# EXACT P-VALUES FOR NETWORK INTERFERENCE

Susan Athey
Dean Eckles
Guido W. Imbens

# Appendix A: Why the Bond et al Randomization P-values are Not Valid

Bond, Fariss, Jones, Kramer, Marlow, Settle, and Fowler (2013), Bond et al. from hereon, are also interested in testing for spillovers (Hypothesis 2). They wish to use testing procedures that are robust to the network structure. We show here analytically that there procedures are not valid in general, and can lead to over-rejections of 0.05-level tests at rates as high as 0.20 because they ignore the variation arising from own treatment effects.

Bond et al. focus on the difference between the average of an ego's outcome over all edges where the alter is exposed, and the average over all edges where the alter is not exposed:

$$T_{\mathrm{B}}(\mathbf{W}, \mathbf{Y}, \mathbf{G}) = \frac{\sum_{i, j \neq i} G_{ij} \cdot W_j \cdot Y_i}{\sum_{i, j \neq i} G_{ij} \cdot W_j} - \frac{\sum_{i, j \neq i} G_{ij} \cdot (1 - W_j) \cdot Y_i}{\sum_{i, j \neq i} G_{ij} \cdot (1 - W_j)}. \tag{A.1}$$

Under Hypothesis 2 the expected value of this statistic is zero, which makes it promising for testing this hypothesis. However, because of the network structure there may dependence between the terms in each of these averages, and its variance is difficult to estimate for a general network structure.

Bond et al. look at a randomization-based distribution for this statistic to test the null hypothesis of no spillovers. The distribution is obtained by re-assigning the treatment vector $\mathbf{W}$, assuming there is no effect of the treatment whatsoever, and deriving from there the quantiles of the $T_{\mathrm{B}}$ distribution. This implicitly assumes for these calculations that there is no effect of the treatment whatsoever (Hypothesis 1), which is stronger than the no-spillover null hypothesis (Hypothesis 2) that they are interested in testing. The reason for this is that if one allows for direct effects of the treatment on the own outcomes, and only assumes no spillovers, one cannot infer the value of the statistic $T_{\mathrm{B}}$ for alternative values of the treatment assignment vector: the no-spillover null hypothesis is not sharp. The concern is that using the randomization that is based on a stronger null hypothesis is not innocuous. Bond et al justify the use of this method using simulations in which the stronger null is true.

Here we show through analytic calculations for a particular example that p-values based on these calculations are not valid, even in large samples, let alone in finite samples, and that the deviations from nominal rejection probabilities can be substantial. In general, because their calculations ignore one source of variation in the distribution of the statistic, the p-values will be too small, leading to rejections of 0.05-level tests at rates as high as 0.20.

We focus on an example with a particular network structure that allows us to simplify the large sample approximations. The population consists of $2 \cdot N$ units, partitioned into $N$ pairs. Out of these $2 \cdot N$ units $N$ units are randomly selected to be exposed to the active treatment. We maintain the assumption that there are no spillovers. The potential outcomes are

$$Y_i(0) = 0, \qquad \text{and} \ \ Y_i(1) = 1,$$

so that the direct treatment effect is equal to 1. The $N$ pairs can be partitioned into three sets: $M_{00}$ pairs with both units exposed to the control treatment, $M_{01}$ pairs with exactly one unit exposed to the control treatment and one unit exposed to the active treatment, and $M_{11}$ pairs with both units exposed to the active treatment. The number of each of these sets, $M_{00}$, $M_{01}$, and $M_{11}$ are random, but, because the total number of pairs is fixed at $N$, it follows that $M_{00} + M_{01} + M_{11} = N$, and because exactly $N$ units are exposed to the active treatment, it must be the case that $M_{00} = M_{11}$. Hence we can rewrite these numbers in terms of a scalar random integer: define $M = M_{11}$, so that $M_{00} = M$, and $M_{01} = N - 2 \cdot M$. The expected value of $M$ is $N \cdot (1/2) \cdot ((N-1)/(2 \cdot N)) \approx N/4$. However, the variance is not $N \cdot (1/4) \cdot (3/4)$ because of the fixed number of treated units. We can approximate the large sample distribution of $\sqrt{N}(M/N - 1/4)$ by looking at the joint distribution for $(\sqrt{N} \cdot (M_{00}/N - 1/4), \sqrt{N} \cdot (M_{01}/N - 1/2), \sqrt{N} \cdot (M_{11}/N - 1/4))$, based on independent random assignment to the treatment for each unit. This leads to

$$\begin{pmatrix} \sqrt{N} \cdot (M_{00}/N - 1/4) \\ \sqrt{N} \cdot (M_{01}/N - 1/2) \\ \sqrt{N} \cdot (M_{11}/N - 1/4) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{16} \cdot \begin{pmatrix} 3 & -2 & -1 \\ & 4 & -2 \\ & & 3 \end{pmatrix} \right).$$

