Supplemental material for "Evaluating How Licensing-Law Strategies Will Impact Disparities in Tobacco Retailer Density: A Simulation in Ohio"

Peter F. Craigmile[1], **Nathaniel Onnen**[1], **Elli Schwartz**[2], **Allison M. Glasser**[2], **Megan E. Roberts**[2].

[1] Department of Statistics, The Ohio State University, Columbus, OH, USA

[2] College of Public Health, The Ohio State University, Columbus, OH, USA

Last updated June 8, 2020

We provide more detail on how we account for the multivariate spatial dependence in the pre- and post-policy counts as we investigate different questions about the equality impact.

Suppose that there are m census tracts, and let Y_{1i} denote the number of establishments in census tract $i = 1, \ldots, m$ before a policy is made – we call $\boldsymbol{Y}_1 = (Y_{11}, \ldots, Y_{1m})^T$ the **pre-policy counts**. After we enact a policy, let Y_{2i} denote the number of establishments in census tract i – we call $\boldsymbol{Y}_2 = (Y_{21}, \ldots, Y_{2m})^T$ the **post-policy counts**. For each census tract let P_i denote the population (in thousands) for tract i. We assume that in each tract the population is unchanged pre- and post- policy.

Investigating the absence of disparities

To investigate whether the establishment densities are different for low and high values of a dichotomous covariate pre- and post-policy we consider the following.

Let d_i denote the value of covariate in census tract i, with $d_i = 0$ at the low level and $d_i = 1$ at the high level. Taking a log transformations of the pre- and post-policy establishment rates, let

$$Z_{ki} = \log((Y_{ki}+1)/P_i)\,, \quad k=1,2,\ i=1,\ldots,m$$

(We add 1 to guard against taking the log of zero). With this transformation, a joint normal distribution is reasonable for $\{Z_{ki} : k=1,2,\ i=1,\ldots,m\}$.

In terms of the covariate d_i, we assume that the mean of the pre-policy log establishment rates in tract i is

$$E(Z_{1i}) = \gamma_1 + \gamma_2 d_i, \quad i=1,\ldots,m. \tag{1}$$

The mean of the post-policy log establishment rates in tract i can be different with

$$E(Z_{2i}) = \gamma_3 + \gamma_4 d_i, \quad i=1,\ldots,m. \tag{2}$$

In these two models for the mean, γ_2 measures the effect of the disparity for the pre-policy log establishment rates and γ_4 measures the effect of the disparity for the post-policy log establishment rates. Thus to compare the disparities post-policy minus pre-policy we need to estimate $\gamma_4 - \gamma_2$. The least squares estimate of γ_2 is

$$\widehat{\gamma}_2 = \text{average}\{Z_{1i} : d_i = 1\} - \text{average}\{Z_{1i} : d_i = 0\},$$

and the least squares estimate of γ_4 is

$$\widehat{\gamma}_4 = \text{average}\{Z_{2i} : d_i = 1\} - \text{average}\{Z_{2i} : d_i = 0\},$$

which means that our estimate of $\gamma_4 - \gamma_2$ is $\widehat{\gamma}_4 - \widehat{\gamma}_2$. (For completeness the least squares estimate of γ_1 is

$$\widehat{\gamma}_1 = \text{average}\{Z_{1i} : d_i = 0\}$$

and the least squares estimate of γ_3 is

$$\widehat{\gamma}_3 = \text{average}\{Z_{2i} : d_i = 0\}.$$

This will be used in the next section.)

