

NBER WORKING PAPER SERIES

REGIONAL DATA IN MACROECONOMICS:
SOME ADVICE FOR PRACTITIONERS

Gabriel Chodorow-Reich

Working Paper 26501
<http://www.nber.org/papers/w26501>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019

I thank Tobias Berg, Bill Dupor, Loukas Karabarbounis, Thuy Lan Nguyen (discussant), Valerie Ramey, David Romer, Ludwig Straub, and the participants at the St Louis Fed-JEDC-SCG-SNB-UniBern Conference and the Miami Empirical Macroeconomics Workshop for helpful comments. Richard Sweeney provided excellent research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Gabriel Chodorow-Reich. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Regional Data in Macroeconomics: Some Advice for Practitioners
Gabriel Chodorow-Reich
NBER Working Paper No. 26501
November 2019
JEL No. E0,R0

ABSTRACT

Cross-sectional or panel studies have joined time series techniques as an important element in empirical macroeconomists' toolkit. The econometric best practices for these studies and their aggregate implications remain active topics of research. In this paper, I offer several pieces of advice for practitioners in this literature. I begin by casting regional analysis in a Rubin (1978) potential outcomes framework. This formalism clarifies three reasons why the estimated impact of a shock on a single region can differ from the aggregate effect of the shock: (i) contamination of the untreated areas through "micro" spillovers, (ii) these spillovers sum to an economically relevant magnitude, and (iii) national variables endogenously respond to national shocks but not to local shocks. I provide several examples to illustrate and discuss how economic theory can sometimes sign the spillovers and bound the difference between the regional and aggregate effects of the shock. I then turn to econometric issues including the choice of endogenous variable in a regional regression and whether or not to weight by population.

Gabriel Chodorow-Reich
Department of Economics
Harvard University
1805 Littauer Center
Cambridge, MA 02138
and NBER
chodorowreich@fas.harvard.edu

1 Introduction

Analysis of regional shocks and outcomes has joined time series techniques as an important element in empirical macroeconomists’ toolkit. Appendix A lists 49 studies published in general interest or macroeconomic field journals between 2012 and 2018 that use regional data to answer macroeconomic questions. The topics include estimation of fiscal multipliers, the impact of foreign trade on U.S. labor markets, the role of household deleveraging in the Great Recession, and the impact of bank credit on the real economy, to name a few. Yet, the econometric best practices for regional analysis and the aggregate implications of these studies remain active topics of research. In this article, I offer several pieces of advice for practitioners in this literature.

Each section of the article addresses a different aspect of regional data. Section 2 begins by casting regional analysis in a Rubin (1978) potential outcomes framework. Unlike in the canonical setting, in regional analysis the potential outcome in each region can depend on the treatments or shocks in all other regions, and national variables can react to treatment applied to all regions. These differences determine relationships among several objects of interest to macroeconomists: the aggregate impact of an aggregate shock (β^{agg}), the economy-wide impact of a local shock ($\beta^{\text{all regions}}$), the impact of a local shock on the treated region only (β^{micro}), and the difference-in-difference estimator commonly used in practice ($\hat{\beta}^{\text{DiD}}$).

The possible difference between $\hat{\beta}^{\text{DiD}}$ and β^{micro} arises because the cross-sectional coefficient measures outcomes in the treated area relative to the “contaminated” untreated area. I refer to such contamination as a violation of the *micro* no-interference Stable Unit Treatment Value Assumption (SUTVA-micro). The wedge disappears as the treated area becomes infinitesimally-sized relative to the untreated area. Using the multi-region New

Keynesian model in Nakamura and Steinsson (2014), I show that in practical settings with geographic units of the size of U.S. states or smaller and demand shocks that do not induce factor mobility, SUTVA-micro violations should have minimal impact on regional estimates and usually may be safely ignored.

Section 3 addresses the difference between the the local impact of a local shock β^{micro} and the aggregate impact of an aggregate shock β^{agg} . Following arguments in Chodorow-Reich (2019), I first argue for comparing regional estimates to a judiciously chosen aggregate benchmark in which monetary policy and other aggregate variables do not respond to the shock. While restrictive, this benchmark can also facilitate comparison to macroeconomic theory by holding fixed factors auxiliary to the shock under study. Second, even when a shock to a single region causes a sufficiently *de minimis* spillover onto each other region as not to contaminate the estimation of the impact of the shock on the treated region, the spillovers may still aggregate to a magnitude that makes the economy-wide impact of the shock, $\beta^{\text{all regions}}$, different from the impact on the treated region only, β^{micro} . I state conditions under which these spillovers make regional estimates a bound for aggregate effects and argue that, where applicable, such a bounding exercise provides a transparent and almost model-free approach to relating β^{micro} to β^{agg} .

Sections 4 and 5 turn to common econometric issues that arise in the estimation of cross-regional regressions. Often in these settings, the exogenous variation affects a known subset of the total variation in the endogenous variable of interest. As one example, studies of regional fiscal multipliers often use excluded instruments that directly affect only part of government spending in an area. Section 4 presents Monte Carlo evidence that in these circumstances, standard instrumental variables estimation can perform poorly. Intuitively, if the excluded instruments explain only a small part of the endogenous variable, then conventional instrumental variables will suffer from a weak instruments problem. In

contrast, imposing the researcher’s knowledge that the exogenous variation affects only a subset of the total can reduce bias and increase efficiency. An application to the 2009 American Recovery and Reinvestment Act (ARRA, or the “Obama stimulus”) and a contrasting set of results reported in Chodorow-Reich (2019) and Ramey (2019) illustrates the importance of this distinction.

Section 5 addresses another common econometric question: should researchers weight regional estimates by population? In the presence of heterogeneous treatment effects, weighting by population can yield a regression coefficient more representative of the population average treatment effect than unweighted regression. However, it can also produce a coefficient less representative than unweighted regression. Moreover, weighting can decrease efficiency and, with instrumental variables, increase bias, especially with skewed weights in small samples. I again use Monte Carlo evidence to assess the practical importance of these considerations and to relate to the literature studying the 2009 ARRA.

I recommend reading section 2 before section 3, and section 4 before section 5. However, readers interested solely in the econometric issues may skip directly to section 4 with little loss of continuity.

2 Potential Outcomes and Spillovers

This section adapts the Rubin (1978) potential outcomes framework to regional analysis. The formalism focuses attention on three reasons why the coefficient from a typical regional difference-in-difference regression can differ from the aggregate response to an aggregate shock. First, the regional coefficient measures outcomes in the treated area relative to the “contaminated” untreated area, resulting in a possibly biased estimate of the impact of the shock on the treated area if “micro” spillovers occur. Second, the aggregate response also incorporates the impact on the untreated areas, or “macro” spillovers. Third, the

aggregate response also potentially reflects changes in variables that occur only at the national level, such as monetary policy.

Section 2.1 presents the potential outcomes framework. Section 2.2 defines the objects of interest in this framework. Section 2.3 illustrates using three economic examples.

2.1 Setup

The economy consists of N units, which I will refer to as “regions”. Let $W_{i,t}$ denote the treatment in region i and period t and $\mathbf{W}_t = \left(W_{1,t} \ \dots \ W_{N,t} \right)'$ the vector of treatments. For notational simplicity, I assume $W_{i,t}$ is a scalar (but not necessarily binary) and ignore lags. The observed outcome $Y_{i,t}^{\text{obs}}$ for region i depends on the vector of treatments and an aggregate treatment $W_{\text{agg},t}$:

$$Y_{i,t}^{\text{obs}} = Y_{i,t}(\mathbf{W}_t', W_{\text{agg},t}). \quad (1)$$

The potential outcome function $Y_{i,t}(\mathbf{W}_t', W_{\text{agg},t})$ defines the outcome for an arbitrary set of treatments, with the fundamental problem of causal identification being that only one realization for unit i is actually observed. To fix ideas, one may think of $W_{i,t}$ as government spending (net of taxes) in region i and $W_{\text{agg},t}$ as monetary policy. The dependence of the potential outcome function on the entire vector of regional treatments \mathbf{W}_t incorporates the possibility of spillovers across regions and differentiates the setup from the textbook case (see e.g. Imbens and Rubin, 2015).

Remark 1 *In the canonical potential outcomes framework, $W_{i,t}$ is binary and $Y_{i,t}$ depends only on $W_{i,t}$. Then the potential outcome function collapses to $Y_{i,t}^{\text{obs}} = Y_{i,t}(W_{i,t}) = Y_{i,t}(1)W_{i,t} + Y_{i,t}(0)(1 - W_{i,t})$.*

Remark 2 *The potential outcomes framework relates $Y_{i,t}$ directly to the treatments in other regions, rather than indirectly through linkages between $Y_{i,t}$ and $Y_{j,t}$ as in much of*

the spatial econometrics literature (see e.g. Anselin, 1988).

