Online Appendix
Assessing the Incidence and Efficiency of a Prominent Place Based Policy

By MATIAS BUSSO, JESSE GREGORY, AND PATRICK KLINE*

This document is a Supplemental Online Appendix of "Assessing the Incidence and Efficiency of a Prominent Place Based Policy" published in the American Economic Review Vol. XX No. YY.

A. *Model Extension with Two Types of Workers*

Let a fixed proportion π_S of the agents be skilled and more productive than their unskilled counterparts who constitute the remaining fraction $\pi_U = 1-\pi_S$ of the population. Write the utility of individual i of skill group $g \in \{S, U\}$ living in community $j \in N$ and working in community $k \in \{\emptyset, \mathcal{N}\}$ and sector $s \in \{1, 2\}$ as:

$$\begin{aligned} u^g_{ijks} &= w^g_{jks} - r_j - \kappa_{jk} + A_j + \varepsilon^g_{ijks} \\ &= v^g_{jks} + \varepsilon^g_{ijks} \end{aligned}$$

where w^g_{jks} is the wage a worker of skill group g from neighborhood j receives when working in sector s of neighborhood k. Define a set of indicator variables $\left\{D^g_{ijks}\right\}$ equal to one if and only if $\max_{j'k's'}\left\{u^g_{ij'k's'}\right\} = u^g_{ijks}$ for worker i and denote the measure of agents of skill group g in each residential/work location by $N^g_{jks} = P\left(D^g_{ijks} = 1| \left\{v^g_{j'k's'}\right\}\right)$.

Suppose that skilled and unskilled workers are perfect substitutes in production so that firm output may be written $B_k\left(qS_{ks} + U_{ks}\right) f\left(\chi_{ks}\right)$ where the S_{ks} and U_{ks} refer to total skilled and unskilled labor inputs respectively, $\chi_{ks} = \frac{K_{ks}}{B_k(qS_{ks}+U_{ks})}$ is the capital to effective labor ratio, and q is the relative efficiency of skilled labor. Now wages will obey

$$\begin{aligned} B_k\left[f\left(\chi_{ks}\right) - \chi_{ks} f'\left(\chi_{ks}\right)\right] &= w^U_{jks}\left(1 - \tau\delta_{jks}\right) \\ w^S_{jks} &= q w^U_{jks} \\ f'\left(\chi_{ks}\right) &= \rho \end{aligned}$$

* Busso: Research Department, Inter-American Development Bank, 1300 New York Ave. NW, Washington, DC 20577 (email: mbusso@iadb.org). Gregory: Department of Economics, University of Wisconsin-Madison, 1180 Observatory Dr., Madison, WI 53706-1393, (email: jmgregory@ssc.wisc.edu). Kline: Department of Economics, UC Berkeley and NBER, 530 Evans Hall 3880, Berkeley, CA 94720-3880 (email: pkline@econ.berkeley.edu).

where w^U_{jks} is the wage for unskilled workers and w^S_{jks} the wage for skilled workers. Note that

$$\frac{d\ln w^U_{jks}}{d\ln B_k} = \frac{d\ln w^S_{jks}}{d\ln B_k} = 1$$

so that productivity increases may still be detected by examining impacts on the wages of commuters. However, productivity effects may also shift the skill composition of local workers and commuters which could lead us to over or understate these effects. For this reason we adjust our wage impacts in the paper for observable skill characteristics.

Our final modification is that with two skill groups, clearing in the housing market requires:

$$H_j = \sum_g \pi_g \sum_k \sum_s N^g_{jks}$$

With these features in place the social welfare function may be written:

$$W = \sum_g \pi_g V^g + \sum_j \left[r_j H_j - \int_0^{H_j} G_j^{-1}(x)\,dx \right]$$

