

Spatial Differencing: Estimation and Inference

Federico Belotti, Edoardo Di Porto & Gianluca Santoni

Highlights

- Spatial differencing is a spatial data transformation pioneered by Holmes (1998) increasingly used to deal with omitted variable bias generated by local or site-specific unobservables in a "boundary-discontinuity" design setting.
- However, the correlation between all differenced observations that share a common sample unit makes inference problematic.
- We show that the dyadic-robust variance matrix estimator proposed by Cameron and Miller (2014) is, in general, a better solution compared to the most commonly used estimators.
- A Stata command implementing the methods described in this paper is available. It can be installed from within Stata by typing `net install sreg`, from (<http://www.econometrics.it/stata>).



Abstract

Spatial differencing is a spatial data transformation pioneered by Holmes (1998) increasingly used to estimate causal effects with non-experimental data. Recently, this transformation has been widely used to deal with omitted variable bias generated by local or site-specific unobservables in a "boundary-discontinuity" design setting. However, as well known in this literature, spatial differencing makes inference problematic. Indeed, given a specific distance threshold, a sample unit may be the neighbor of a number of units on the opposite side of a specific boundary inducing correlation between all differenced observations that share a common sample unit. By recognizing that the spatial differencing transformation produces a special form of dyadic data, we show that the dyadic-robust variance matrix estimator proposed by Cameron and Miller (2014) is, in general, a better solution compared to the most commonly used estimators.

Keywords

Spatial Differencing, Boundary Discontinuity, Robust Inference, Dyadic Data.

JEL

C12, C21.

Working Paper

CEPII

CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) is a French institute dedicated to producing independent, policy-oriented economic research helpful to understand the international economic environment and challenges in the areas of trade policy, competitiveness, macroeconomics, international finance and growth.

CEPII Working Paper
Contributing to research in international economics

© CEPII, PARIS, 2017

All rights reserved. Opinions expressed in this publication are those of the author(s) alone.

Editorial Director:
Sébastien Jean

Production:
Laure Boivin

No ISSN: 1293-2574

CEPII
113, rue de Grenelle
75007 Paris
+33 1 53 68 55 00

www.cepii.fr
Press contact: presse@cepii.fr

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY



Spatial differencing: estimation and inference¹

Federico Belotti* Edoardo Di Porto† Gianluca Santoni‡

¹We thank participants at the 1st CESifo Economic Studies Conference On the Use of Geocoded Data in Economic Research for their critics and comments. All remaining errors are ours.

*Corresponding author. Department of Economics and Finance, University of Rome Tor Vergata. E-mail: federico.belotti@uniroma2.it

†University of Naples Federico II, CSEF and UCFS Uppsala University. E-mail: edoardo.diporto@unina.it

‡CEPII. E-mail: gianluca.santoni@cepii.fr

1. Introduction

In this paper, we provide a practitioners' corner on spatial differencing estimation with a special focus on inference. Spatial differencing is a spatial data transformation pioneered by Holmes (1998) increasingly used thereafter to deal with omitted variable bias generated by local or site-specific unobservables arbitrarily correlated with the explanatory variables. This approach is at the heart of the so-called "boundary-discontinuity" design, a regression discontinuity design, more and more popular in public and urban economics literature, exploiting the spatial discontinuity that occurs at administrative boundaries to identify the treatment effect of interest. In this framework, spatial differencing consists in taking the difference between each sample unit located near a specific (administrative) area boundary and any other neighboring unit that lies on the other side of that boundary at a distance less than a specified threshold. Under the assumption that site-specific unobservables vary smoothly across space, i.e. very close units share the same site-specific unobservables, this transformation allows to rule out these unobservables from the estimating equation. In presence of panel data, it is even more appealing given that it can be applied after the standard within-group transformation to partial out site-specific time-varying unobservables (Duranton et al., 2011). Typical applications include studies of the effects of school quality on house prices (Black, 1999, Fack and Grenet, 2010, Gibbons et al., 2013, Kortelainen et al., 2014), the effect of local taxes on firm performance (Belotti et al., 2016), the effects of tax policies at county level (Chirinko and Wilson, 2008), the evaluation of area-based interventions (Einio and Overman, 2016), the effect of pollution havens on cancer deaths (Kahn, 2004).