2.2 Objects of Interest

I now define four objects of interest. The first three objects characterize causal effects at different levels of aggregation:

$$\beta^{\text{agg}} = \sum_{j=1}^N \frac{Y_{j,t} \left((\mathbf{W}_t + \iota_N \Delta)' , \tilde{W}_{\text{agg},t} \right) - Y_{j,t} (\mathbf{W}_t', W_{\text{agg},t})}{N\Delta}, \quad (2)$$

$$\beta^{\text{all regions}} = \sum_{j=1}^N \frac{Y_{j,t} \left(W_{1,t} \dots \tilde{W}_{i,t} \dots W_{N,t}, W_{\text{agg},t} \right) - Y_{j,t} \left(W_{1,t} \dots W_{i,t} \dots W_{N,t}, W_{\text{agg},t} \right)}{\Delta}, \quad (3)$$

$$\beta^{\text{micro}} = \frac{Y_{i,t} \left(W_{1,t} \dots \tilde{W}_{i,t} \dots W_{N,t}, W_{\text{agg},t} \right) - Y_{i,t} \left(W_{1,t} \dots W_{i,t} \dots W_{N,t}, W_{\text{agg},t} \right)}{\Delta}, \quad (4)$$

where $\Delta \equiv \tilde{W}_{i,t} - W_{i,t}$ is the difference in “dosage” between two possible sets of treatments and ι_N is a vector of ones of length N . The first object, β^{agg} , gives the causal effect of increasing the treatment in all regions symmetrically and is often the object of ultimate interest to macroeconomists. For example, if Δ represents an increase in government spending (net of taxes), then β^{agg} gives the aggregate government spending multiplier.

The second object, $\beta^{\text{all regions}}$, is the impact on the whole economy of varying the treatment in region i only, holding fixed the treatment status of all other regions and the aggregate treatment. My first result gives sufficient conditions under which $\beta^{\text{all regions}}$ equals β^{agg} :

Result 1 *With linear treatment effects and symmetry across regions, that is, if $Y_{i,t} = \alpha_i + \delta W_{i,t} + \gamma \sum_{j \neq i}^N W_{j,t} + \eta W_{\text{agg},t} \forall i$, then $\beta^{\text{all regions}}$ differs from β^{agg} only because of the*

change in $W_{agg,t}$.

For example, if $W_{agg,t}$ is aggregate monetary policy, then Result 1 equates $\beta^{\text{all regions}}$ to the aggregate government spending multiplier when monetary policy does not respond to the spending.

The third object, β^{micro} , is the causal effect of varying $W_{i,t}$ on $Y_{i,t}$, holding fixed the treatment status of all other regions and the aggregate treatment.¹ The difference between $\beta^{\text{all regions}}$ and β^{micro} stems from macro spillovers, i.e. from summing the impact over all other regions of varying the treatment in i only.

The fourth object, $\hat{\beta}^{\text{DiD}}$, is the empirical object typically estimated in regional analysis:

$$\hat{\beta}^{\text{DiD}} = \frac{[Y_{i,t}^{\text{obs}} - Y_{i,t-1}^{\text{obs}}] - \frac{1}{N-1} \sum_{j \neq i} [Y_{j,t}^{\text{obs}} - Y_{j,t-1}^{\text{obs}}]}{\Delta}, \quad (5)$$

where $W_{i,t} - W_{i,t-1} = \Delta$ and $W_{j,t} = W_{j,t-1} \forall j \neq i$. Thus, $\hat{\beta}^{\text{DiD}}$ is the difference-in-difference estimator for the effect on $Y_{i,t}$ of varying the treatment in i only, and is given by comparing the change in the observed outcome in i to the average change across all other regions. Clearly, for $\hat{\beta}^{\text{DiD}}$ to be a good estimator of β^{micro} , standard causal identification assumptions must apply, including unconfoundedness of treatment (the potential outcome function remains stable between $t - 1$ and t across regions) and separability in the potential outcome function between the vector of regional treatments \mathbf{W}_t and the (possibly changing) aggregate treatment $W_{agg,t}$. In addition, the difference-in-difference approach to estimating β^{micro} requires an assumption of *de minimis* spillovers on each individual unit $j \neq i$, since $\hat{\beta}^{\text{DiD}}$ measures the change in outcomes in i relative to the average change in outcomes in other regions. Formally:

¹Sarto (2019) refers to β^{micro} as the “micro-local” elasticity.

Result 2 Assume standard causal identification conditions apply. Then:

$$|\hat{\beta}^{DiD} - \beta^{micro}| = \frac{1}{(N-1)\Delta} \left| \sum_{j \neq i}^N Y_{j,t} \left(W_{1,t} \dots \tilde{W}_{i,t} \dots W_{N,t}, W_{agg,t} \right) - \sum_{j \neq i}^N Y_{j,t} \left(W_{1,t} \dots W_{i,t} \dots W_{N,t}, W_{agg,t} \right) \right|.$$

According to Result 2, $\hat{\beta}^{DiD} \rightarrow \beta^{micro}$ as the average impact of the treatment on the control group, $\frac{1}{N-1} \sum_{j \neq i}^N dY_{j,t}^{obs}/dW_{i,t}$, goes to zero. This condition closely resembles the textbook no-interference Stable Unit Treatment Value Assumption (SUTVA), which states that “the potential outcomes for any unit do not vary with the treatments assigned to other units” (Imbens and Rubin, 2015, p. 10). For example, a SUTVA violation in a clinical trial occurs if administering a drug to Jane affects Joe’s blood pressure, where Joe is in the control group. Result 2 differs in that it requires only *de minimis* spillovers on each region j , so that the *average* spillover is small. *Summing* over regions j , as in equation (3) for $\beta^{all\ regions}$, these *de minimis* regional spillovers may aggregate to economically relevant macro spillovers. In particular, as $N \rightarrow \infty$, $\hat{\beta}^{DiD} \rightarrow \beta^{micro}$ but β^{micro} is not necessarily equal to $\beta^{all\ regions}$. I will refer to cases where $\hat{\beta}^{DiD}$ differs from β^{micro} as violations of SUTVA-micro.

I can now restate formally the three issues discussed at the outset. The problem of spillovers causing a biased estimate of the impact of a shock on the treated region concerns the case where a violation of SUTVA-micro results in $\hat{\beta}^{DiD}$ yielding a biased estimate of β^{micro} . Even when SUTVA-micro holds, β^{micro} can differ from $\beta^{all\ regions}$ because of macro spillovers to other regions. Finally, $\beta^{all\ regions}$ may differ from β^{agg} because treating all regions can induce an endogenous response of $W_{agg,t}$ that does not occur when only a single region receives treatment.

The remainder of this section illustrates these issues in the context of three economic examples. The first two examples permit analytical characterizations of the micro and macro spillovers, while the third provides a numeric evaluation of the importance of violations of SUTVA-micro in practice. The next section focuses on how to relate β^{micro} to β^{agg} .

2.3 Examples

Example I. Allocation of labor with production subsidies. A simple economic example concerns the allocation of labor across regions in the presence of production subsidies. Suppose a producer in each region $i = 1, 2, \dots, N$ has a revenue function $y_i = (1 + \tau_i) L_i^{1-\frac{1}{\gamma}}$, where y_i denotes total revenue, τ_i is a production subsidy, and L_i is labor input. Labor is perfectly mobile across regions but fixed in the aggregate and perfectly inelastically supplied at a quantity normalized to one. With a competitive factor market, the allocation of labor satisfies:²

$$L_i^{\text{obs}} = L_i \left(\tau_1 \quad \dots \quad \tau_N \right) = \frac{(1 + \tau_i)^\gamma}{\sum_{j=1}^N (1 + \tau_j)^\gamma}. \quad (6)$$

Starting from a symmetric equilibrium with $\tau_j = 0$ for all j , consider estimation of the semi-elasticity of the employment response to introducing a subsidy $\tau_i > 0$ in region i only:

$$\beta^{\text{micro}} = \frac{d \ln L_i}{d \tau_i} = \left(\frac{N-1}{N} \right) \gamma. \quad (7)$$

²Equating the marginal product to the wage in each region and solving for labor, $\left(1 - \frac{1}{\gamma}\right) (1 + \tau_i) L_i^{-\frac{1}{\gamma}} = W \Rightarrow L_i = \left(\frac{(1-\frac{1}{\gamma})(1+\tau_i)}{W} \right)^\gamma$. Summing over regions, equating to the aggregate population level of one, and solving for W yields $W = \left(1 - \frac{1}{\gamma}\right) \left(\sum_{j=1}^N (1 + \tau_j)^\gamma \right)^{\frac{1}{\gamma}}$. Substituting this expression back into the earlier one and solving for L_i gives equation (6) in the main text. This example also implicitly assumes lump-sum taxation to finance the subsidies.

The difference-in-difference estimator $\frac{d \ln L_i}{d \tau_i} - \frac{1}{N-1} \sum_{j \neq i}^N \frac{d \ln L_j}{d \tau_i}$ would instead yield $\hat{\beta}^{\text{DiD}} = \gamma$. The difference between this value and β^{micro} stems from the violation of SUTVA-micro, $d \ln L_j / d \tau_i = -\gamma / N < 0$.

With infinitesimal areas ($N \rightarrow \infty$), the SUTVA-micro violation disappears ($d \ln L_j / d \tau_i \rightarrow 0$) and $\hat{\beta}^{\text{DiD}} = \beta^{\text{micro}}$. Intuitively, a shock to an infinitesimal area has no measurable impact on any other single area. Yet, β^{micro} still does not equal the response of aggregate employment to the policy change:

$$\beta^{\text{all regions}} = \frac{d \ln \left(\sum_{j=1}^N L_j \right)}{d \tau_i} = \left(\frac{N-1}{N} \right) \gamma - (N-1) \left(\frac{1}{N} \right) \gamma = 0. \quad (8)$$

Mechanically, the difference between β^{micro} and $\beta^{\text{all regions}}$ occurs because the summing of spillovers that are *de minimis* in each region results in an aggregate spillover of the same order of magnitude as the response in the treated region i . Economically, the difference occurs because labor supply to a single region is perfectly elastic while aggregate labor supply is perfectly inelastic.

