mathbb{E}_0\left(\frac{3}{n_g}-\frac{2}{(n_g-1)}-1\right)\sigma_g^4 - \mathbb{E}_0\left(\frac{1}{n_g}\right)\gamma_g^4, \tag{A.9}$$

and gives (up to the factor 2) and expression for the second term in (A.7).

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

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

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

The first expectation is

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

The second expectation is

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

The third expectation is

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

This then yields

$$\mathbb{E}_0\left(\sum_{i=1}^{n_g}\sum_{i'=1}^{n_g}\frac{\tilde{x}_{g,i}\,\tilde{x}_{g,i'}\,x_{g,i}\,x_{g,i'}}{(n_g-1)^2}\right) = \mathbb{E}_0\left(\frac{n_g}{(n_g-1)}\left(1-\frac{2}{n_g}+\frac{3}{n_g^2}\right)\right)\sigma_g^4 + \mathbb{E}_0\left(\frac{n_g}{(n_g-1)^2}\left(1-\frac{2}{n_g}+\frac{1}{n_g^2}\right)\right)\gamma_g^4 \tag{A.10}$$

for the third and final term in (A.7).

Now, collecting results by combining (A.7) with (A.8)–(A.10) yields

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

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

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

by (A.6). This is the variance formula stated in Equation (A.1) in the main text. □

**Proof of Equation (A.2).** Suppose that

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

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

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

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

$$\boldsymbol{G}_g := \boldsymbol{D}_g^{-1} \boldsymbol{A}_g, \qquad \boldsymbol{D}_g := \operatorname{diag}(m_g(1), \ldots, m_g(n_g)),$$

we have the linear system

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

which has reduced form

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

Moreover,

$$x_{g,i} = \varepsilon_{g,i} + \rho\, \frac{1}{m_g(i)} \sum_{j=1}^{n_g} (\boldsymbol{A}_g)_{i,j}\, \varepsilon_{g,j} + \rho^2\, \frac{1}{m_g(i)} \sum_{j=1}^{n_g} \sum_{k=1}^{n_g} (\boldsymbol{A}_g)_{i,j}\, (\boldsymbol{A}_g)_{j,k}\, \varepsilon_{g,k} + \cdots. \tag{A.11}$$

We then obtain

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

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

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

which mimics (A.5). Next we calculate

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

and, finally,

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

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

$$\mathbb{E}_{\frac{\varrho}{\sqrt{r}}}\left(x_{g,i}\,x_{g,i'}|\boldsymbol{A}_g\right) = \left\{\begin{array}{ll}\sigma_g^2 & \text{if } i=i' \\ \frac{\varrho}{\sqrt{r}}\left(\frac{(\boldsymbol{A}_g)_{i,i'}}{m_g(i)}+\frac{(\boldsymbol{A}_g)_{i',i}}{m_g(i')}\right)\sigma_g^2 & \text{if } i\neq i'\end{array}\right. . \tag{A.12}$$

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

The bias is

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

We calculate each of the expectations in turn.

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

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

because,

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

and

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

where we have made extensive use of (A.12).

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

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

again exploiting (A.12).

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

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

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

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

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 (A.2). □

**Proof of Equation (A.3).** To see that endogenous effects and contextual effects are locally asymptotically equivalent note that the latter violation of the null is of the form

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

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 coincides with the reduced form in (A.11), up to first-order. Consequently, it is immediate that $\mathbb{E}_{\frac{\vartheta}{\sqrt{r}}}\left(x_{g,i}\, x_{g,i'} | \boldsymbol{A}_g\right)$ satisfies the expansions in (A.12). This, then, implies that the asymptotic bias induced by contextual effects, too, is identical. □

**Proof of Equation (A.4).** Consider non-overlapping peer groups. Peers are subject to a common random effect drawn from a distribution with zero mean and variance $\sigma_\eta^2$. Consequently,

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

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

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

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

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

This follows from the observations that

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

and that

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

here, the last equality exploits the fact that peer groups do not overlap by appealing to (A.6).

Second, we have

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

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. □

**Proof of Equation** (2.4)**.** 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

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

to

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

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

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

is equal to

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

Here we have used that

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

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

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

This is Equation (2.4) 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. □

**Proof of (2.5).** A short calculation shows that, in (2.4), 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

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

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 $s_r^{\mathrm{HO}}$. Moreover, because $p_n \in (0,1)$,

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

so that

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

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

**Proof of Equation (3.7).** The proof mimics the proof of Equation (1.2). The only difference arises in the second term where, now,

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

from which the result follows. □

**Proof of Equation (3.8).** For urn $g$ write the $n_g \times n_g$ matrix that transforms observations into deviations from the within-urn mean as

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

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

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

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

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

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

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

An unbiased plug-in estimator of the bias in (3.7) thus is

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

on verifying that, indeed,

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

This then immediately also implies the unbiasedness of $q_r^{\mathrm{HC}}$ as defined in (3.8). The proof is complete. □

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

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

and note that these weights are again uniformly bounded. □

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

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

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

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

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

Let $\ddot{x}_{g,i} := \tilde{x}_{g,i} - \tilde{\boldsymbol{w}}'_{g,i}\,\boldsymbol{\beta}_1$ and $\ddot{\bar{x}}_{g,[i]} := \tilde{\bar{x}}_{g,[i]} - \tilde{\boldsymbol{w}}'_{g,i}\,\boldsymbol{\beta}_2$ 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 $\boldsymbol{w}_{g,i}$. Then

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_1 + \left( \sum_{g=1}^{r} \sum_{i=1}^{n_g} \tilde{\boldsymbol{w}}_{g,i}\, \boldsymbol{w}'_{g,i} \right)^{-1} \left( \sum_{g=1}^{r} \sum_{i=1}^{n_g} \tilde{\boldsymbol{w}}_{g,i}\, \ddot{x}_{g,i} \right),$$

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

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

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

Here, the last equality is a consequence of

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

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

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

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

With the representation

$$
\sum_{g=1}^{r}\sum_{i=1}^{n_g} \bar{x}_{g,[i]}\,\dot{x}_{g,i} = \sum_{g=1}^{r}\sum_{i=1}^{n_g} \ddot{x}_{g,i}\,\bar{x}_{g,[i]} + o_p(\sqrt{r}) \tag{A.13}
$$

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

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

Further, we remark that, under the null, the projection coefficient $\hat{\boldsymbol{\beta}}_1$ 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

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

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

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

The transitions in this display have made use of

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

where

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

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

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 $\hat{q}_r^{\mathrm{HO}}$.

From the argument above,

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

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

# References


Caeyers, B. and M. Fafchamps (2020). Exclusion bias in the estimation of peer effects. NBER Working Paper No. 22565.

Godfrey, L. G. (1981). On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis. *Econometrica 49*, 1443–1455.

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.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*, 531–542.