How to Use Spatial Instruments

James Sundquist*

July 9, 2021

Abstract

Betz, Cook, and Hollenbach (2018) offer two notes of caution with regard to spatial instruments — instrumental variables based on the value of the treatment variable in other units. Their first point is well taken: such strategies merit special attention to problems of spillover and interdependent outcomes. Yet their second point, that for such instruments "the exclusion restriction is, necessarily and by construction, violated," is overstated. I show that spatial instruments can be consistent estimators when they work by serving as a proxy for an unobserved instrument. These results clarify when such strategies are appropriate and suggest that researchers justify them by specifying the unobserved instrument that underlies the strategy and presenting qualitative evidence of its strength.

*Ph.D. Candidate, Department of Political Science, Yale University (james.sundquist@yale.edu)

1 Introduction

It is increasingly common to see researchers construct an instrumental variable (IV) whose value for each unit i is calculated by summing or otherwise combining the treatment variable for some set of other units $\neg i$. Because these instruments are calculated in a cross-sectional way (most often by averaging the treatment of neighboring units), they are sometimes called "spatial instruments," though the term has also been applied to other instrumental variables with a spatial component. The technique can appear to be a form of statistical alchemy, creating an instrumental variable out of nothing more than a set of confounded treatments.

Betz, Cook, and Hollenbach (2018) warn that the technique is indeed too good to be true, categorically denying the validity of such instruments. They assert that "if [such an instrument] is strong, it violates the exclusion restriction; if the instrument does not violate the exclusion restriction, it is irrelevant." They conclude that spatial instruments "cannot produce valid inferences."

In this article, I show that this critique assumes that treatment values form a cyclic graph; that is, $D_1 = f(D_2)$ and $D_2 = f(D_1)$. In such a case, inference is indeed impossible. However, there is a way for D_1 and D_2 to be correlated that does not violate the exclusion restriction: if both are influenced by some causally prior variable Z. I argue that some spatial instrument strategies implicitly assume the existence of an unobserved instrument Z, and work by using aggregated treatments as a proxy for Z. I show that this strategy assumes that the hidden instrument affects groups of units (such as all units in a given region), and that it uses comparisons between different groups to identify the causal effect of interest. In other words, if the average levels of the treatment D are shifted across regions by Z and only by Z, the average of D in a given region can proxy for the instrument Z.

In the third section, I introduce a procedure to estimate the value of an unobserved instrument using confounded treatment variables. I show that, under the assumption that errors are mean zero in each group, the estimator is unbiased. The proxied instrument can then be used to estimate a causal effect of interest. The online appendix presents the results of a simulation study that characterizes the performance of a spatial instruments strategy.

These results suggest that researchers should think carefully about the causal graph they claim underlies their spatial instruments strategy, and when possible, support this claim with qualitative evidence. If causal arrows exist between treatment values, spatial instruments are unlikely to be helpful. If the underlying causal diagram presumes an unobserved instrument, spatial instruments are potentially useful, and all the more convincing if additional evidence attests to the existence and exogeneity of this unobserved instrument.

2 When Spatial Instruments Can and Cannot Be Used

The core critique leveled by Betz et al. is that "feedback from the endogenous predictor to the instrument makes the instrument a function of the source of endogeneity — that is, it makes the instrument itself endogenous." While they make use of causal diagrams elsewhere, this key section restricts itself to systems of equations. The use of an equality sign rather than a directed relationship obscures the multiple ways in which treatment variables can be causally related, with important consequences for what a spatial instrument is capturing.

If the language of "from…to" is interpreted in a causal sense, simultaneity bias does indeed frustrate any attempt at inference. Figure 1 makes this immediately apparent: a cycle exists between D_1 and D_2. Furthermore, no instrumental variable appears at all in the graph! Because there is no underlying natural experiment, there is no way to separate exogenous variation in D from its endogenous component.