Example II. Old Keynesian government spending multiplier. My next example offers an analytical characterization of spillovers and the SUTVA-micro violation in the context of the government spending multiplier with “Old Keynesian” rule-of-thumb agents. The economy again consists of N regions, each with fixed size $1/N$ (no inter-regional migration, unlike the example above). Let $\mathbf{c} = \left(c_1 \dots c_N \right)'$ denote the vector of consumption expenditures in each region, $\mathbf{y} = \left(y_1 \dots y_N \right)'$ the vector of outputs, and $\mathbf{g} = \left(g_1 \dots g_N \right)'$ the vector of government purchases, where each variable c_i, y_i, g_i is the (level) deviation from its steady state value. A representative agent in each region allocates $1 - \alpha$ of her expenditure to locally-produced output and $\frac{\alpha}{N-1}$ of her expenditure

to output produced in each other region. Market clearing then requires:

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \mathbf{g}, \quad (9)$$

where:

$$\mathbf{A} = \begin{pmatrix} 1 - \alpha & \frac{\alpha}{N-1} & \cdots & \frac{\alpha}{N-1} \\ \frac{\alpha}{N-1} & 1 - \alpha & \frac{\alpha}{N-1} \cdots & \frac{\alpha}{N-1} \\ \vdots & \frac{\alpha}{N-1} & \ddots & \vdots \\ \frac{\alpha}{N-1} & \cdots & \frac{\alpha}{N-1} & 1 - \alpha \end{pmatrix}.$$

The agent also has a marginal propensity to consume out of income of ρ :

$$\mathbf{c} = \rho\mathbf{y}. \quad (10)$$

Substituting equation (10) into equation (9):

$$\mathbf{y} = \rho\mathbf{A}\mathbf{y} + \mathbf{g} = \mathbf{B}\mathbf{g}, \quad (11)$$

where:

$$\mathbf{B} = (\mathbf{I} - \rho\mathbf{A})^{-1}$$

and element i, j of \mathbf{B} is:

$$b_{i,j} = \begin{cases} \frac{1}{m} + \frac{1}{Nm - \frac{N-1}{\alpha\rho}m^2}, & i = j, \\ \frac{1}{Nm - \frac{N-1}{\alpha\rho}m^2}, & i \neq j, \end{cases} \quad (12)$$

where:

$$m = 1 - \rho(1 - \alpha) + \frac{\alpha\rho}{N-1}.$$

Thus, $y_i^{\text{obs}} = y_i \begin{pmatrix} g_1 & \cdots & g_N \end{pmatrix} = \sum_{j=1}^N b_{i,j} g_j$.

As $N \rightarrow \infty$, the off-diagonal elements of \mathbf{B} converge to zero while the diagonal elements converge to the ‘‘Old Keynesian’’ open economy government spending multiplier, $\beta^{\text{micro}} = \frac{1}{1 - \rho(1 - \alpha)}$. Additionally, since the off-diagonal elements converge to zero, SUTVA-micro holds and the difference-in-difference estimator $\hat{\beta}^{\text{DiD}}$ converges to β^{micro} .

The column sum of \mathbf{B} gives the impact of local expenditure on the aggregate economy: $\beta^{\text{all regions}} = \sum_{i=1}^N b_{i,j} = \frac{1}{1-\rho}$, which also equals the “Old Keynesian” closed economy government spending multiplier β^{agg} , since, intuitively, the aggregate response occurs in a closed economy. Once again, $\beta^{\text{all regions}}$ differs from β^{micro} because the cumulation of spillovers that are *de minimis* in each region results in an aggregate spillover of the same order of magnitude as the response in the treated region i .

Example III. New Keynesian government spending multiplier. In practice, geographic units of analysis are not infinitesimal, making SUTVA-micro violations potentially quantitatively relevant. My third example assesses the magnitude of these violations using a benchmark multi-region New Keynesian model developed in Nakamura and Steinsson (2014). I briefly describe the setup and refer the reader to their article for the equations.

The model economy consists of two regions, H (for “Home”) and F (for “Foreign”).³ The regions differ only in population size, which remain fixed. Production of constant elasticity of substitution (CES) varieties occurs in each region, with consumers exhibiting “home bias” in preferences for goods produced in their local region. Trade in varieties occurs frictionlessly across regions, such that a law-of-one-price holds for each variety. This price may exhibit stickiness due to Calvo adjustment frictions. A monetary authority sets the nominal interest rate, which is common across regions. I keep all parameters the same as in Nakamura and Steinsson (2014) except the population of H relative to the total economy, denoted by n .⁴

³One can reformulate what follows as an economy with N ex ante identical regions each with size $n = 1/N$, of which one region gets treated. By assumption, all variables in the $N - 1$ untreated regions evolve identically following the treatment. Therefore, they can be represented as a single region F . In particular, the spillover to each of the $N - 1$ untreated regions, measured on a per capita basis, equals the spillover to the combined region F .

⁴The authors have posted their code at <https://eml.berkeley.edu/~enakamura/papers/FiscalDataPrograms.zip>.

The model experiment I consider concerns the impact of government purchases on private spending. Specifically, I compute the impulse responses of private expenditure on output produced in H and F to an increase in government purchases in H . This experiment corresponds to an empirical setting in which the researcher has identified plausibly exogenous variation in local government spending. In econometric terms, H is the treated region, F is the untreated region, and the difference-in-difference estimator of β^{micro} is the change in private expenditure on output produced in H less the change in output in F .

Table 1 reports the 1 year horizon difference-in-difference estimator separately for separable and GHH (Greenwood, Hercowitz, and Huffman, 1988) preferences, spending financed locally and by the federal government, and four relative sizes of the home region. Separable preferences refer to how consumption and leisure enter the period utility function, while with GHH preferences consumption and leisure enter as substitutes. The first row reports the relative response of private expenditure when H is approximately infinitesimal ($n=0.001$). The increase in government purchases in H causes the price of output produced in H to rise, which in turn causes residents of both regions to expenditure-switch toward output produced in F . With separable preferences and locally-financed spending (column (1)), the private output multiplier is therefore negative. However, the magnitude of the crowd-out of private output is small; when government purchases of H output rise by 1 percentage point of steady state output, private purchases of H output fall by 0.14 percentage point.⁵ The second row increases n to 0.02, such that H is the average size of a U.S. state. The private output multiplier does not change up to the third decimal place. Rows 3 and 4 increase n to 0.1 and 0.2, respectively, still with very little change in the

⁵More precisely, government purchases rise by 1 percentage point in quarter 1 and subsequently decay with auto-regressive parameter 0.933. Thus, the annualized increase in government purchases in year 1 is $(1 - 0.93^4)/(1 - 0.93) \times (1/4) = 0.90$ percentage point of output, in response to which output purchased by private agents falls by 0.14 percentage point.

Table 1: Private Output Response to Government Spending by Region Size

Preferences:	Separable		GHH	
Financing:	Local	Federal	Local	Federal
	(1)	(2)	(3)	(4)
Region size n :				
0.001	-0.144	-0.019	0.568	0.931
0.020	-0.144	-0.023	0.554	0.904
0.100	-0.145	-0.041	0.493	0.787
0.200	-0.147	-0.063	0.410	0.640

Notes: The table reports the difference in output purchased by private agents in H and F in the year following an increase in government purchases of output in H , using the Nakamura and Steinsson (2014) model. In columns (1) and (2), the period utility function is separable between consumption and leisure, while in columns (3) and (4) the period utility function features substitutability between consumption and leisure. In columns (1) and (3), the increased government purchases are financed by lump-sum taxes on residents of H only, while in columns (2) and (4) they are financed by lump-sum taxes levied proportionally on residents of H and F .

relative response of private output across regions.

Column (2) changes the financing from lump-sum taxes on residents of H only to lump-sum taxes levied on residents of both H and F in proportion to their respective population sizes. Imposing some of the tax burden on residents of F reduces the difference in private output responses in H and F (the entries in column 2 are less negative than those in column 1), although the difference is small relative to the increase in government purchases. The role of region size increases with federal financing. However, region size now matters beyond just the SUTVA-micro violation. A federally-financed government purchase of H output can be split into a locally-financed purchase (the experiment in column (1)) and a transfer from the federal government to residents of H equal to the share of the tax burden borne by residents of F . The role of region size in column (2) therefore combines the SUTVA-micro violation in column (1), the impact of higher taxes in F to finance the transfer, and the fact that the magnitude of the transfer falls as n grows. Combining these effects, the crowd-out magnitudes with $n = 0.001$ and $n = 0.02$ do not differ up to

the second decimal place and the differences between $n = 0.001$ and even $n = 0.2$ remain small.

Columns (3) and (4) turn to GHH preferences. The preference substitutability between leisure and consumption causes H agents to increase their consumption when hours worked go up to produce the additional output purchased by the government. Because these agents have home bias for output produced in H , private expenditure on H output becomes positive. The importance of the SUTVA-micro violation remains quite small when $n = 0.02$, but becomes more notable with $n = 0.1$ and especially with $n = 0.2$.⁶ The differences in response to locally-financed and federally-financed increases in government purchases become larger with GHH preferences, as does the dependence of the private output response on region size. In the extreme, this response falls from 0.93 with $n = 0.001$ to 0.64 with $n = 0.2$.