This implies that

$$\begin{pmatrix} \sqrt{N} \cdot (M_{11}/N - 1/4) \\ \sqrt{N} \cdot (2 \cdot M_{11}/N + M_{01}/N) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{16} \cdot \begin{pmatrix} 3 & 4 \\ & 12 \end{pmatrix} \right).$$

Now define $M = M_{11}$ and condition on $M_{01}/N + 2 \cdot M_{11}/N = 0$. Because the correlation between $\sqrt{N} \cdot (M_{11}/N - 1/4)$ and $\sqrt{N} \cdot (M_{01}/N + 2 \cdot M_{11}/N$ is $\rho = 4/sqrt24$, the conditional variance of $\sqrt{N} \cdot (M_{11}/N - 1/4)$

given $\sqrt{N} \cdot (M_{01}/N + 2 \cdot M_{11}/N = 0$ is $(3/16) \cdot (1 - \rho^2) = 1/16$, and

$$\sqrt{N} \cdot \left( \frac{M}{N} - \frac{1}{4} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{1}{16} \right).$$

Now consider the statistic $T_{\mathrm{B}}$. We calculate first the actual distribution of this statistic under the randomization distribution. Then we compare this to the distribution Bond et al use for the calculation of p-values.

There are $2 \cdot N$ edges. Out of these $N$ have treated alters and $N$ have control alters. For the $N$ edges with treated alters $2 \cdot M_{11} = 2 \cdot M$ have treated egos, and so have realized outcome equal to $Y_i(1) = 1$, and $M_{01} = N - 2 \cdot M$ have control egos, and so have realized outcomes equal to $Y_i(0) = 0$. The average realized outcome for egos with treated alters is therefore $2 \cdot M/N$. Similarly, for the $N$ edges with control alters, there are $2 \cdot M_{00} = 2 \cdot M$ edges with control egos and realized outcomes $Y_i(0) = 0$, and $M_{01} = N - 2 \cdot M$ edges with treated egos and thus $Y_i(1) = 1$, leading to an average realized outcome equal to $1 - 2 \cdot M/N$. Hence the value of the statistic is

$$T_{\mathrm{B}} = 2 \cdot \frac{M}{N} - \left( 1 - 2 \cdot \frac{M}{N} \right) = 4 \cdot \left( \frac{M}{N} - \frac{1}{4} \right).$$

The actual distribution of the normalized statistic, under random assignment, is

$$\sqrt{N} \cdot T_{\mathrm{B}} = \sqrt{N} \cdot \left( \frac{4 \cdot M}{N} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Now consider the distribution used by Bond et al for the calculation of their p-values. They calculate the randomization distribution, assuming that there are no effects of the treatment whatsoever. Under this randomization distribution, there are always $N$ egos with treated alters, and $N$ egos with control alters. Out of the $2 \cdot N$ units there are $N$ with realized outcome equal to 1 and $N$ with realized outcome equal to 0, so that the total average outcome is exactly $1/2$. Hence, if the average of the outcome for the egos with treated alters is equal to $\overline{Y}_{\mathrm{t}}$, the average of the outcome for egos with control alters is equal to $\overline{Y}_{\mathrm{c}} = 1 - \overline{Y}_{\mathrm{t}}$. Therefore the difference in the average outcome for egos with treated alters and the average outcome for egos with control alters is equal to $2 \cdot \overline{Y}_{\mathrm{t}} - 1$. To infer the randomization distribution used by Bond et al, we need to infer the distribution of $\overline{Y}_{\mathrm{t}}$ under their randomization distribution. We can write $\overline{Y}_{\mathrm{t}}$ as

$$\overline{Y}_{\mathrm{t}} = \frac{1}{N} \sum_{i=1}^{2N} W_i^p \cdot Y_i,$$

where $W_i^p$ is an indicator for unit $i$ having a treated alter. We are interested in this distribution under random assignment of $Z_i$, with $\sum_{i=1}^{2N} Z_i = N$, for fixed $\mathbf{Y}$. (It is the treating of $\mathbf{Y}$ as fixed that is not correct here – if we change the treatment of the alter for unit $i$ we may be changing the value of the outcome for unit $i$'s alter. Thus the $Y_i$ are stochastic, leading to additional variation in the test statistic that is not taken into account in the B procedure.) Note that $\sum_{i=1}^{2 \cdot N} Y_i = N$ and $\sum_{i=1}^{2 \cdot N} W_i^p = N$. The treatments (and thus the peer treatments) are randomly assigned, with $\mathrm{pr}(W_i^p = 1) = 1/2$ and $\mathrm{pr}(W_i^p = 1 | W_j^p = 1) = (N-1)/(2 \cdot N - 1)$. Define $D_i = 2 \cdot W_i^p - 1$ so that $W_i^p = (D_i + 1)/2$, and

$$\mathbb{E}[D_i] = 0, \qquad D_i^2 = 1, \qquad \mathbb{E}[D_i \cdot D_j] = -\frac{1}{2 \cdot N - 1}, \text{ for } j \neq i.$$