To calculate the standard error for $\widehat{\gamma}_4 - \widehat{\gamma}_2$, we need to assume a joint distribution for our log transformed data. We assume a separable bivariate conditional autoregressive (CAR) model (e.g., Banerjee et al., 2014, Section 7.4) for the counts using vector-matrix notation. Let $\boldsymbol{Z}_1 = (Z_{1i})^T$ denote the vector of pre-policy log establishment rates and $\boldsymbol{Z}_2 = (Z_{2i})^T$ denote the vector of post-policy log establishment rates. We then specify a $m \times m$ spatial proximity matrix $\boldsymbol{W}$ as follows: the (i, j) element of $\boldsymbol{W}$ is equal to 1 if census tract j is a neighbor of tract i, and zero otherwise (we assume each tract cannot be a neighbor of themselves). Next, let $\boldsymbol{C}$ denote a diagonal $m \times m$ matrix, where the ith diagonal element is equal to the number of neighbors that census tract i has. Then we assume

$$\begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{pmatrix} \sim N_{2m}(\boldsymbol{D}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$$

where the covariance matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \tau_1^2(\boldsymbol{C} - \delta\boldsymbol{W})^{-1} & \rho\tau_1\tau_2(\boldsymbol{C} - \delta\boldsymbol{W})^{-1} \\ \rho\tau_1\tau_2(\boldsymbol{C} - \delta\boldsymbol{W})^{-1} & \tau_2^2(\boldsymbol{C} - \delta\boldsymbol{W})^{-1} \end{bmatrix}.$$

Our parameter vector that appears in the mean is $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$ and the design matrix is

$$\boldsymbol{D} = \begin{bmatrix} \mathbf{1}_m & \boldsymbol{x} & \mathbf{0}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{0}_m & \mathbf{1}_m & \boldsymbol{x} \end{bmatrix},$$

where $\mathbf{0}_m$ is a vector of m zeros, $\mathbf{1}_m$ is a vector of m ones, and $\boldsymbol{x} = (x_1, \ldots, x_m)^T$ is the covariate vector.

In the covariance matrix $\boldsymbol{\Sigma}$, the parameter $-1 < \delta < 1$ is known as the *spatial dependence parameter* and in this model measures the spatial dependence in neighboring census tracts both within and between the pre- and post-policy counts. The parameter $-1 < \rho < 1$ is a correlation parameter, $\tau_1^2 > 0$ is a variance parameter for the pre-policy log establishment rates, and $\tau_2^2 > 0$ is a variance parameter for the post-policy log establishment rates.

Then the standard error for $\widehat{\gamma}_4 - \widehat{\gamma}_2$ is

$$\sqrt{\boldsymbol{a}^T(\boldsymbol{D}^T\boldsymbol{D})^{-1}(\boldsymbol{D}^T\boldsymbol{\Sigma}\boldsymbol{D})(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{a}},$$

where $\boldsymbol{a} = (0, -1, 0, 1)^T$. In practice we use an estimate of $\boldsymbol{\Sigma}$ calculated with maximum likelihood estimates of δ, τ_1^2, τ_2^2 and ρ based on the residuals, calculated using the least squares estimates of $\boldsymbol{\gamma}$.

Investigating the percentage reduction

From (1) and (2) the mean log establishment rate at the low level of the covariate pre-policy (i.e., when $d_i = 0$) is γ_1 and at the low level of the covariate post-policy is γ_3. Thus the percentage reduction in the log establishment rate at low level of the covariate can be defined to be

$$\left(\frac{\gamma_1 - \gamma_3}{\gamma_1}\right) \times 100\% \;=\; \left(1 - \frac{\gamma_3}{\gamma_1}\right) \times 100\%.$$

Similarly from (1) and (2), the percentage reduction in the log establishment rate at high level of the covariate (i.e., when $d_i = 1$) can be defined to be

$$\left(\frac{[\gamma_1 + \gamma_2] - [\gamma_3 + \gamma_4]}{[\gamma_1 + \gamma_2]}\right) \times 100\% \;=\; \left(1 - \frac{[\gamma_3 + \gamma_4]}{[\gamma_1 + \gamma_2]}\right) \times 100\%.$$