Overall, a few patterns emerge. The importance of the SUTVA-micro violation (and the effect on the size of transfers in columns (2) and (4)) remains quantitatively small when $n = 0.02$. In practical terms, this result suggests that researchers may ignore the impact of SUTVA-micro violations on estimation of β^{micro} when using geographic data at the state level or finer to study demand shocks that do not induce factor reallocation.⁷ The importance of SUTVA-micro violations with larger units of analysis appears more sensitive to details of the economy.

⁶Auclert and Rognlie (2017) criticize the use of GHH preferences in the context of fiscal multipliers and, implicitly, of general demand shocks.

⁷Berg and Streitz (2019) introduce an important caveat to this conclusion — inclusion of region (or in a panel setting region×time) fixed effects at a coarser level than the unit of analysis shrinks the effective size of the control group if spillovers concentrate on adjacent units. As an extreme example, if all spillovers from treating region 1 fall on region 2, then including a fixed effect for belonging to region 1 or 2 increases the importance of the SUTVA-micro violation to the equivalent of $n = 1/2$ in the simulations table 1. Intuitively, in this case the combined entity of region 1 and region 2 behaves as a single closed economy. Berg and Streitz (2019) also discuss how to estimate SUTVA-micro violations in settings with random assignment of treatment across varying levels of geography.

3 Setting Boundaries

Even without SUTVA-micro violations, the local or cross-sectional effect may differ from the national effect, for two main reasons. First, national variables such as monetary policy may endogenously respond to national shocks but do not react to local shocks. In the potential outcomes notation, $W_{agg,t}$ may endogenously respond to a change in all elements of \mathbf{W} . Second, the impact of a shock to local area i on the rest of the economy may be of the same order of magnitude as the impact on i , so that $\beta^{\text{all regions}} \neq \beta^{\text{micro}}$.

The first difference is easily dealt with by comparing β^{micro} to the aggregate response without the endogenous reaction of national variables. That is, even if $W_{agg,t}$ endogenously reacts to the treatment, the researcher can always compare β^{micro} to the theoretical object $\beta^{\text{agg, fixed policy}}$, defined as:

$$\beta^{\text{agg, fixed policy}} = \sum_{j=1}^N \frac{Y_{j,t} ((\mathbf{W}_t + \iota_N \Delta)') - Y_{j,t} (\mathbf{W}_t', W_{agg,t})}{N \Delta}. \quad (13)$$

The expression in equation (13) differs from the definition of β^{agg} in equation (2) only in holding $W_{agg,t}$ fixed across the two treatment scenarios. For example, Chodorow-Reich (2019) emphasizes the comparison of cross-sectional fiscal spending multipliers to national multipliers when monetary policy does not react. A zero lower bound provides a leading example where this condition might occur; however, the condition is stronger, since it requires a constant nominal yield curve at all maturities. The comparison also must exclude the reaction of variables that do not vary with the regional variation in the shock, such as confidence in government following a policy change. As emphasized by Nakamura and Steinsson (2014) and Nakamura and Steinsson (2018), limiting the scope of the comparison can actually increase the informativeness for macroeconomic theory by holding fixed factors auxiliary to the shock under study.

The difference between β^{micro} and $\beta^{\text{all regions}}$, which arises because of the impact of a shock to a local area on the rest of the economy, creates more difficulty for drawing inference about aggregate effects from regional regressions.⁸ In effect, the magnitude of these spillovers constitutes a sufficient statistic for going from β^{micro} to $\beta^{\text{all regions}}$. Much extant research proceeds in one of two directions: (i) ignore general equilibrium and report only the “partial equilibrium” or “direct” impact β^{micro} ; or (ii) calculate the aggregate effect using a fully-specified general equilibrium model with multiple regions, partly disciplined by the empirical regional estimate of β^{micro} . Each of these approaches has drawbacks.

The “partial equilibrium” object on its own lacks grounding in economic theory, precisely because it offers no guidance on general equilibrium channels. Consider again the example of the allocation of labor across regions from the previous section. There the partial equilibrium local elasticity equals γ (in the limit as the SUTVA-micro violation disappears), while the economy-wide elasticity equals zero. Why? When local region i is small, it faces an (almost) perfectly elastic labor supply curve, drawing in labor from the rest of the economy when labor demand rises due to the production subsidy. In contrast, the aggregate economy faces a perfectly inelastic labor supply curve. Therefore, in general equilibrium the shock causes a wage increase just large enough to discourage hiring in the other regions by an amount equal to the additional hiring in the treated region. This general equilibrium price response explains the difference between the local and economy-wide effects.⁹

A fully specified model accounts for any general equilibrium channels. The results

⁸The careful reader will note that I have elided differences between $\beta^{\text{all regions}}$ and $\beta^{\text{agg.fixed policy}}$ that may arise if the conditions in Result 1 do not hold, for example because of heterogeneity across regions in their sensitivity to shocks or in inter-regional linkages. Adão, Arkolakis, and Esposito (2019) explore departures from these assumptions in a quantitative structural framework.

⁹Similar issues arise with aggregation of household or firm-level regressions. See Chodorow-Reich (2014, Appendix A), Chodorow-Reich and Falato (2019), Sraer and Thesmar (2018), Blattner, Farinha, and Rebelo (2019) and Baqaee and Farhi (2019) for alternative approaches to aggregation of firm-level estimates.

reported in table 1 offer an example of output from such a model.¹⁰ However, the aggregate response then depends heavily, and sometimes non-transparently, on the ingredients in the model as well as the particular parameterization.

Here I advocate a third way: in many cases, theory provides guidance on the sign of the macro spillovers. In these cases, β^{micro} provides a bound for $\beta^{\text{all regions}}$. Relative to simply reporting the partial equilibrium impact, this approach appropriately relates the partial equilibrium object to its general equilibrium counterpart. However, by avoiding the need to fully specify a structural model, it preserves much of the transparency and robustness of the direct approach.

The key insight is that the examples in the previous section generalize in important respects in providing guidance for such a bound. In general, with factors of production immobile across regions, demand shocks in one region have positive spillovers onto factor demand in other regions. The intuition (and formalism) follows standard trade and open economy theory (Nakamura and Steinsson, 2014; Farhi and Werning, 2016; Chodorow-Reich, 2019; Chodorow-Reich, Gopinath, et al., Forthcoming; Dupor, Karabarbounis, et al., 2018). Higher demand in region i raises local incomes and prices. Higher incomes increase consumption expenditure, some of which falls on tradable output produced in other regions. Higher prices cause expenditure-switching to other regions. Both forces raise output in other regions; $\beta^{\text{all regions}} > \beta^{\text{micro}}$ in the Old Keynesian (income channel) and New Keynesian (price channel) examples of the previous section.

In contrast, with mobile factors an increase in factor demand in region i raises factor prices and draws in factors from other regions. The movement of factors makes spillovers

¹⁰This approach has come to define the field of quantitative spatial economics (Redding and Rossi-Hansberg, 2017; Caliendo, Dvorkin, and Parro, 2019), although models in that tradition have yet to incorporate the nominal rigidities central to many macroeconomic applications. Chodorow-Reich and Wieland (Forthcoming) argue that nominal wage rigidity impedes factor reallocation, with implications for the transition paths in these models.

potentially negative, as in the labor allocation example of the previous section. Thus, this force operates in an opposite direction. However, in many circumstances the mobility of at least one important factor — labor — is directly quantifiable, since changes in population in local region i are often observable. When factors do not move in response to the shock, perhaps because of fixed costs of moving, the spillovers onto production in other regions from demand shocks are unambiguously positive. In that case, the partial equilibrium, direct output effect provides a lower bound for the aggregate output effect of the shock.

An important element of interpretation concerns the case where multiple shocks occur simultaneously. For example, the previous section simulated the response to an increase in government spending in region H financed by the federal government. This policy change combines two shocks, one the increase in government purchases of output from H , and the other the transfer from F to H to pay for the increase in spending. Applying the lower bound result for aggregate government spending (holding constant nominal interest rates) requires isolating the regional output response to the government spending shock only, since no simultaneous transfer of resources occurs at the national level. (In practice, Chodorow-Reich (2019) argues that for transitory increases in spending the output response associated with the transfer component is too small to undo the lower bound, but this is a quantitative statement.)

I summarize the preceding discussion with a claim:

Result 3 *When factors of production are immobile, a demand shock in one region causes positive spillovers to production in other regions. Therefore, regional analysis of output effects provides a lower bound for the aggregate effect of the demand shock, holding constant the reaction of variables that vary only at the national level such as nominal interest rates or confidence in the government.*

Remark 3 *The distinction between measuring regional output and absorption responses matters to this result. For example, in an economy with fixed regional endowments, a shock to consumption demand in a single region will cause negative consumption spillovers to other regions but zero production spillovers.*

Remark 4 *Even when factors are mobile, if their response can be quantified, then a lower bound result can be applied to regional estimates adjusted for the factor response.*

A number of other approaches also merit mention. Building on Mian and Sufi (2014), Chodorow-Reich, Nenov, and Simsek (2019) and Auclert, Dobbie, and Goldsmith-Pinkham (2019) show how estimation of local responses in the nontradable sector map into the aggregate, since, intuitively, in an aggregate closed economy all sectors are nontradable. Chodorow-Reich, Nenov, and Simsek (2019) and Guren et al. (in progress) relate regional estimates to household-level responses by establishing an equivalence between the local general equilibrium multiplier and the local fiscal multiplier. Beraja, Hurst, and Ospina (2016) and Sarto (2019) use regional elasticities as identification restrictions in aggregate simultaneous equation systems. Huber (2018), Berg and Streitz (2019), Auerbach, Gorodnichenko, and Murphy (2019), Dupor and McCrory (2018), Dupor and Guerrero (2017), McCrory (2019), and Conley et al. (2019) all propose different ways of estimating indirect effects of a shock directly. Hornbeck and Moretti (2019) use a spatial equilibrium structure in which local labor demand and housing supply elasticities are sufficient statistics for calculating spillovers.