It is straightforward then to verify that for some community m :

$$\begin{aligned}
\left.\frac{d}{dB_m}W\right|_{\tau=0} &= \sum_g \pi_g \sum_j \sum_k \sum_s N^g_{jks} \frac{dw^g_{jks}}{dB_m} \\
&= \left[\pi_U N^U_{.m} + q\pi_S N^S_{.m}\right] R(\rho) \\
\left.\frac{d}{dA_m}W\right|_{\tau=0} &= N_{m.}
\end{aligned}$$

where $N^g_{.m} = \sum_j \sum_s N^g_{jms}$ and $N_{m.} = \sum_g \pi_g \sum_k \sum_s N^g_{mks}$. Furthermore we may write the deadweight losses attributable to taxes as:

$$\begin{aligned}
DWL_\tau &= \sum_g \pi_g \sum_{j\in\mathcal{N}_1} \sum_{k\in\mathcal{N}_1} N^g_{jk1} w^g_{jk1} \int_0^{d\tau} t \frac{d\ln N^g_{jk1}}{dt} dt \\
&\approx \frac{1}{2}\psi d\tau^2 \sum_g \pi_g \sum_{j\in\mathcal{N}_1} \sum_{k\in\mathcal{N}_1} N^g_{jk1} w^g_{jk1}
\end{aligned}$$

where in the second line we have assumed a constant semi-elasticity of local employment $\psi = \frac{d\ln N^g_{jk1}}{d\tau}$. This formula is effectively the same as that in equation (10) of the paper, relying on the total covered wage bill and the elasticity ψ. Were the elasticity to vary by type we would simply need to compute the deadweight loss separately within skill group and average across groups using the marginal frequencies π_g. Finally, we may write the deadweight attributable to the block grants as:

$$\begin{aligned} DWL_G &\approx C\left[1-\lambda_a \sum_{j\in\mathcal{N}_1} \left.\frac{dW}{d\ln A_j}\right|_{\tau=0} - \lambda_b \sum_{k\in\mathcal{N}_1} \left.\frac{dW}{d\ln B_k}\right|_{\tau=0}\right] \\ &= C\left[1-\lambda_a \sum_{j\in\mathcal{N}_1} A_j N_{j.} - \lambda_b \sum_g \pi_g \sum_j \sum_{k\in\mathcal{N}_1} \sum_s N^g_{jks} w^g_{jks}\right] \end{aligned}$$

As before, the deadweight loss computation relies on the parameters λ_a and λ_b. Heterogeneity provides no essential complication to the exercise since, with knowledge of these parameters, one only needs to know the total wage bill and population inside of the zone to compute DWL_G.

B. Monte Carlo Experiments

We simulated hierarchical datasets of 64 zones with a random number of tracts N_z within each zone. The number of tracts per zone was generated according to $N_z = 10+\widetilde{\eta}_z$ where $\widetilde{\eta}_z$ is a Negative Binomial distributed random variable with the first two moments matching the ones observed in the data (i.e. a mean 21 and a standard deviation of 16 tracts). Hence, each simulated sample was expected to yield approximately 1,344 census tracts with no zone containing less than 10 tracts in any draw.

Outcomes were generated according to the model:

$$Y_{tz} = \beta_z T_z + \alpha^x_t X_{tz} + \alpha^p_z P_z + \xi_z + e_{tz}$$

where T_z is an EZ assignment dummy, X_{tz} is a tract level regressor, P_z is a zone level regressor, ξ_z a random zone effect, and e_{tz} an idiosyncratic tract level error. We assume throughout that:

$$\begin{bmatrix} X_{tz} \\ P_z \\ e_{tz} \end{bmatrix} \sim N\left(0, I_3\right)$$

To build in some correlation between the covariates and EZ designation, and to reflect the fact that treated zones tend to be larger, we model the EZ assignment mechanism as:

$$\begin{aligned} T_z &= I\left(rank\left(T^*_z\right) \leq 6\right) \\ T^*_z &= \overline{X}_z + P_z + 0.008 \times N_z + u_z \\ u_z &\sim N\left(0, 1\right) \end{aligned} \tag{1}$$

where $\overline{X}_z = \frac{1}{N_z}\sum_{t\in z} X_{tz}$ and the $rank\left(.\right)$ function ranks the T^*_z in descending order. Note that this assignment process imposes that exactly six zones will be treated. Hence, each simulation sample will face the inference challenges present in our data.

The nature of the coefficients $\left(\beta_z, \alpha_t^x, \alpha_z^p\right)$ and the random effect ξ_z vary across our Monte Carlo designs as described in the following table. We have two sets of results. In the first set, which we label symmetric, ξ_z follows a normal distribution. In a second set of results, which we label asymmetric, ξ_z follows a χ^2 distribution.

