Economics Letters 137 (2015) 123–126

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction☆

Michael S. Delgado [a,*], Raymond J.G.M. Florax [a,b]

[a] Department of Agricultural Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907, USA

[b] Department of Spatial Economics, Vrije Universiteit Amsterdam, and Tinbergen Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

HIGHLIGHTS

- Develop a difference-in-differences estimator for spatial data.
- Allow for local spatial interaction in potential outcomes.
- Identify direct and indirect treatment effects.
- Monte Carlo simulations show good finite sample performance.

ARTICLE INFO

Article history:
Received 24 May 2015
Received in revised form
24 September 2015
Accepted 26 October 2015
Available online 30 October 2015

JEL classification:
C21
C53

Keywords:
Difference-in-differences
Monte Carlo simulation
Program evaluation
Spatial autocorrelation
Spatial interaction

ABSTRACT

We consider treatment effect estimation via a difference-in-difference approach for spatial data with local spatial interaction such that the potential outcome of observed units depends on their own treatment as well as on the treatment status of proximate neighbors. We show that under standard assumptions (common trend and ignorability) a straightforward spatially explicit version of the benchmark difference-in-differences regression is capable of identifying both direct and indirect treatment effects. We demonstrate the finite sample performance of our spatial estimator via Monte Carlo simulations.

1. Introduction

The linear difference-in-differences (DID) model is a benchmark tool in the program evaluation literature (e.g., Ashenfelter, 1978; Ashenfelter and Card, 1985). At its core, a treatment effect is the difference between two potential outcomes, with potential outcomes being a function of treatment status (Rubin, 1974). The fundamental problem is that units are never observed in both treated and untreated states (Holland, 1986), and identification requires comparison of treated units to untreated (control) units. In the standard DID design, observations $i = (1, 2, \ldots, n)$ are observed in two time periods, $T \in \{0, 1\}$, and are grouped via $D \in \{0, 1\}$ such that $D_i = 1$ indicates treatment. Given a vector of time-varying covariates, X_{it}, the standard DID equation is:

$$y_{it} = \alpha_0 + \alpha_1 X_{it} + \alpha_2 D_{it} + \alpha_3 T_{it} + \alpha_4 D_{it} T_{it} + \varepsilon_{it}, \quad (1)$$

in which ε_{it} is a mean-zero error term that is uncorrelated with D_{it} and T_{it}. It is straightforward to accommodate additional time periods. The identifying assumptions require correct linear specification of the conditional mean, a homogeneous effect of treatment, and the parallel-trends assumption denoting that absent treatment both treated and untreated units evolve along the same temporal path. Strong or weak ignorability (unconfoundedness) is assumed as well, implying that treatment assignment is independent of the

☆ This research was supported in part by computational resources provided by Information Technology at Purdue (ITAP)–Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, USA. The authors would like to thank Arnab Bhattacharjee, Julie Le Gallo, and an anonymous reviewer for excellent comments.

* Corresponding author.
E-mail addresses: delgado2@purdue.edu (M.S. Delgado), rflorax@purdue.edu (R.J.G.M. Florax).

http://dx.doi.org/10.1016/j.econlet.2015.10.035
0165-1765/

outcome y_{it}, eventually conditional on X_{it}. Maintaining these assumptions and suppressing the subscript, the conditional average treatment effect, $ATE(x)$ is:

$$\begin{aligned} ATE(x) = & \{\mathrm{E}[y|X=x, D=1, T=1] \\ & - \mathrm{E}[y|X=x, D=1, T=0]\} \\ & - \{\mathrm{E}[y|X=x, D=0, T=1] \\ & - \mathrm{E}[y|X=x, D=0, T=0]\}, \end{aligned} \tag{2}$$

or the difference in the conditional differences over time between the treated and control units. From Eqs. (1) and (2), we can see that $ATE = \alpha_4$. Typically, Eq. (1) is estimated with ordinary least squares (OLS), and a rejection of $H_0 : \alpha_4 = 0$ via a t-test is evidence of a significant causal effect.