4 Which Endogenous Variable?

I now turn to a few important issues that can arise in practical estimation of β^{micro} . In many settings, a researcher has excluded instruments that directly affect only a known

subset of the total variation in the endogenous variable. A leading example, which I use throughout this section, occurs when a subset of total government spending gets allocated randomly across space. Concretely, a number of articles reviewed in Chodorow-Reich (2019) exploit provisions of the 2009 fiscal stimulus, known as the American Recovery and Reinvestment Act (ARRA), that created plausibly exogenous variation in the spatial distribution of funds. While these provisions strongly predict ARRA spending, it may be natural to assume that they are orthogonal to the remainder of state and local (S&L) expenditure. In the extreme, the excluded instruments can be arbitrarily weak for predicting total expenditure, even though they are strong instruments for a particular subset. What specification should a researcher use in this case?

Formally, consider the system:

$$y_i = \beta x_i + u_i, \tag{14}$$

$$x_i = \sum_{\ell=1}^L z_i^\ell + e_i, \tag{15}$$

$$E[z_i^\ell u_i] = 0, \ell = 1, \dots, L, \tag{16}$$

$$\text{Corr}(u_i, e_i) = \rho. \tag{17}$$

Here y_i denotes the outcome for area $i = 1, 2, \dots, N$, x_i the endogenous variable, z_i^1, \dots, z_i^L are the excluded instruments, and β is the parameter of interest.¹¹ For example, if y_i is output growth and x_i is the change in government purchases as a share of output, then β is the impact fiscal spending multiplier. The parameter ρ governs the amount of endogeneity in the residual part of x_i . Let $\bar{z}_i = \sum_{\ell=1}^L z_i^\ell$ denote the sum of the excluded instruments. The econometrician observes y_i, x_i , each z_i^ℓ , and \bar{z}_i . In the ARRA example, one may think

¹¹For simplicity, the system omits any included instruments (i.e. covariates). All of what follows continues to hold if one re-interprets the variables as the part orthogonal to any included instruments. I also write β instead of β^{micro} in equation (14) and assume SUTVA-micro is satisfied.

of each z_i^ℓ as one excluded instrument and \bar{z}_i as total ARRA spending induced by these instruments, where for simplicity each z_i^ℓ enters with a coefficient of one. I assume for now that the econometrician knows the first stage coefficients with certainty. Finally, let bold letters denote matrices, i.e. $\mathbf{y} = \begin{pmatrix} y_1 & \dots & y_N \end{pmatrix}'$, and define $\mathbf{Z} = \begin{pmatrix} \mathbf{z}^1 & \dots & \mathbf{z}^L \end{pmatrix}$ and $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

I am interested in the properties of three estimators:

$$\text{OLS:} \quad \hat{\beta}^{ols} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}, \quad (18)$$

$$\text{IV-2sls:} \quad \hat{\beta}^{iv} = (\mathbf{x}'\mathbf{P}_Z\mathbf{x})^{-1} \mathbf{x}'\mathbf{P}_Z\mathbf{y}, \quad (19)$$

$$\text{Reduced form:} \quad \hat{\beta}^{rf} = (\bar{\mathbf{z}}'\bar{\mathbf{z}})^{-1} \bar{\mathbf{z}}'\mathbf{y}. \quad (20)$$

Clearly, $\hat{\beta}^{ols}$ is biased and inconsistent when $\rho \neq 0$. In fact, $\hat{\beta}^{ols} - \beta \rightarrow \frac{\sigma_e\sigma_u\rho}{L\sigma_z^2 + \sigma_e^2}$, where σ_v is the standard deviation of variable v . As is well known, $\hat{\beta}^{iv}$ is also biased and converges to $\hat{\beta}^{ols}$ when the explanatory power of the excluded instruments in the first stage approaches zero, as will occur when $\sigma_e/\sigma_z \rightarrow \infty$. Perhaps surprisingly, $\hat{\beta}^{rf}$ remains unbiased and consistent even in this case:

$$\hat{\beta}^{rf} = (\bar{\mathbf{z}}'\bar{\mathbf{z}})^{-1} \bar{\mathbf{z}}'(\mathbf{x}\beta + \mathbf{u}) = (\bar{\mathbf{z}}'\bar{\mathbf{z}})^{-1} \bar{\mathbf{z}}'((\bar{\mathbf{z}} + \mathbf{e})\beta + \mathbf{u}) = \beta + (\bar{\mathbf{z}}'\bar{\mathbf{z}})^{-1} \bar{\mathbf{z}}'\mathbf{u},$$

where $\bar{\mathbf{z}}'\mathbf{e} = 0$ by construction and $E[\bar{\mathbf{z}}'\mathbf{u}] = 0$ by equation (16). Thus, under the maintained assumptions, $\hat{\beta}^{rf}$ has better properties than $\hat{\beta}^{iv}$. The intuition is straightforward: weak instrument bias in IV occurs because of sampling uncertainty in the first stage. Hence, when the researcher has external information on the first stage coefficients — here because they affect a subset of spending dollar-for-dollar and do not impact the remainder — there are gains from imposing this information on the estimation.

I next perform Monte Carlo simulations to assess the importance of these properties in practice. I choose parameters to approximate the cross-state ARRA analysis in Chodorow-

Reich (2019). In particular, I set the parameter β to 2 (the results that follow scale proportionally with this parameter), the number of excluded instruments L to 3, and the bias parameter ρ to -0.8 . I normalize the standard deviation of each z_i to 1. As the value of σ_e governs the strength of the first stage, I vary that parameter and set $\sigma_u = -\beta\sigma_e/\rho$, which creates an OLS bias of $-\beta$ as $\sigma_e \rightarrow \infty$. For each value of σ_e , I randomly draw 500 samples of 50 observations each from a multivariate normal distribution.

Panel A of table 2 reports results for $\sigma_e = 1$. In this case, most (3/4) of the variance in x_i comes from the exogenous component \bar{z}_i . As a result, the bias in OLS is not that large ($plim\hat{\beta}^{ols} = 1.5$), the first stage IV remains strong (median F statistic of 49), and $\hat{\beta}^{iv}$ is centered around the true $\beta = 2$. Of course, imposing the true first stage coefficients still leads to efficiency gains with β^{rf} .

The contrast appears much starker in Panel B, where $\sigma_e = 5$. Now most (25/28) of the variance of x_i comes from the endogenous component e_i , making $\hat{\beta}^{ols}$ severely biased ($plim\hat{\beta}^{ols} = 0.22$). Moreover, the first stage is weak, with a median F statistic of 2.73. As a result, the median $\hat{\beta}^{iv}$ falls to 1.41. In contrast, $\hat{\beta}^{rf}$ remains mean and median unbiased.

So far I have assumed the first stage coefficients multiplying the z s are known. A small variant of this example, and one of practical import, occurs when the excluded instruments powerfully explain a subset of spending but with the coefficients unknown. Formally, replace equation (15) with:

$$x_i = \bar{z}_i + e_i, \tag{21}$$

$$\bar{z}_i = \sum_{\ell=1}^L \pi^\ell z_i^\ell + w_i, \tag{22}$$

where $E[z_i^\ell w_i] = 0 \forall \ell$. Equation (22) introduces additional, unobserved variation w_i into \bar{z}_i . For simplicity, this additional variation has the same correlation of ρ with the structural

Table 2: Monte Carlo with Weak Instruments

Parameter	Median	Mean	SD	P10	P90	Pooled	Pooled SE	ARRA data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: $\sigma_e = 1$:</i>								
$\hat{\beta}^{ols}$	1.50	1.50	0.17	1.30	1.71	1.50	0.01	-0.00
$\hat{\beta}^{iv}$	1.97	1.98	0.21	1.72	2.25	2.00	0.01	1.31
$\hat{\beta}^{rf}$	1.99	2.00	0.13	1.84	2.16	2.00	0.01	
1 st stage $F(\hat{\beta}^{iv})$	49.33	53.15	18.07	33.49	76.26			1.89
<i>Panel B: $\sigma_e = 5$:</i>								
$\hat{\beta}^{ols}$	0.21	0.22	0.22	-0.07	0.51	0.22	0.01	-0.00
$\hat{\beta}^{iv}$	1.41	1.54	0.91	0.63	2.57	1.98	0.04	1.31
$\hat{\beta}^{rf}$	1.97	2.01	0.63	1.20	2.82	2.01	0.03	
1 st stage $F(\hat{\beta}^{iv})$	2.73	3.18	2.09	1.01	6.18			1.89
<i>Panel C: $\sigma_e = 5, \sigma_w = 1.04$:</i>								
$\hat{\beta}^{ols}$	0.32	0.33	0.17	0.10	0.56	0.33	0.01	-0.00
$\hat{\beta}^{iv}$	1.36	1.46	0.89	0.68	2.40	1.98	0.04	1.31
$\hat{\beta}^{niv}$	1.95	1.98	0.63	1.19	2.82	2.01	0.03	2.01
1 st stage $F(\hat{\beta}^{iv})$	2.16	2.61	1.86	0.69	5.09			1.89
1 st stage $F(\hat{\beta}^{niv})$	36.61	38.77	13.99	23.23	57.54			37.49

Notes: The table reports results from Monte Carlo estimation of each estimator shown. The true β is 2. Each cell in columns (1)-(5) provides a summary statistic based on estimating the parameter shown in the leftmost column in 500 samples of 50 observations each. Columns (6) and (7) report the coefficient and standard error from pooling across all 500 samples, respectively. Column (8) reports empirical estimates based on the cross-state analysis of the ARRA in Chodorow-Reich (2019).

residual u (none of what follows hinges on this parameter choice). The researcher does not know the true values of the coefficients $\{\pi^\ell\}$, although he/she continues to observe both \bar{z}_i and each z_i^ℓ . In the ARRA example in Chodorow-Reich (2019), y_i is cumulative employment growth over the 24 months following passage of the ARRA, x_i is total state and local (S&L) spending in state i , \bar{z}_i is total spending financed by the ARRA, and each z_i is an excluded instrument based on provisions of the ARRA that used pre-existing formulae to allocate funds across geographic areas.