Now

$$\overline{Y}_{\mathrm{t}} = \frac{1}{N} \sum_{i=1}^{2N} Y_i \cdot \frac{D_i + 1}{2}$$

$$= \frac{1}{N} \sum_{i=1}^{2N} Y_i \cdot \frac{1}{2} + \frac{1}{2N} \sum_{i=1}^{2N} Y_i \cdot D_i = \frac{1}{2} + \frac{1}{2N} \sum_{i=1}^{2N} Y_i \cdot D_i.$$

Then

$$\mathbb{E}\left[ \overline{Y}_{\mathrm{t}} \right] = 1/2,$$

and

$$\mathbb{V}\left( \overline{Y}_{\mathrm{t}} \right) = \frac{1}{4 \cdot N^2} \cdot \mathbb{E}\left[ \left( \sum_{i=1}^{2N} Y_i \cdot D_i \right)^2 \right] = \frac{1}{4 \cdot N^2} \cdot \mathbb{E}\left[ \sum_{i=1}^{2N} D_i^2 \cdot Y_i^2 + \sum_{i=1}^{2N} \sum_{j \neq i} D_i \cdot D_j \cdot Y_i \cdot Y_j \right]$$

[32]

$$= \frac{1}{4 \cdot N^2} \cdot \sum_{i=1}^{2N} Y_i + \frac{1}{4 \cdot N^2} \cdot \sum_{i=1}^{2N} \sum_{j \neq i} Y_i \cdot Y_j \cdot \mathbb{E}[D_1 \cdot D_2]$$

$$= \frac{1}{4 \cdot N} - \frac{1}{4 \cdot N^2} \cdot N \cdot (N-1) \cdot \frac{1}{2 \cdot N - 1}$$

$$= \frac{1}{4 \cdot N} - \frac{1}{4 \cdot N} \cdot \frac{N-1}{2 \cdot N - 1} \approx \frac{1}{8 \cdot N}.$$

Hence the variance of $N \cdot \overline{Y}_t$ is equal to $1/8$, and thus the variance of Bond et al randomization distribution is $4 \cdot N \cdot \mathbb{V}(\overline{Y}_t)$ which is equal to 0.5. The actual distribution has variance equal to 1, which is twice as large. The implication is that the for a two-sided test at the 0.05 level the rejection probability based on using the incorrect Bond et al randomization distribution is 0.157. Bond et al implicitly use the wrong variance of 0.5 for the test statistic, leading to

$$\mathrm{pr}\left(\sqrt{2} \cdot |T_B| > 1.96\right) = \mathrm{pr}\left(|T_B| > \sqrt{2} \cdot 1.96\right)$$

$$= \mathrm{pr}\left(|T_B| > \frac{1.96}{\sqrt{2}}\right) \approx \mathrm{pr}\left(|T_B| > 1.386\right) \approx 0.157.$$

We carried out a small simulation study to verify these analytic calculations. We use $N = 1000$ pairs, 10,000 replications, and use 1,000 draws from the randomization distribution. We reject the null hypothesis if the Bond et al p-value is less than 0.05. This leads us to reject at a rate equal to 0.153, close to the theoretical rejection rate we calculated above which is equal to 0.157. (A 95% confidence interval for the rejection rate is $(0.144, 0.163)$).

Appendix B: Derivation of the Score Test Statistic for the Null of No Spillovers

In terms of the potential outcomes the linear-in-means model in (5.9) corresponds to

$$\mathbf{Y}(\mathbf{w}) = \alpha_0 \cdot \left(I - \tau_{\mathrm{endog}} \cdot \overline{\mathbf{G}}\right)^{-1} \cdot \iota_N + \tau_{\mathrm{direct}} \cdot \left(I - \tau_{\mathrm{endog}} \cdot \overline{\mathbf{G}}\right)^{-1} \mathbf{w} + \left(I - \tau_{\mathrm{endog}} \cdot \overline{\mathbf{G}}\right)^{-1} \varepsilon. \tag{B.1}$$

The expected value of the observed outcomes given the assignment is, given the random assignment,

$$\mathbb{E}[\mathbf{Y}^{\mathrm{obs}}|\mathbf{W} = \mathbf{w}] = \mathbb{E}[\mathbf{Y}(\mathbf{w})] = \alpha_0 \cdot \left(I - \tau_{\mathrm{endog}} \cdot \overline{\mathbf{G}}\right)^{-1} \iota_N + \tau_{\mathrm{direct}} \cdot \left(I - \tau_{\mathrm{endog}} \cdot \overline{\mathbf{G}}\right)^{-1} \mathbf{w}. \tag{B.2}$$