Another key assumption needed for identification is the stable unit treatment value assumption (SUTVA) (Rubin, 1978, 1990) which "implies that potential outcomes for person i are unrelated to the treatment status of other individuals" (Angrist et al., 1996, p. 446). Violation of the SUTVA assumption, referred to as "interference" or "social interaction", invalidates identification of causal effects in the traditional DID setup, and adjustments are clearly needed. It no longer suffices to only consider one's own treatment status, but the treatment status of other units has to be taken into account as well. For example, imagine a program designed to improve crop yields by enhancing farming ability through education or experience. Communication among neighboring farmers may evoke indirect treatment effects, in which treated and untreated farmers benefit from treatment (knowledge) of others. Similar examples are prevalent throughout economics.

We develop a spatial DID model that accounts for the possibility of spatial correlation in treatments and spatial interaction in treatment responses.[1] The term "spatial interaction" is deliberate and refers to the fact that both direct and indirect treatment effects exist, because potential outcomes are not independent. Indirect treatment effects are sometimes referred to as "spillovers" or "network effects", resulting from contagion, displacement, communication, social comparison, or signaling (Gerber and Green, 2012). Such spillovers create spatial correlation in the treatment response. It is important to note that we are explicitly interested in spatial correlation caused by dependence in potential outcomes, as opposed to spatial correlation stemming from spatially correlated treatments or any other sort of spatial sorting process. Spatial interaction in treatment responses requires a plausible causal mechanism to be at work; existing spatial patterns due to spatial sorting need to be excluded as a source of spatially correlated treatment responses. This is akin to the well known ignorability assumption evoked in the traditional DID setup (see Imbens and Wooldridge, 2009).

We assume treatment is binary, and restrict our focus to spatial interaction in the treatment results that are "local", as opposed to "global". This means that the spatial effects are restricted to immediate neighbors, defined on the basis of contiguity or distance.[2] In effect, this amounts to assuming that SUTVA holds outside the bounds of immediate neighbors. In the context of the farming example one might assume that SUTVA holds *between* villages, but local spatial interaction and hence SUTVA violation is allowed for *within* villages. Generally, however, the only requirement for our approach to be valid is that SUTVA holds across some particular dimension of the sample space; it is not necessarily restricted to a discernable within–between situation. Under these general circumstances, a simple extension of the standard DID setup using spatial econometric tools allows us to estimate the *ATE*, which we decompose into a direct ("own") and indirect ("neighbor") treatment effect. Our spatial DID model is straightforward to implement, and simple Monte Carlo exercises illustrate good finite sample performance.

2. Spatial difference-in-differences

Imagine a setup with n cross-sectional, spatial observations observed over two time periods that can represent points or areas. We operationalize local spatial interaction in treatment responses using the spatial lag operator L^s, defined as W_s, where W_s is a $(2n \times 2n)$ block-diagonal row-standardized spatial weights matrix containing non-zero elements for spatial units belonging to contiguity class s (Anselin, 1988). The contiguity class can vary, as long as the number of neighbors is restricted and does not extend to the entire spatial system. Spatial interaction in outcomes is defined as $D_L T = (I + \rho L^s)D \circ T = (I + \rho W_s)D \circ T$, where I is the identity matrix, ρ a spatial autoregressive parameter, and $\circ$ signals element-by-element multiplication (Hadamard product).

One question is whether we need to distinguish random treatment from spatially correlated treatment.[3] In the case of spatial correlation in treatment the SUTVA assumption is not violated. Moreover, even with spatially correlated exogenous covariates the unbiasedness, consistency, and efficiency properties of OLS are unaltered (Anselin, 1988). Hence the standard DID approach leads to proper identification and estimation of the *ATE*, and we do not require spatially explicit reparameterization of Eq. (1).

The situations with spatial interaction in the responses are different because SUTVA is violated and indirect effects should therefore be explicitly modeled. One common approach is to meticulously identify all treatment and control groups and apply a difference-in-difference-in-differences technique (Imbens and Wooldridge, 2009). There are several disadvantages to such an approach. First, the number of groups becomes unwieldy and the estimator becomes inefficient in small samples.[4] Second, since there are no restrictions on the estimated parameters one may be confronted with illogical "bouncing beta's", such as the case where the treatment response of having two treated neighbors may be greater than that of having four treated neighbors. We therefore suggest a spatially structured approach.