The changes in equations (21) and (22) render estimation of $\hat{\beta}^{rf}$ undesirable, because the component w_i introduces endogeneity bias and the coefficients multiplying the z s are unknown. However, these coefficients can be powerfully estimated in a first stage where the endogenous variable is the narrow component of spending \bar{z}_i (i.e. total ARRA spending) as in equation (22), even when the first stage using total spending x_i is weak. This insight motivates a fourth estimator, $\hat{\beta}^{niv}$, which keeps the same excluded instruments $\{z_i^\ell\}$ but makes the narrow component of spending, \bar{z}_i , the endogenous variable, and is defined formally as:

$$\text{Narrow IV:} \quad \hat{\beta}^{niv} = (\bar{\mathbf{z}}' \mathbf{P}_z \bar{\mathbf{z}})^{-1} \bar{\mathbf{z}}' \mathbf{P}_z \mathbf{y}. \quad (23)$$

Panel C of table 2 extends the Monte Carlo exercise to this case. I keep all parameters from Panel B unchanged (including setting the true $\pi^1 = \pi^2 = \pi^3 = 1$) and set σ_w , the standard deviation of the residual part of \bar{z}_i , to target a first stage F statistic for estimation of $\hat{\beta}^{niv}$, $F(\beta^{niv})$, of 37, as in the ARRA analysis.¹² The additional variation in the endogenous component of x_i , now $w_i + e_i$, has only a small quantitative impact on the estimates of $\hat{\beta}^{ols}$ or $\hat{\beta}^{iv}$, which remain severely biased. The new estimator, $\hat{\beta}^{niv}$, performs well and indeed behaves almost identically to $\hat{\beta}^{rf}$ from Panel B. Intuitively, the first stage is strong enough that the coefficients multiplying the z s can essentially be treated as known, as they are in Panel B. Therefore, $\hat{\beta}^{niv}$ inherits the superiority of $\hat{\beta}^{rf}$ over $\hat{\beta}^{iv}$.

These differences can matter in practice. As an illustration, the degree of bias of $\hat{\beta}^{ols}$ relative to $\hat{\beta}^{niv}$, the magnitude of the first stage F statistics for $\hat{\beta}^{iv}$ and $\hat{\beta}^{niv}$, and, importantly, the bias of $\hat{\beta}^{iv}$ in the Monte Carlo exercise are all close to the empirical values from the ARRA analysis, shown in the last column of the table. While the bias

¹²This requires setting $\sigma_w = \sqrt{N\sigma_z^2/F(\beta^{niv})}$. The first stage F statistic of 37 corresponds to the “effective” F statistic defined in Montiel Olea and Pflueger (2013) and advocated by Andrews, Stock, and Sun (2019) as the proper metric of first stage strength in the presence of heteroskedastic residuals. This statistic coincides with the conventional F statistic when residuals are homoskedastic, as in the Monte Carlo simulation.

of OLS and the strength of the first stages were targeted moments in the calibration of the Monte Carlo parameters, the second stage coefficients $\hat{\beta}^{iv}$ and $\hat{\beta}^{niv}$ were not directly targeted. Therefore, this exercise suggests that weak instrument bias in $\hat{\beta}^{iv}$ can potentially account for the entire difference between a multiplier of 2.01 using $\hat{\beta}^{niv}$, as reported in Chodorow-Reich (2019), and the value of 1.31 that arises from repeating the same empirical specification but making the endogenous variable total S&L spending and estimating $\hat{\beta}^{iv}$, as advocated by Ramey (2019).¹³

Two final caveats are in order. First, even when the excluded instruments directly impact only a subset of the variation in the endogenous variable, they may nonetheless induce changes in the remainder. The fiscal literature refers to such induced changes as the crowding-out or in of other governmental expenditure. Even so, it may be preferable to report the reduced form specification and assess the amount of crowd-out or in separately, to avoid the weak instrument problem.¹⁴ Second, the weak instrument problem is closely tied to the small sample size — in the panel B estimation pooling all samples together, $\hat{\beta}^{iv} = 1.98$. Of course, in most circumstances the question is what to do with the data available, and not how to expand the sample size by 500-fold. In these circumstances, researchers may want to put emphasis on $\hat{\beta}^{rf}$ or $\hat{\beta}^{niv}$.

5 HATE means Weight?

In the presence of Heterogeneous Average Treatment Effects (HATE), researchers sometimes weight their regressions with the intention of making the estimates nationally rep-

¹³It also bears mentioning that the confidence intervals of these two estimands in the ARRA analysis overlap, even without applying weak instrument coverage adjustments to the estimate of $\hat{\beta}^{iv}$.

¹⁴For example, using the same Monte Carlo parameterization as in Panel C of table 2, a regression of the total endogenous variable x_i on the subcomponent \bar{z}_i , using z_i^1, \dots, z_i^L as excluded instruments, yields a mean (median) coefficient of 1.06 (1.04), indicating no bias in the estimation of crowd-out. In the context of the ARRA, Leduc and Wilson (2017) present evidence of crowd-in of highway expenditure, a result disputed by Dupor (2017). Chodorow-Reich (2019) estimates a statistically insignificant crowd-in of \$0.20 of state and local expenditure per dollar of total ARRA using the same excluded instruments described above.

representative.¹⁵ For example, Ramey (2019) advocates applying population weights to the ARRA specification from Chodorow-Reich (2019).

In their excellent practitioner’s guide to weighted regression, Solon, Haider, and Wooldridge (2015) offer a note of caution when weighting to estimate the population-representative treatment effect in the presence of unmodeled treatment heterogeneity. As they show, weighting by population does not necessarily yield a population-average treatment effect and can yield a coefficient *farther* from the population-average treatment effect than occurs in unweighted regression. The reason is that least squares regression already weights observations according to their contribution to the overall variance of the regressors.¹⁶ Therefore, the effect of further weighting by observation size depends on the joint distribution of size and extremity of the regressors. A similar argument applies to instrumental variables regression, with the variance contributions replaced by Rotemberg (1983) weights. Solon, Haider, and Wooldridge (2015) instead recommend exploring treatment effect heterogeneity explicitly.

Weighted regression can also decrease efficiency and, with instrumental variables, increase bias, especially with skewed weights in small samples. As an extreme example, consider an economy consisting of one large region of measure 1 and a continuum of infinitesimal regions indexed on the unit interval. An unweighted regression on these data has an infinite sample size. With a weighted regression, the large region has a weight of 0.5

¹⁵Another rationale for weighting involves greater efficiency in samples with heteroskedastic residuals. The arguments in this section do not apply to this case.

¹⁶Formally, consider a sample S divided into two subsamples S_1 and S_2 , with regressors $\mathbf{X} = (\mathbf{X}_1' \quad \mathbf{X}_2')'$ and dependent variable $\mathbf{y} = (\mathbf{y}_1' \quad \mathbf{y}_2')'$. Letting π denote the population weight for the observations in subsample 1 and β_1 and β_2 the average treatment effects in sample 1 and sample 2, respectively, the population average treatment effect is $\pi\beta_1 + (1 - \pi)\beta_2$. The least squares coefficient $\hat{\beta}$ for the full sample can be written as $\hat{\beta} \equiv (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}_1'\mathbf{X}_1) \hat{\beta}_1 + (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}_2'\mathbf{X}_2) \hat{\beta}_2$, where $\hat{\beta}_i$ is the least squares coefficient for subsample i . Thus, even without explicit weighting, the least squares regression on the pooled sample already assigns a weight of $(\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}_i'\mathbf{X}_i)$ to the treatment effect in subsample i . Whether weighting the regression by population moves the pooled sample estimate closer to or further from the population-weighted average treatment effect therefore depends on the joint distribution of these variance contributions and the population weights.

and the effective sample size shrinks to the equivalent of a four observation sample.¹⁷ The concomitant decrease in power inflates the standard errors. With instrumental variables, the loss of power in the first stage increases bias.