However, spatial differencing introduces some concerns for statistical inference. Indeed, given a specific distance threshold, a sample unit may be the neighbor of a number of units on the opposite side of a specific boundary. This induces correlation between all differenced observations that share a common sample unit. Under homoskedasticity, the best solution to perform valid inference is represented by the analytically corrected variance matrix proposed by Duranton et al. (2011). As reported by Kortelainen et al. (2014) in this special issue, an alternative solution is represented by the one-way cluster-robust variance matrix at the boundary level, which allows for arbitrary correlation between all differenced observations on either side of a given boundary as well as for heteroskedasticity-robust inference. By recognizing that the spatial differencing transformation produces a special form of dyadic data, we show that the dyadic-robust variance matrix estimator proposed by Cameron and Miller (2014) is, in general, the right choice in order to perform valid inference. By running Monte Carlo simulations we compare, in the context of a "boundary-discontinuity" cross-sectional design, the finite sample properties of the dyadic-robust variance matrix estimator with those of the aforementioned

commonly employed estimators. The results indicate that dyadic-robust standard errors are well calibrated, outperforming all the alternatives in almost all considered design.

The remainder of the paper proceeds as follows. After a review of the spatial differencing estimation approach in the context of a “boundary-discontinuity” cross-sectional design in Section 2, Section 3 presents the set-up and the results of a set of Monte Carlo experiments. Finally, Section 4 concludes.

2. Spatial Differencing

2.1. Estimation

Let us consider a sample of N units drawn at random from a given population. The statistical model for the data is a linear model with local (or site-specific) unobserved effects

$$y_{ic} = \theta_z + \mathbf{x}'_{ic}\boldsymbol{\beta} + \epsilon_{ic}, \quad (1)$$

where y_{ic} is the outcome of unit i located in area c with $c = 1, \dots, C$, \mathbf{x}_{ic} is a k -vector of exogenous covariates, ϵ_{ic} is the idiosyncratic error and θ_z is an unobserved local effect for the unobserved location z , $z = 1, \dots, Z$, possibly at a finer spatial scale than c . Estimating (1) by ordinary least squares (OLS) ignoring θ_z gives a consistent estimate of $\boldsymbol{\beta}$ only if $\mathbb{E}(\theta_z|\mathbf{x}_{ic}) = 0$. If we set aside the latter unrealistic assumption by allowing for arbitrary correlation between the local unobservables and the explanatory variables, i.e. $\mathbb{E}(\theta_z|\mathbf{x}_{ic}) \neq 0$, a non-experimental approach to estimating equation (1) involves, in some way, transforming the data to rule out θ_z . An increasingly common way to deal with this issue in a “boundary-discontinuity” cross-sectional design is the so-called “spatial differencing” approach (Holmes, 1998).

Given a certain distance threshold d , denote with S_i^d the set of unit i 's neighbors lying on the other side of the area c 's boundary where i is located, and with $\Delta_{ij}^d = \mathbf{x}_{ic} - \mathbf{x}_{jc'}$ the spatial difference operator, which takes the difference between each unit i located in c and any other unit j located in c' (at a distance less than d from unit i), with c and c' being two contiguous areas. Under the assumption that site-specific effects vary smoothly across space and $d = d^*$, by applying the spatial difference operator to (1) gives

$$\Delta_{ij}^d y_{ic} = \Delta_{ij}^d \mathbf{x}'_{ic} \boldsymbol{\beta} + \Delta_{ij}^d \epsilon_{ic}. \quad (2)$$