Without loss of generality the model with spatial interaction in the responses is:

$$\begin{aligned} y &= \alpha_0 + \alpha_1 X + \alpha_2 D + \alpha_3 T + \alpha_4 D_L \circ T + \varepsilon, \\ &= \alpha_0 + \alpha_1 X + \alpha_2 D + \alpha_3 T + \alpha_4 (I + \rho W) D \circ T + \varepsilon, \\ &= \alpha_0 + \alpha_1 X + \alpha_2 D + \alpha_3 T + \alpha_4 D \circ T + \alpha_5 WD \circ T + \varepsilon, \end{aligned} \tag{3}$$

in which $\alpha_5 = \rho\alpha_4$, and D represents either random or spatially correlated treatments. We assume $\rho \neq 0$ so that the model does not revert to the standard DID equation. Omitting $WD \circ T$ renders

[1] In spatial econometrics, "autocorrelation" and "dependence" are often treated as synonymous, even though dependence is a characteristic of the joint probability density function and can only be measured as correlation under assumptions such as normality, stationarity, etc. Spatial patterns on a map can be generated through a dependence mechanism that may or may not be known, or result from spatial heterogeneity (Anselin, 1988).

[2] In the global case, indirect effects propel through the entire spatial system; everybody is a neighbor of everybody else, with interactions (or correlations) subject to distance decay.

[3] Given that D is binary, join count statistics can be used to determine whether there is spatial autocorrelation in treatment (Cliff and Ord, 1981).

[4] A spatial setting with local indirect effects implies that groups would have to be defined such that units with different numbers of treated neighbors would fall into different treatment or control groups. In the simple case of a (10×10) spatial grid with adjacency defined as the sharing of a border or vertex, the number of groups based on spatially interactive responses is 9×2, the number of treated neighbors $(0, 1, \ldots, 8)$ multiplied by 2 to incorporate whether own treatment is 0 or 1.

the standard DID estimator biased and inconsistent, because of the omitted variable problem.

Given Eq. (3), it is straightforward to derive a conditional *ATE*. Treated and untreated units may be subject to indirect treatment effects caused by treated neighbors ($wd \in WD$ for $0 < wd \leq 1$), but the control group is the group that is not treated at all, neither directly nor indirectly ($D = 0, WD = 0$). Formally:

$$\begin{aligned} ATE(wd) &= \{\mathrm{E}[y|X, D = 1, T = 1, WD = wd] \\ &\quad - \mathrm{E}[y|X, D = 1, T = 0, WD = wd]\} \\ &\quad - \{\mathrm{E}[y|X, D = 0, T = 1, WD = 0] \\ &\quad - \mathrm{E}[y|X, D = 0, T = 0, WD = 0]\} \\ &= \alpha_4 + \alpha_5 wd \\ &= \alpha_4(1 + \rho wd), \end{aligned} \tag{4}$$

which leads to the average treatment effect:

$$ATE = \mathrm{E}[ATE(wd)|WD] = \alpha_4\left(1 + \rho\overline{WD}\right), \tag{5}$$

where $\overline{WD}$ is the average proportion of treated neighbors.

With spatially interactive responses the *ATE* becomes a function of the magnitude of the direct effect of treatment, the strength of the local spatial interaction, and the proportion of treated neighbors. The standard DID estimator, $ATE = \alpha_4$, is clearly biased. Further, given (4), one can write the *ATE* as a dose–response-type function of indirect treatment that can be evaluated at any $wd \in (0, 1)$. This gives rise to interesting comparisons, for instance, the impact of (only) indirect treatment for some $0 < wd \leq 1$ versus the impact of direct treatment with $wd = 0$. Finally, the *ATE* in Eq. (5) can be decomposed into an average *direct* treatment effect (*ADTE*)

$$\begin{aligned} ADTE &= \{\mathrm{E}[y|X, D = 1, T = 1, WD = 0] \\ &\quad - \mathrm{E}[y|X, D = 1, T = 0, WD = 0]\} \\ &\quad - \{\mathrm{E}[y|X, D = 0, T = 1, WD = 0] \\ &\quad - \mathrm{E}[y|X, D = 0, T = 0, WD = 0]\} \\ &= \alpha_4, \end{aligned} \tag{6}$$