Table 3 illustrates the problem in a more realistic setting by introducing regression weights into the Monte Carlo exercise reported above. Specifically, keeping the parameters the same as in Panel C of table 2, for each sample I randomly draw (without replacement) from the actual 2010 distribution of state population sizes. Row 1 repeats the summary statistics for $\hat{\beta}^{niv}$ without regression weights. Row 2 reports summary statistics for regressions weighted by the the state population drawn for each observation. Three issues stand out. First, the weighted $\hat{\beta}^{niv}$ is biased, albeit slightly, toward the OLS value. The median bias increases much more for $\hat{\beta}^{iv}$ (shown in row 5), with the distribution of the weighted coefficient centered around 1.05, or roughly half of the true value. Notably, the first stage F statistics *rise* slightly in the weighted specifications (as also occurs with the ARRA data); however, the decline in the effective sample size nonetheless results in larger bias. Second, the distributions of $\hat{\beta}^{niv}$ and $\hat{\beta}^{iv}$ across samples widen considerably in the weighted specifications. The decrease in efficiency occurs because each Monte Carlo observation is drawn from an identical distribution, so that unweighted regression is the best linear unbiased estimator of each stage of 2sls. Third, as shown in the third and sixth rows of the table, the unweighted and weighted coefficients can differ substantially even within the same Monte Carlo sample, despite there being no underlying heterogeneity in treatment effects.

The possibility of highly influential observations in weighted regression also emerges in

¹⁷This calculation follows from an application of a formula first derived by Kish (1965) for sample means. Let $x_i \sim iid(\mu, \sigma^2)$, w_i denote the weight assigned to x_i , with $\sum_{i=1}^N w_i = 1$, and $\hat{\mu}(\{w_i\}) = \sum_{i=1}^N w_i x_i$ the sample mean using weights $\{w_i\}$. Then $\text{Var}(\hat{\mu}(\{w_i\})) = \text{Var}\left(\sum_{i=1}^N w_i x_i\right) = \sum_{i=1}^N \text{Var}(w_i x_i) = \sigma^2 \sum_{i=1}^N w_i^2$. The effective sample size is defined as the number of unweighted observations that would yield the same variance as the weighted mean, or $N_{eff} = 1 / \sum_{i=1}^N w_i^2$.

Table 3: Monte Carlo with Population Weights

Parameter	Weighted	Median	Mean	SD	P10	P90	ARRA data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\beta}^{niv}$	No	1.95	1.98	0.63	1.19	2.82	2.01
$\hat{\beta}^{niv}$	Yes	1.83	1.89	0.89	0.77	3.04	1.15
Difference	.	0.10	0.10	0.64	-0.78	0.89	0.86
$\hat{\beta}^{iv}$	No	1.36	1.46	0.89	0.68	2.40	1.31
$\hat{\beta}^{iv}$	Yes	1.05	1.09	1.12	0.11	2.22	0.89
Difference	.	0.26	0.37	1.31	-0.60	1.48	0.42
$F(\hat{\beta}^{niv})$	No	36.61	38.77	13.99	23.23	57.54	37.49
$F(\hat{\beta}^{niv})$	Yes	39.85	44.90	23.18	21.37	75.15	62.91
$F(\hat{\beta}^{iv})$	No	2.16	2.61	1.86	0.69	5.09	1.89
$F(\hat{\beta}^{iv})$	Yes	2.96	3.89	3.26	0.84	7.75	5.48

Notes: The table reports results from Monte Carlo estimation of each estimator shown. The true β is 2. Each cell in columns (3)-(7) provides a summary statistic based on estimating the parameter shown in the leftmost column in 500 samples of 50 observations each. If indicated in column (2), each regression is weighted by randomly assigned state population. The third and fifth rows of the table report summary statistics for the difference between the unweighted and weighted coefficient in each sample.

the ARRA data. As first shown in Ramey (2019) and repeated in the last column of table 3, weighting the baseline specification from Chodorow-Reich (2019) by population reduces $\hat{\beta}^{niv}$ from 2.01 to 1.15. The Monte Carlo evidence makes clear that such a difference can arise as the result of applying skewed population weights to a 50 observation sample. In the actual data, a single observation — Florida — appears to drive the decline. Removing Florida from the sample and re-estimating $\hat{\beta}^{niv}$ with population weights yields a coefficient of 1.80. By contrast, removing Florida has a much smaller impact in the unweighted specification, with the coefficient rising only slightly from 2.01 to 2.14. More systematically, trimming the two most positive and two most negative influential observations from the weighted specification results in a coefficient of 1.90, very close to the full sample value.

I conclude that weighting by population can introduce substantial cost in small samples in the form of larger bias and estimation variability. While researchers should report both

unweighted and weighted estimates, they may emphasize one or the other in light of these considerations. In the context of the ARRA analysis, when combined with the analysis of the previous section, the unweighted $\hat{\beta}^{niv}$ estimate of 2 may be closer to the truth than the weighted estimate of $\hat{\beta}^{iv}$ of 0.89. Incidentally, this conclusion also better accords with the larger body of evidence on cross-sectional multipliers reviewed in Chodorow-Reich (2019) as well as with the handful of estimates of government spending multipliers in zero lower bound episodes (Ramey and Zubairy, 2018; Miyamoto, Nguyen, and Sergeyev, 2018), which suggest national multipliers in fixed interest rate periods of 1.5 or above.¹⁸ However, clearly the limited sample size of cross-sectional state-level regressions leaves room for other interpretations of the ARRA data. With finer geographic units such as counties, the case for emphasizing weighted estimates grows stronger because the small effective sample size problem becomes less acute.

6 Conclusion

As macroeconomists turn to regional data, many questions arise. In this article, I have offered a framework for understanding the aggregate implications of regional estimates using potential outcomes and practical advice for several aspects of research design and interpretation.

¹⁸Indeed, adjusting the estimate of $\hat{\beta}^{niv} = 2$ by the point estimate of crowd-in of other state and local expenditure reported in footnote 14 of 0.2 yields a local government purchases multiplier from the ARRA of $2/(1+0.2)=1.7$, which falls in the middle of the range of this other evidence.

References

- Adão, Rodrigo, Costas Arkolakis, and Federico Esposito (Feb. 2019). *Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence*. Working Paper 25544. National Bureau of Economic Research.
- Andrews, Isaiah, James Stock, and Liyang Sun (2019). “Weak Instruments in IV Regression: Theory and Practice”. *Annual Review of Economics* 11.
- Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Springer Netherlands. ISBN: 978-94-015-7799-1.
- Auclert, Adrien, Will Dobbie, and Paul Goldsmith-Pinkham (2019). “Macroeconomic Effects of Debt Relief: Consumer Bankruptcy Protections in the Great Recession”.
- Auclert, Adrien and Matthew Rognlie (2017). “A Note on Multipliers in NK Models with GHH Preferences”.
- Auerbach, Alan, Yuriy Gorodnichenko, and Daniel Murphy (2019). “Local Fiscal Multipliers and Fiscal Spillovers in the United States”.
- Baqae, David and Emmanuel Farhi (2019). “Productivity and Misallocation in General Equilibrium”.
- Beraja, Martin, Erik Hurst, and Juan Ospina (2016). “The Aggregate Implications of Regional Business Cycles”.
- Berg, Tobias and Daniel Streitz (2019). “Handling Spillover Effects in Empirical Research: An Application Using Credit Supply Shocks”.
- Blattner, Laura, Luísa Farinha, and Francisco Rebelo (2019). “When Losses Turn Into Loans: The Cost of Undercapitalized Banks”.
- Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019). “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock”. *Econometrica* 87.3, pp. 741–835.
- Chodorow-Reich, Gabriel (2014). “The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008-9 Financial Crisis”. *The Quarterly Journal of Economics* 129.1, pp. 1–59.
- (May 2019). “Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?” *American Economic Journal: Economic Policy* 11.2, pp. 1–34.
- Chodorow-Reich, Gabriel and Antonio Falato (2019). “The Loan Covenant Channel: How Bank Health Transmits to Nonfinancial Firms”.
- Chodorow-Reich, Gabriel, Gita Gopinath, et al. (Forthcoming). “Cash and the Economy: Evidence from India’s Demonetization”. *Quarterly Journal of Economics* 135.1.

- Chodorow-Reich, Gabriel, Plamen Nenov, and Alp Simsek (2019). “Stock Market Wealth and the Real Economy: A Local Labor Market Approach”.
- Chodorow-Reich, Gabriel and Johannes Wieland (Forthcoming). “Labor Reallocation and Business Cycles”. *Journal of Political Economy*.
- Conley, Timothy et al. (2019). “The Technology Multiplier for Employment”.
- Dupor, Bill (2017). “So, Why Didn’t the 2009 Recovery Act Improve the Nation’s Highways and Bridges?” *Federal Reserve Bank of St. Louis Review*, pp. 169–182.
- Dupor, Bill and Rodrigo Guerrero (2017). “Local and aggregate fiscal policy multipliers”. *Journal of Monetary Economics* 92, pp. 16–30. ISSN: 0304-3932.
- Dupor, Bill, Marios Karabarbounis, et al. (2018). *Regional Consumption Responses and the Aggregate Fiscal Multiplier*. Unpublished.
- Dupor, Bill and Peter McCrory (2018). “A Cup Runneth Over: Fiscal Policy Spillovers from the 2009 Recovery Act”. *The Economic Journal* 128.611, pp. 1476–1508.
- Farhi, Emmanuel and Ivan Werning (2016). “Fiscal Multipliers”. *Handbook of Macroeconomics* 2, pp. 2417–2492. ISSN: 1574-0048.
- Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman (1988). “Investment, Capacity Utilization, and the Real Business Cycle”. *The American Economic Review* 78.3, pp. 402–417.
- Guren, Adam et al. (in progress). “What Do We Learn From Cross-Sectional Empirical Estimates in Macroeconomics?”
- Hornbeck, Richard and Enrico Moretti (2019). “Estimating Who Benefits from Productivity Growth: Direct and Indirect Effects of City Manufacturing TFP Growth on Wages, Rents, and Inequality”.
- Huber, Kilian (Mar. 2018). “Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties”. *American Economic Review* 108.3, pp. 868–98.
- Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Kish, Leslie (1965). *Survey Sampling*. John Wiley & Sons, Inc.
- Leduc, Sylvain and Daniel Wilson (May 2017). “Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act”. *American Economic Journal: Economic Policy* 9.2, pp. 253–92.
- McCrory, Peter (2019). “Tradable Spillovers of Fiscal Policy: Evidence from the 2009 Recovery Act”.