Under the null hypothesis that $\tau_{\mathrm{endog}} = 0$, the least squares estimates for the remaining parameters based on outcomes for focal units are

$$\hat{\alpha}_0 = \overline{Y}_{F,0}^{\mathrm{obs}}, \qquad \text{and} \quad \hat{\tau}_{\mathrm{direct}} = \overline{Y}_{F,1}^{\mathrm{obs}} - \overline{Y}_{F,c}^{\mathrm{obs}},$$

where, for $w = 0, 1$, $\overline{Y}_{F,w}^{\mathrm{obs}}$ is the average outcome for focal units with $W_i = w$,

$$\overline{Y}_{F,w}^{\mathrm{obs}} = \frac{1}{N_{F,w}} \sum_{i:F_i=1, W_i=w} Y_i^{\mathrm{obs}},,$$

and $N_{F,w}$ is the number of focal units with $W_i = w$. Hence the residual under the null is

$$\hat{\varepsilon}_i^{\mathrm{null}} = Y_i^{\mathrm{obs}} - \hat{\alpha}_0 - W_i \cdot \hat{\tau}_{\mathrm{direct}}.$$

Under normality of the outcome the score for $\tau_{\mathrm{endog}} = 0$ is proportional to the covariance of the residual under the null and the derivative of the expectation in (B.2), with respect to $\tau_{\mathrm{endog}}$, evaluated at $\tau_{\mathrm{endog}} = 0$. The derivative of the expectation at $\tau_{\mathrm{endog}} = 0$ is

$$\frac{\partial}{\partial \tau_{\mathrm{direct}}} \mathbb{E}[\mathbf{Y}^{\mathrm{obs}}|\mathbf{W}] = \alpha_0 \cdot \overline{\mathbf{G}} \iota_N + \tau_{\mathrm{direct}} \cdot \overline{\mathbf{G}} \mathbf{W} = \alpha_0 \cdot \overline{\mathbf{G}}(\iota_N - \mathbf{W}) + (\tau_{\mathrm{direct}} + \alpha_0) \cdot \overline{\mathbf{G}} \mathbf{W}.$$

Substituting $\overline{Y}_{F,0}^{\mathrm{obs}}$ for $\alpha_0$ and $\overline{Y}_{F,1}^{\mathrm{obs}} - \overline{Y}_{F,0}^{\mathrm{obs}}$ for $\tau_{\mathrm{direct}}$ suggests that a natural test statistic would be the covariance of the residual under the null and $\overline{Y}_{F,0}^{\mathrm{obs}} \cdot \overline{\mathbf{G}}(\iota_N - \mathbf{W}) + \overline{Y}_{F,1}^{\mathrm{obs}} \cdot \overline{\mathbf{G}} \mathbf{W}$. This leads to the following average score:

$$\frac{1}{N_F} \sum_{i \in \mathbb{P}_F} \left\{ \left(Y_i^{\mathrm{obs}} - \overline{Y}_{F,0}^{\mathrm{obs}} - W_i \cdot (\overline{Y}_{F,1}^{\mathrm{obs}} - \overline{Y}_{F,0}^{\mathrm{obs}})\right) \cdot \sum_{j=1}^{N} \overline{G}_{ij} \cdot \left((1 - W_j) \cdot \overline{Y}_{F,0}^{\mathrm{obs}} + W_j \cdot \overline{Y}_{F,1}^{\mathrm{obs}}\right) \right\}.$$

[33]

Because $\sum_{j=1}^{N} \overline{G}_{ij} = 1$, in combination with the fact that the residuals average to zero, it follows that the score statistic is proportional to the covariance between the residual under the null and $\sum_{j=1}^{N} \overline{G}_{ij} \cdot W_j$, which is the fraction of treated neighbors, leading to the score statistic

$$T_{\text{score}} = \text{Cov}\left(Y_i^{\text{obs}} - \overline{Y}_{F,0}^{\text{obs}} - W_i \cdot (\overline{Y}_{F,1}^{\text{obs}} - \overline{Y}_{F,0}^{\text{obs}}), \sum_{j=1}^{N} W_j \cdot \overline{G}_{ij} \,\middle|\, \sum_{j=1}^{N} G_{ij} > 0, F_i = 1\right)$$

$$= \text{Cov}\left(Y_i^{\text{obs}} - \hat{\alpha} - \hat{\tau}_{\text{direct}} \cdot W_i, \sum_{j=1}^{N} W_j \cdot \overline{G}_{ij} \,\middle|\, \sum_{j=1}^{N} G_{ij} > 0, F_i = 1\right),$$

which is the expression in (5.8).