Table S1: Data Generating Processes

	Symmetric $\xi_z \sim N(0,1)$	Asymmetric $\xi_z \sim \chi^2(4)$
1. Baseline	$\beta_z = 0, \alpha_t^x = \alpha_z^p = 1$	$\beta_z = 0, \alpha_t^x = \alpha_z^p = 1$
2. Random Coefficient on X_{tz}	Same as 1) but, $\alpha_t^x \sim N(1,1)$	Same as 1) but, $(\alpha_t^x + 3) \sim \chi^2(4)$
3. Random Coefficient on P_z	Same as 1) but, $\alpha_z^p \sim N(1,1)$	Same as 1) but, $(\alpha_z^p + 3) \sim \chi^2(4)$
4. Random Coefficient on T_z	Same as 1) but, $\beta_z \sim N(0,1)$	Same as 1) but, $(\beta_z + 4) \sim \chi^2(4)$
5. All deviations from baseline	(2) + (3) + (4)	(2) + (3) + (4)

Note that the null of zero average treatment effect among the treated is satisfied in each simulation design. Specification 1) corresponds to the relatively benign case where our regression model is properly specified and the errors are homoscedastic. Specification 2) allows for heteroscedasticity with respect to the tract level regressor, while specification 3) allows some heteroscedasticity in the zone level regressor. Specification 4) allows heteroscedasticity with respect to the treatment, or alternatively, a heterogeneous but mean zero treatment effect. Specification 5) combines all of these complications so that heteroscedasticity exists with respect to all of the regressors.

For each Monte Carlo design we compute three sets of tests of the true null that EZ designation had no average effect on treated tracts. The first (analytical) uses our analytical cluster-robust standard error to construct a test statistic $\widehat{t} = \left|\frac{\widehat{\beta}}{\widehat{\sigma}}\right|$ where and rejects when $\widehat{t} > 1.96$. The second (wild bootstrap-se) uses a clustered wild bootstrap procedure to construct a bootstrap standard error σ^* and rejects when $\left|\frac{\widehat{\beta}}{\sigma^*}\right| > 1.96$. The third approach (wild bootstrap-t) estimates the wild bootstrap distribution $F_t^*(.)$ of the test statistic $\widehat{t} = \left|\frac{\widehat{\beta}}{\widehat{\sigma}}\right|$ and rejects when $\widehat{t} > F_t^{*-1}(0.95)$ – where $F_t^{*-1}(0.95)$ denotes the 95th percentile of the bootstrap distribution of t statistics. Both the bootstrap-se and bootstrap-t procedures simulate the bootstrap distribution imposing the null that $\beta = 0$ as recommended by Cameron, Gelbach, and Miller (2008). The false rejection rates for these three tests in each of the five simulation designs are given in the table below.

Table S2: False Rejection Rates in Monte Carlo Simulations

Tract-level models

	Analytical s.e. OLS	Analytical s.e. PW	Wild BS-s.e. OLS	Wild BS-s.e. PW	Wild BS-t OLS	Wild BS-t PW
Symmetric						
Baseline	0.126	0.074	0.039	0.111	0.054	0.053
Random Coefficient on X_{tz}	0.125	0.075	0.036	0.113	0.056	0.051
Random Coefficient on P_z	0.124	0.077	0.041	0.110	0.055	0.048
Random Coefficient on T_z	0.123	0.073	0.041	0.110	0.055	0.053
All	0.124	0.080	0.042	0.110	0.059	0.051
Asymmetric						
Baseline	0.123	0.106	0.037	0.138	0.055	0.056
Random Coefficient on X_{tz}	0.121	0.109	0.041	0.136	0.047	0.049
Random Coefficient on P_z	0.123	0.111	0.039	0.139	0.054	0.052
Random Coefficient on T_z	0.132	0.111	0.039	0.142	0.053	0.056
All	0.125	0.111	0.038	0.128	0.051	0.051

Standard error based methods tend to overreject in both designs save for in the case of OLS where the wild bootstrapped standard errors perform well. However the wild bootstrapped-t procedure yields extremely accurate inferences for both the OLS and PW estimators across all designs.