Figure 1: Cyclic graph

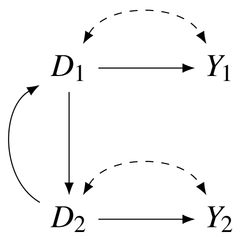

Figure 2: Identification possible

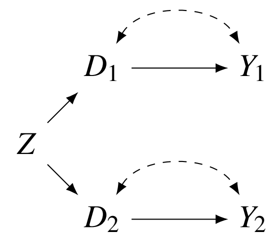

Figure 2, however, presents an alternative situation in which D_1 and D_2 are correlated, but no feedback loop exists: a classic instrumental variable, Z, affects outcomes through and only through treatment. If Z were observed, it could be used in a traditional IV setup. Yet if it is unobserved, spatial instruments can use average differences in D to infer differences in Z. Crucially, the strategy can only work if this hidden instrument is the *only* unobserved cause each unit has in common.

The simplest application of this insight would involve cross-sectional data, in which Z takes different values in different regions. The mean of D for all other units in a given region would proxy for the unobserved regional Z, and cross-regional variation in the spatial instrument would identify the effect of D on Y. This technique has seen some use in biostatistics, where varying regional preferences for a particular medical procedure have been used as an instrument to evaluate the procedure's effectiveness (Baiocchi, Cheng, and Small, 2014, p. 2303).

Alternatively, Z could vary across time periods instead of across regions. Assuming that the hidden instrument is the only unobserved factor shifting average levels of treatment over time, these over-time changes can be attributed to exogenous variation in Z, and used for causal identification. As the following section explains in more detail, this assumption poses an obstacle for spatial instruments strategies, but still allows for many types of confounding ruled out by other inferential strategies, such as two-way fixed effects. In circumstances where it holds, spatial instruments are in fact valid. The key is to recognize that they are not creating an instrument where there was none, but revealing the shape of one that was already there, though unobserved.

3 Spatial Instruments as a Proxying Strategy

I begin with the cross-sectional case. Given the system of equations with k regions indexed by j, each with n units indexed by i:

$$D_i = \beta_1 Z_{ij} + u_i \tag{1}$$

$$Y_i = \beta_2 D_i + \varepsilon_i \tag{2}$$

The estimand of interest is β_2, but let us assume that estimation via ordinary least squares is confounded by correlation between D_i and ε_i. Z_{ij} operates at the regional level, taking the same value for every unit in a given region j, and satisfies the two requirements of an instrument: $Cov(Z_{ij}, \varepsilon_i) = 0$ and $Cov(Z_{ij}, D_i) \neq 0$. However, it is unobserved. Under the additional assumption that errors in equation (1) are mean zero within each region, $E[u_i|Z_{ij}] = 0$, I show that the instrument can be proxied by a spatial instrument: the average treatment of all *other* units in the region.

Proposition. *Define* γ_i *as* $\sum^{\neg i} D_i$ *(where ¬i refers only to units in i's same region)*

If $E[u_i|Z_{ij}] = 0$*, then* $\frac{\gamma_i}{n-1}$ *is an unbiased estimator for* $\beta_1 Z_{ij}$*.*

Proof.

$$\gamma_i = \sum^{\neg i} D_i$$

$$\gamma_i = \sum^{\neg i} \beta_1 Z_{ij} + \sum^{\neg i} u_{it}$$

$$E[\gamma_i] = (n-1)(\beta_1)E[Z_{ij}] + (n-1)E[u_i]$$

$$\frac{E[\gamma_i]}{n-1} = \beta_1 E[Z_{ij}]$$

□

We can then use γ_i as a proxy for the unobserved instrument in a conventional IV setup, where the second equality follows from the fact that Z_{ij} does not vary within regions, and so $Z_{ij} = E[Z_{ij}]$.

$$\frac{Cov(\gamma_i, Y_i)}{Cov(\gamma_i, D_i)} = \frac{\beta_1 Cov(E[Z_{ij}], Y_i)}{\beta_1 Cov(E[Z_{ij}], D_i)} = \frac{Cov(Z_{ij}, Y_i)}{Cov(Z_{ij}, D_i)} = \beta_2$$

This approach to spatial instruments makes three things clear: first identification of β_2 depends on regional variation in the hidden instrument Z. Because we can only infer Z

at the level of the region, it is essential to observe multiple regions with varying levels of Z. Second, regional differences in average treatment D that remain after controlling for any observable confounders must be attributable to Z. Regional fixed effects are impossible because they would be perfectly collinear with Z_{ij}. Third, despite this limitation, spatial instruments greatly narrow the scope of threats to inference. Whereas controlling strategies block specific confounders but remain vulnerable to all others, this strategy can overcome all forms of confounding except for the specific problem of unobserved differences between regions.