and an average *indirect* treatment effect (*AITE*)

$$\begin{aligned} AITE(wd) &= \{\mathrm{E}[y|X, D = 0, T = 1, WD = wd] \\ &\quad - \mathrm{E}[y|X, D = 0, T = 0, WD = wd]\} \\ &\quad - \{\mathrm{E}[y|X, D = 0, T = 1, WD = 0] \\ &\quad - \mathrm{E}[y|X, D = 0, T = 0, WD = 0]\} \\ &= \alpha_5 wd, \end{aligned} \tag{7}$$

which yields $AITE = \alpha_5\overline{WD}$. The *ADTE* is identified by setting indirect treatment *WD* to zero, and in all cases the necessary control group consists of units that are neither treated directly nor indirectly.

3. Monte carlo simulations

We consider the following data generating process:

$$y_{it} = \alpha_0 + \alpha_1 X_{it} + \alpha_2 D_{it} + \alpha_3 T_{it} + \alpha_4 R_{it} + \varepsilon_{it}, \tag{8}$$

with R_{it} for the responses, and other notation as before. We generate $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ and include:

(a) $\dot{X}_{it} = a\dot{X}_{it-1} + \nu_i$	a common trend,
(b) $X = (I + \ell W)\dot{X}$	random or autocorrelated sorting,
(c) $D_t = D_{t-1} \sim$ Bern$(n, \Phi(X_{t-1}))$	random or autocorrelated binary treatment,
(d) $R = (I + \rho W)D \circ T$	autocorrelated response due to interaction,

where $\dot{X}_{it-1} \sim \mathcal{N}(0, 1)$ and we calculate $\dot{X}_{it}$ using a common trend for the treatment and control group with $a = 1.02$ and $\nu_i \sim \mathcal{N}(0, (a-1)^2)$. We further allow X to be spatially autocorrelated through the autoregressive parameter ℓ and a row-standardized block-diagonal matrix with first-order queen weights on a regular lattice. The parameter ℓ is not estimated, but allows us to incorporate spatial sorting into the model. The sorting variable is clearly correlated with the binary treatments, which are generated using a Bernoulli distribution for the normally distributed treatment probabilities $\Phi(X_{t-1})$. As a result, we investigate violations of the ignorability assumption, because D_{it} and y_{it} are only conditionally independent and X_{it} should therefore be included in the DID equation to maintain unconfoundedness. Finally, ρ governs spatial interaction in the response. We use 1000 replications, maintaining $\alpha_k = 1, \forall k$, and varying $\ell = \{0, 0.9\}$, $\rho = \{0, 0.50, 0.90\}$ and $n = \{100, 900, 2500\}$.[5]

We compare the traditional DID estimator from Eq. (1) and our suggested spatial DID estimator given in Eq. (3), using the labels DID and SDID, respectively. In addition, we use a specification in which we add X to the estimated equations, labeling these estimators DIDX and SDIDX. Table 1 presents the simulation results in terms of bias and root mean squared error (RMSE) of the four estimators.

When treatment is random and there is no spatial interaction in outcomes, all estimators perform well. DIDX achieves the smallest bias and the lowest RMSE because it is perfectly specified. This superior performance of DIDX is reinforced if we allow treatments to be spatially correlated. However, in that case our SDID estimator is biased because the ignorability assumption is violated. The SDID estimator incorrectly interprets spatial sorting in X to be spatial interaction in the response. In the case where there is spatial interaction, DID and DIDX are biased and inconsistent, because SUTVA is violated and the spatial interaction is not accounted for. In the case where there is both spatial sorting and spatial interaction, the SDID estimator is biased and inconsistent, because although it accounts for the spatial interaction it fails to recognize the spatial sorting. The effect of spatial sorting is picked up as a treatment effect, which it is not. Overall, spatial correlation in treatments does not substantially affect the different estimators, except for SDID that erroneously attributes spatial sorting to spatial interaction. In cases where both spatial sorting and spatial interaction in treatment responses are relevant, SDIDX is superior.