- Mian, Atif and Amir Sufi (2014). “What Explains the 2007-2009 Drop in Employment?” *Econometrica* 82.6, pp. 2197–2223.
- Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitriy Sergeyev (July 2018). “Government Spending Multipliers under the Zero Lower Bound: Evidence from Japan”. *American Economic Journal: Macroeconomics* 10.3, pp. 247–77.
- Montiel Olea, José Luis and Carolin Pflueger (2013). “A Robust Test for Weak Instruments”. *Journal of Business & Economic Statistics* 31.3, pp. 358–369.
- Nakamura, Emi and Jón Steinsson (2014). “Fiscal Stimulus in a Monetary Union: Evidence from US Regions”. *American Economic Review* 104.3, pp. 753–92.
- (Aug. 2018). “Identification in Macroeconomics”. *Journal of Economic Perspectives* 32.3, pp. 59–86.
- Ramey, Valerie A. (May 2019). “Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?” *Journal of Economic Perspectives* 33.2, pp. 89–114.
- Ramey, Valerie A. and Sarah Zubairy (2018). “Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data”. *Journal of Political Economy* 126.2, pp. 850–901.
- Redding, Stephen J. and Esteban Rossi-Hansberg (2017). “Quantitative Spatial Economics”. *Annual Review of Economics* 9.1, pp. 21–58.
- Rotemberg, Julio (1983). “Instrumental Variable Estimation of Misspecified Models”.
- Rubin, Donald B. (Jan. 1978). “Bayesian Inference for Causal Effects: The Role of Randomization”. *Ann. Statist.* 6.1, pp. 34–58.
- Sarto, Andres (2019). “Recovering Macro Elasticities from Regional Data”.
- Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge (2015). “What Are We Weighting For?” *Journal of Human Resources* 50.2, pp. 301–316.
- Sraer, David and David Thesmar (2018). “A Sufficient Statistics Approach for Aggregating Firm-Level Experiments”.

A Selected Articles Using Regional Data to Address Macroeconomic Questions

American Economic Review: Kerwin Kofi Charles, Erik Hurst, and Matthew J. Notowidigdo (2018 [108:10], “Housing Booms and Busts, Labor Market Opportunities, and College Attendance”); Ferdinando Monte, Stephen J. Redding, and Esteban Rossi-Hansberg (2018 [108:12], “Commuting, Migration, and Local Employment Elasticities”); Paul Beaudry, David A. Green, and Ben M. Sand (2018 [108:9], “In Search of Labor Demand”); Brad Hershbein and Lisa B. Kahn (2018 [108:7], “Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings”); Kilian Huber (2018 [108:3], “Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties”); Clemens Fuest, Andreas Peichl, and Sebastian Sieglöcher (2018 [108:2], “Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany”); Enrico Moretti and Daniel J. Wilson (2017 [107:7], “The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists”); Aditya Aladangady (2017 [107:11], “Housing Wealth and Consumption: Evidence from Geographically-Linked Microdata”); Rafael Dix-Carneiro and Brian K. Kovak (2017 [107:10], “Trade Liberalization and Regional Dynamics”); Juan Carlos Suárez Serrato and Owen Zidar (2016 [106:9], “Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms”); Rafael Lalive, Camille Landais, and Josef Zweimüller (2015 [105:12], “Market Externalities of Large Unemployment Insurance Extension Programs”); Antonio Acconcia, Giancarlo Corsetti, and Saverio Simonelli (2014 [104:7], “Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-experiment”); Richard Hornbeck and Suresh Naidu (2014 [104:3], “When the Levee Breaks: Black Migration and Economic Development in the American South”); Emi Nakamura and Jón Steinsson (2014 [104:3], “Fiscal Stimulus in a Monetary Union: Evidence from US Regions”); David H. Autor, David Dorn, and Gordon H. Hanson (2013 [103:6], “The China Syndrome: Local Labor Market Effects of Import Competition in the United States”); David H. Autor and David Dorn (2013 [103:5], “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market”); Thomas Chaney, David Sraer, and David Thesmar (2012 [102:6], “The Collateral Channel: How Real Estate Shocks Affect Corporate Investment”); Chris Forman, Avi Goldfarb, and Shane Greenstein (2012 [102:1], “The Internet and Local Wages: A Puzzle”).

Quarterly Journal of Economics: Martin Beraja et al. (2019 [134:1], “Regional Heterogeneity and the Refinancing Channel of Monetary Policy*”); Gabriel Chodorow-Reich, John Coglianesi, and Loukas Karabarbounis (2019 [134:1], “The Macro Effects of Unemployment Benefit Extensions: a Measurement Error Approach”); Xavier Giroud and Holger M. Mueller (2016 [132:1], “Firm Leverage, Consumer Demand, and Employment Losses During the Great Recession*”); Efraim Benmelech, Ralf R. Meisenzahl, and Rodney Ramcharan (2016 [132:1], “The Real Effects of Liquidity During the Financial Crisis: Evidence from Automobiles*”); Dave Donaldson and Richard Hornbeck (2016 [131:2], “Railroads and American Economic Growth: A ”Market Access” Approach”); Patrick Kline and En-

rico Moretti (2013 [129:1], “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *”); Atif Mian, Kamalesh Rao, and Amir Sufi (2013 [128:4], “Household Balance Sheets, Consumption, and the Economic Slump*”); Konrad B. Burchardi and Tarek A. Hassan (2013 [128:3], “The Economic Impact of Social Ties: Evidence from German Reunification*”); Nicola Gennaioli et al. (2013 [128:1], “Human Capital and Regional Development”); Atif Mian and Amir Sufi (2012 [127:3], “The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program*”).

Journal of Political Economy: Johannes Stroebel and Joseph Vavra (2019 [127:3], “House Prices, Local Demand, and Retail Prices”); Owen Zidar (2019 [127:3], “Tax Cuts for Whom? Heterogeneous Effects of Income Tax Changes on Growth and Employment”); Andrew C. Johnston and Alexandre Mas (2018 [126:6], “Potential Unemployment Insurance Duration and Labor Supply: The Individual and Market-Level Response to a Benefit Cut”).

Econometrica: Atif Mian and Amir Sufi (2014 [82:6], “What Explains the 2007-2009 Drop in Employment?”).

Review of Economic Studies: Lorenzo Caliendo et al. (2017 [85:4], “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy”); Hunt Allcott and Daniel Keniston (2018 [85:2], “Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America”).

Journal of Monetary Economics: Kyle F. Herkenhoff, Lee E. Ohanian, and Edward C. Prescott (2018 [93:], “Tarnishing the golden and empire states: Land-use restrictions and the U.S. economic slowdown”); Bill Dupor and Rodrigo Guerrero (2017 [92:], “Local and aggregate fiscal policy multipliers”); Daniel Shoag and Stan Veuger (2016 [84:], “Uncertainty and the geography of the great recession”); Nicholas Coleman and Leo Feler (2015 [71:], “Bank ownership, lending, and local economic performance during the 2008-2009 financial crisis”); Gerald A. Carlino and Robert P. Inman (2013 [60:5], “Local deficits and local jobs: Can US states stabilize their own economies?”).

American Economic Journal: Macroeconomics: Viktor Slavtchev and Simon Wiederhold (2016 [8:2], “Does the Technological Content of Government Demand Matter for Private R&D? Evidence from US States”).

American Economic Journal: Economic Policy: Nathan Falk and Cameron A. Shelton (2018 [10:4], “Fleeing a Lame Duck: Policy Uncertainty and Manufacturing Investment in US States”); Daniele Coen-Pirani and Michael Wooley (2018 [10:2], “Fiscal Centralization: Theory and Evidence from the Great Depression”); Sumit Agarwal, Nathan Marwell, and Leslie McGranahan (2017 [9:4], “Consumption Responses to Temporary Tax Incentives: Evidence from State Sales Tax Holidays”); Decio Coviello and Stefano Gagliarducci (2017 [9:3], “Tenure in Office and Public Procurement”); Richard Hornbeck and Pinar Keskin (2015 [7:2], “Does Agriculture Generate Local Economic Spillovers? Short-Run and Long-Run Evidence from the Ogallala Aquifer”); Sascha O. Becker, Peter H. Egger, and Maximilian von Ehrlich (2013 [5:4], “Absorptive Capacity and the Growth and Investment Effects

of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects”); Daniel J. Wilson (2012 [4:3], “Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act”); Gabriel Chodorow-Reich, Laura Feiveson, et al. (2012 [4:3], “Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act”); Jeffrey Clemens and Stephen Miran (2012 [4:2], “Fiscal Policy Multipliers on Subnational Government Spending”).