We convert to the original scale instead, noting first that on the log scale that the mean log rate is equal to the median log rate. Taking exponentials, we have that exp(median log rate) is the median rate on the original scale. Thus, to calculate the percentage reduction in the establishment rate at the low level of the covariate is

$$\mathrm{PR}_0 \;=\; \left(1 - \frac{\exp(\gamma_3)}{\exp(\gamma_1)}\right) \times 100\%,$$

and at the high level of the covariate is

$$\mathrm{PR}_1 \;=\; \left(1 - \frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)}\right) \times 100\%.$$

We estimate these quantities from the data using the least squares estimates of γ_1, γ_2, γ_3, and γ_4 defined in the previous section. Then using the delta method (e.g., Casella and Berger, 2002, p.240), the standard error for our estimator of

$$\text{PR}_1 - \text{PR}_0 \quad = \quad 100 \left(\frac{\exp(\gamma_3)}{\exp(\gamma_1)} - \frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)} \right)$$

is

$$\sqrt{\boldsymbol{b}^T (\boldsymbol{D}^T \boldsymbol{D})^{-1} (\boldsymbol{D}^T \boldsymbol{\Sigma} \boldsymbol{D}) (\boldsymbol{D}^T \boldsymbol{D})^{-1} \boldsymbol{b}}$$

where

$$\boldsymbol{b} = 100 \left(-\frac{\exp(\gamma_3)}{\exp(\gamma_1)} + \frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)}, \frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)}, \frac{\exp(\gamma_3)}{\exp(\gamma_1)} - \frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)}, -\frac{\exp(\gamma_3 + \gamma_4)}{\exp(\gamma_1 + \gamma_2)} \right)^T .$$

Investigating weakened associations

To investigate the difference between associations pre- and post-policy, we consider two marginal models for the post and pre-counts.

As with the Adibe et al. (2019) we fit a negative binomial model to the pre-policy counts, assuming a working covariance of independence between the counts. We also fit a negative binomial model to the post-policy counts, making the same working independence covariance assumption. Again we use a sandwich estimator to correct the covariance of the model coefficient estimator, but now account for cross-dependence between the spatial counts pre- and post-policy.

In the negative binomial model for the pre-counts we assume $E(Y_{1i}) = \mu_{1i}$, where for a set of covariates of length p, $\boldsymbol{x}_{1i}$ and coefficients $\boldsymbol{\beta}_1$,

$$\eta_{1i} = \log(\mu_{1i}) = \log(P_i) + \boldsymbol{x}_{1i}^T \boldsymbol{\beta}_1.$$

Similarly for the post-counts we assume $E(Y_{2i}) = \mu_{2i}$ with

$$\eta_{2i} = \log(\mu_{2i}) = \log(P_i) + \boldsymbol{x}_{2i}^T \boldsymbol{\beta}_2.$$

Allowing for potentially different over-dispersion in the pre- and post-policy counts our model for the variance is

$$\mathrm{var}(Y_{ki}) = V_k(\mu_{ki}) = \mu_{ki} + \frac{\mu_{ki}^2}{\theta_k},$$

where $\theta_k > 0$ are the pre- ($k = 1$) and post-policy ($k = 2$) dispersion parameters. To account for the spatial dependence for the pre- and post-policy counts over the census tracts we assume for $k = 1, 2$, $k' = 1, 2$, $i = 1, \ldots, m$, and $j = 1, \ldots, m$,

$$\mathrm{cov}(Y_{ki}, Y_{k'j}) \;\; = \;\; \begin{cases} \sqrt{V_k(\mu_{ki})V_{k'}(\mu_{kj})}\; \rho_{ij}, & k = k'; \\ \sqrt{V_k(\mu_{ki})V_{k'}(\mu_{k'j})}\; \lambda\, \rho_{ij}, & k \neq k', \end{cases}$$

where ρ_{ij} parameterizes the correlation in the counts between tracts i and j. Our model for ρ_{ij} is the same as for Adibe et al. (2019). We estimate the parameter spatial dependence parameter that appears in the model for ρ_{ij} and the correlation parameter $-1 < \lambda < 1$ using maximum likelihood estimation based on the Pearson residuals calculated from both the pre- and post-policy counts.