4. Conclusion

We develop a straightforward DID estimator for spatial data and allow for the possibility of spatially correlated treatments and local spatial interactions in treatment responses. The *ATE* depends on both the average direct effect of own treatment as well as the average indirect effect from the share of proximate neighbors that receive treatment. Monte Carlo simulations illustrate that in the case of spatial data an appropriate *ATE* estimator should account for potential outcomes which are a function of treatment status of other units because the SUTVA assumption is likely violated, and spatial correlation in treatments should be conditioned out if it results from a violation of the unconfoundedness assumption.

[5] A more extensive set of simulation results is available from the authors upon request.

Table 1
Bias and RMSE for different *ATE* estimators.

ℓ	ρ	ATE estimator	Bias			RMSE		
			$n = 100$	$n = 900$	$n = 2500$	$n = 100$	$n = 900$	$n = 2500$
0.0	0.0	DID	0.0350	0.0214	0.0249	0.2791	0.0956	0.0607
.		DIDX	0.0121	−0.0012	0.0023	0.2771	0.0930	0.0553
.		SDID	0.0342	0.0207	0.0306	0.4372	0.1512	0.0972
.		SDIDX	0.0069	−0.0062	0.0078	0.3818	0.1273	0.0791
.	0.5	DID	−0.2227	−0.2254	−0.2281	0.3660	0.2447	0.2350
.		DIDX	−0.2455	−0.2480	−0.2507	0.3803	0.2655	0.2570
.		SDID	0.0078	0.0273	0.0215	0.4623	0.1536	0.0917
.		SDIDX	−0.0042	0.0037	−0.0008	0.3975	0.1325	0.0768
.	0.9	DID	−0.4420	−0.4270	−0.4277	0.5290	0.4374	0.4317
.		DIDX	−0.4647	−0.4495	−0.4503	0.5483	0.4594	0.4540
.		SDID	0.0151	0.0281	0.0249	0.4693	0.1543	0.0943
.		SDIDX	−0.0002	0.0051	0.0018	0.3969	0.1316	0.0795
0.9	0.0	DID	0.0289	0.0205	0.0283	0.2780	0.0967	0.0631
.		DIDX	0.0043	−0.0040	0.0040	0.2765	0.0945	0.0565
.		SDID	0.5589	0.5491	0.5607	0.7027	0.5670	0.5673
.		SDIDX	0.0000	−0.0059	0.0071	0.3532	0.1178	0.0713
.	0.5	DID	−0.1807	−0.1832	−0.1840	0.3498	0.2058	0.1927
.		DIDX	−0.2055	−0.2080	−0.2087	0.3629	0.2280	0.2164
.		SDID	0.5513	0.5596	0.5560	0.7069	0.5765	0.5621
.		SDIDX	−0.0072	0.0031	−0.0009	0.3645	0.1163	0.0703
.	0.9	DID	−0.3614	−0.3500	−0.3510	0.4655	0.3622	0.3560
.		DIDX	−0.3868	−0.3750	−0.3760	0.4854	0.3864	0.3807
.		SDID	0.5562	0.5605	0.5573	0.6961	0.5767	0.5640
.		SDIDX	0.0016	0.0050	0.0014	0.3581	0.1136	0.0722

References

Angrist, J., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Amer. Statist. Assoc. 91, 444–472.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. Rev. Econ. Stat. 60, 47–57.

Ashenfelter, O., Card, D., 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. Rev. Econ. Stat. 67, 648–660.

Cliff, A.D., Ord, J.K., 1981. Spatial Processes: Models and Applications. Pion, London.

Gerber, A.S., Green, D.P., 2012. Field Experiments: Design, Analysis, and Interpretation. W.W. Norton and Company.

Holland, P., 1986. Statistics and causal inference. J. Amer. Statist. Assoc. 81, 945–970.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47, 5–86.

Rubin, D., 1974. Estimating causal effects of treatments in randomized and non-randomized studies. J. Educ. Psychol. 66, 688–701.

Rubin, D.B., 1978. Bayesian inference for causal effects: the role of randomization. Ann. Statist. 6, 34–58.

Rubin, D.B., 1990. Formal modes of statistical inference for causal effects. J. Statist. Plann. Inference 25, 279–292.