This approach can also be applied to panel data. Just as the cross-sectional version assumed that the unobserved instrument explained cross-regional variation, the panel data version assumes that the instrument shifts average levels of treatment over time. Panel data have the additional advantage of being able to condition on a unit fixed effect a_i to satisfy the conditional independence assumption $E[u_{it}|a_i, Z_{it}] = 0$.

While this procedure is consistent, it suffers from a problem in samples with few groups: within each group (either region or year), the unit with the highest treatment value will have the smallest γ_i, while the largest γ_i will belong to the unit with the lowest treatment value. The resulting negative correlation between γ_i and D_i within groups weakens the instrument, which is motivated by positive correlation with D_i across groups. This problem is particularly acute in situations with few groups, where there is less cross-group variation to compensate for the problem. Figure 3 illustrates these cross-cutting correlations. With only two small groups (the filled circles), the correlation between γ_i and D_i is weak. Adding more units to these two groups (denoted by hollow circles) does not address this problem — though by shrinking the variance of $E[\overline{u}|Z_{ij}]$, it does reduce the degree to which random variation can violate the assumption $E[u_i|Z_{ij}] = 0$. Additional groups (denoted by red circles) are necessary to strengthen the correlation between the instrument and the treatment variable.[1] As a consequence of this quirk of spatial instruments, researchers should try to include as many groups of units as possible and report the first-stage F-statistic to confirm that the instrument is sufficiently strong.

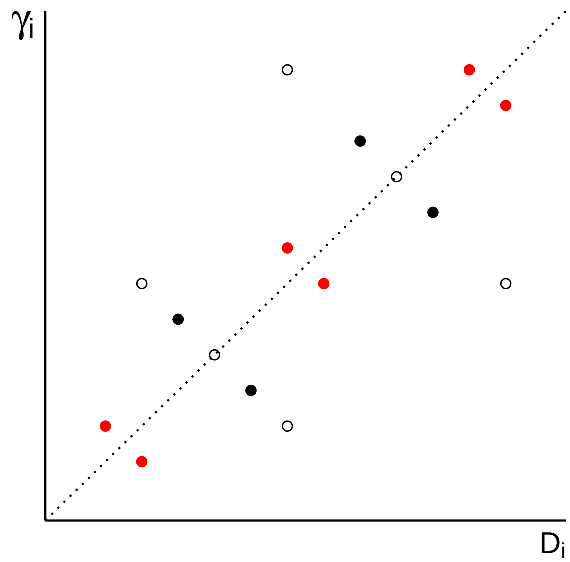

Figure 3: Adding additional observations to the initial two groups (open circles) does not strengthen the correlation between γ_i and D_i across groups, but including more groups with varying levels of Z (red dots) does.

Another solution to the problem of a weak spatial instrument is to introduce a degree of bias by including unit i's treatment value in the construction of its instrument. This modified instrument $\Gamma_i = \sum^i D_i$ or $\Gamma_{it} = \sum^i D_{it}$ will take the same value for every unit in a group. Simulations confirm that this procedure enjoys lower variance in situations with few groups, at the cost of bias from including a unit's own confounded treatment in

[1] Speaking more generally, the partial correlation between the two variables after partialling out any other covariates.

its instrument. This bias also shrinks as groups contain more units, because each unit's treatment exerts less effect over the group average. Yet because bias is still unwelcome as a source of error, researchers can take advantage of the estimators' complementary qualities and use both. If the results substantively agree, there is likely to be enough across-group variation to ensure a strong instrument when using γ_i and enough units per group to shrink bias when using Γ_i. If they return meaningfully different results, at least one of the problems is severe. Because having few units per group compromises the key assumption of errors within each group summing to zero, it poses problems for both γ_i and Γ_i. In the case of few groups with many units, Γ_i may be accurate enough to be useful. In the online appendix, a simulation study provides evidence in support of these assertions.