For $k = 1, 2$, let $\boldsymbol{X}_k$ denote an $m \times p$ design matrix with ith row $\boldsymbol{x}_{ki}$ and let $\boldsymbol{V}_k$ denote a diagonal matrix with (i, i) element $V_k(\mu_{ki})$. Let $\boldsymbol{G}_k$ be an $m \times p$ matrix with (i, j) element

$$[G_k]_{ij} \;\; = \;\; \frac{\partial \mu_{ki}}{\partial \beta_j} \;\; = \;\; \frac{\partial \mu_{ki}}{\partial \eta_{ki}} \frac{\partial \eta_{ki}}{\partial \beta_j} \;\; = \;\; \mu_{ki}\, [x_{ki}]_j,$$

since

$$\frac{\partial \mu_{ki}}{\partial \eta_{ki}} \;\; = \;\; \left[\frac{\partial \eta_{ki}}{\partial \mu_{ki}}\right]^{-1} \;\; = \;\; \left[\frac{\partial \log(\mu_{ki})}{\partial \mu_{ki}}\right]^{-1} \;\; = \;\; \mu_{ki}.$$

Then the sandwich estimator of the covariance of $\widehat{\boldsymbol{\beta}}_k$ is

$$\mathrm{cov}(\widehat{\boldsymbol{\beta}}_k) \;\; = \;\; (\boldsymbol{G}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{G}_k)^{-1} \boldsymbol{G}_k^T \boldsymbol{V}_k^{-1} \mathrm{cov}(\boldsymbol{Y}_k) \boldsymbol{V}_k^{-1} \boldsymbol{G}_k (\boldsymbol{G}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{G}_k)^{-1}, \quad k = 1, 2,$$

with cross-covariance

$$\mathrm{cov}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2) \;\; = \;\; (\boldsymbol{G}_1^T \boldsymbol{V}_1^{-1} \boldsymbol{G}_1)^{-1} \boldsymbol{G}_1^T \boldsymbol{V}_1^{-1} \mathrm{cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2) \boldsymbol{V}_2^{-1} \boldsymbol{G}_2 (\boldsymbol{G}_2^T \boldsymbol{V}_2^{-1} \boldsymbol{G}_2)^{-1}.$$

Then our sandwich estimator of the covariance of the estimated model coefficients can be written as

$$\widehat{\text{cov}}(\widehat{\boldsymbol{\beta}}_k) \;\; = \;\; \boldsymbol{J}_k^{-1} \boldsymbol{B}_k^T (\boldsymbol{C} - \widehat{\alpha} \boldsymbol{W}) \boldsymbol{B}_k \boldsymbol{J}_k^{-1}, \quad k = 1, 2,$$

with

$$\widehat{\text{cov}}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2) \;\; = \;\; \lambda \, \boldsymbol{J}_1^{-1} \boldsymbol{B}_1^T (\boldsymbol{C} - \widehat{\alpha} \boldsymbol{W}) \boldsymbol{B}_2 \boldsymbol{J}_2^{-1},$$

where here $\boldsymbol{J}_k = \boldsymbol{G}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{G}_k$ and $\boldsymbol{B}_k = \text{diag}\Big(\widehat{\mu}_{ki} / \sqrt{V_k(\widehat{\mu}_{ki})} \Big) \, \boldsymbol{X}$.

References

Adibe, C., Craigmile, P. F., Onnen, N., Schwartz, E., and Roberts, M. E. (2019). The relationship between tobacco retailer density and neighborhood demographics in Ohio. *Ohio Journal of Public Health*, 4.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, FL.

Casella, G. and Berger, R. L. (2002). *Statistical Inference, Second Edition*. Duxbury, Pacific Grove, CA.