Finally, some readers may wonder: given the strong assumption that cross-group variation is driven by an exogenous instrument, why not simply average treatment and outcomes by group and estimate β_2 using OLS at the group level? I believe that there are at least three reasons to prefer a spatial instruments approach. First, even when clustering standard errors by group to account for co-determined treatment status, the spatial instruments approach can return more precise estimates by drawing on a larger number of observations. Second, the instrumental variables framework encourages more attention to the underlying natural experiment and questions of research design. Most important, however, is the fact that spatial instruments can be useful even when this assumption does not hold for all units. An instructive example comes from Dube and Naidu (2015). Studying the effect of US military aid on paramilitary violence in Colombia, they reason that this aid is influenced by fluctuations in the Pentagon's total annual military assistance budget, but as a small percentage of the overall budget, is unlikely to drive these fluctuations. The spatial instrument is particularly defensible in this single-unit case. An in-between approach is also possible: if the assumption that budget fluctuations are exogenous is appropriate for some countries other than Colombia, but not appropriate for another set of countries (e.g., Iraq and Afghanistan in Dube and Naidu's case), all units' treatment information can be used in constructing the instrument, but only the outcomes for countries that satisfy the assumption used to estimate β_2.

4 Discussion

What does this mean for the use of spatial instruments in social science? Researchers who hope to employ such a strategy must be explicit about what hidden instrument underlies their study. They cannot fall victim to the fallacy of composition that can occur when checking IV assumptions one-by-one, and report a strong first stage for a spatial instrument while arguing that spillover is unlikely. Together, these facts suggest that treatment assignments share a causal ancestor. Theory and qualitative evidence should support this ancestor's existence and exogeneity. If it is not exogenous, it is not a valid instrument.

Among the papers cited by Betz et al., two attempts to instrument for democracy prove illuminative. Ansell (2008) uses regional Polity scores as an instrument for each country's own Polity score, without attention to what causes these measures to be correlated, and why this cause should be considered exogenous. This is an example of checking IV assumptions one-by-one, without developing a theory of the underlying identifying variation. Acemoglu et al. (2019) on the other hand, use regional waves of democratization as an instrument for democracy. By focusing on waves, readers can scrutinize whether the identifying variation satisfies the independence assumption.

In conclusion, spatial instruments cannot be used to create an instrument where none exists, but can be used to proxy for one that is unobserved. In such cases, spatial instruments can in fact produce valid inferences. However, spatial instruments rule out the use of region or time fixed effects, requiring strong assumptions about the absence of confounders at the group level. Thus, spatial instruments demand more, not less, substantive knowledge of causal processes.

Acknowledgments

I am grateful to Matthew Graham, Peter Aronow, and Winston Lin for helpful comments.

Data Availability

The `R` code used to generate and analyze the data for the simulation study is available as part of the online appendix.

Conflicts of Interest

The author declares that there is no conflict of interest.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy* 127(1): 47–100.

Ansell, Ben W. 2008. "Traders, Teachers, and Tyrants: Democracy, Globalization, and Public Investment in Education." *International Organization* 62(2): 289–322.

Baiocchi, Michael, Jimg Cheng, and Dylan S. Small. 2014. “Instrumental variable methods for causal inference.” *Statistics in Medicine* 33: 2297–2340.

Betz, Timm, Scott J. Cook, and Florian M. Hollenbach. 2018. “On the Use and Abuse of Spatial Instruments.” *Political Analysis* 26(4): 474–479.

Dube, Oeindrila, and Suresh Naidu. 2015. “Bases, Bullets, and Ballots: The Effect of US Military Aid on Political Conflict in Colombia.” *The Journal of Politics* 77(1): 249–267.