In other words $\Delta_{ij}^d \theta_z \approx 0$ if d is small enough to allow the site-specific unobserved effect hitting two units located at a distance less than d to be (approximately) the same. Given model (2),

the SD estimator can be defined as

$$\hat{\boldsymbol{\beta}}_{SD} = \left[\sum_{i=1}^n \sum_{j \in S_i^d} \Delta_{ij}^d \mathbf{x}_{ic} \Delta_{ij}^d \mathbf{x}_{ic} \right]^{-1} \sum_{i=1}^n \sum_{j \in S_i^d} \Delta_{ij}^d \mathbf{x}_{ic} \Delta_{ij}^d y_{ic}. \quad (3)$$

Clearly, under the assumption that $E(\epsilon_{ic} | \mathbf{x}_{ic}, \theta_z) = 0$ and that θ_z is smooth over space, there exists $\hat{\boldsymbol{\beta}}_{SD}$ such that, as $N \rightarrow \infty$, $\hat{\boldsymbol{\beta}}_{SD}$ is consistent for $\boldsymbol{\beta}$, and $N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{SD} - \boldsymbol{\beta}) \sim \mathcal{N}(0, V)$.

2.2. Inference

While solving the main identification issue related to the presence of local unobservables correlated with the regressor(s) of interest, spatial differencing introduces some problems for statistical inference. The estimator of the asymptotic variance matrix, say \hat{V} , needs indeed to account for the fact that the error of unit i enters the error of m_i pairs, where m_i is the number of unit i 's neighbours. By recognizing that the spatially differenced data is a special form of dyadic data, we propose to use the dyadic-robust estimator described in Cameron and Miller (2014). To clarify why this estimator should perform better than others, consider the random sample $\{(x_i, y_i), i = 1, \dots, 4\}$ reported in Table 1

Table 1 – Original data.

id	y	x	c
1	-1.83	0.37	1
2	-0.71	0.65	2
3	0.56	0.03	3
4	-1.23	0.68	4

Suppose that the optimal distance $d^* = 1\text{km}$, i.e. by applying the spatial differencing transformation one is able to completely rule-out local unobservables, and that $S_1^1 = \{2, 3\}$, $S_2^1 = \{1, 3\}$, $S_3^1 = \{1, 2, 4\}$ and $S_4^1 = \{3\}$. Then, the spatially paired data will be as reported in Table 2.

It is worth noting that the number of unit-pairs observations are strictly less than $n(n-1)/2$, i.e. the number of pairs in presence of fully dyadic data.² In this example, there are 4 unit-pair observations ($< (4(4-1)/2) = 6$), namely (1,2), (1,3), (2,3) and (3,4), among which we do not include pairs (i, j) for which $i > j$ to avoid duplications. If we apply the spatial operator to the paired data reported in Table 2, we will obtain the spatial differenced data reported in Table 3.

²The only (unrealistic) case in which the spatially paired data is fully dyadic is when each unit in the sample is a neighbor of all the others, given a specific distance threshold d . In this case the number of unit-pairs observations will be equal to $n(n-1)/2$.

Table 2 – Spatially paired data at $d = 1\text{km}$.

pair	id	y	x	c
1	1	-1.83	0.37	1
1	2	-0.71	0.65	2
2	1	-1.83	0.37	1
2	3	0.56	0.03	3
3	2	-0.71	0.65	2
3	3	0.56	0.03	3
4	3	0.56	0.03	3
4	4	-1.23	0.68	4

Table 3 – Spatially differenced data (at $d = 1\text{km}$).

pair	g	h	Δy	Δx	c
1	1	2	-1.12	-.28	1
2	1	3	-2.39	.34	1
3	2	3	-1.27	.62	2
4	3	4	1.79	-.65	3

As can be noted by looking at Table 4, clustering at unit-pair level (up in Table 4), which coincides with using the White (1980)'s heteroskedastic robust variance estimator, only controls for correlation when $(g, h) = (g', h')$. The case of one-way clustering on the area boundary c (c in Table 4), controls for correlation when $g = g'$ but only if g is located in c . Two-way clustering on g and on h , additionally controls for correlation when $g = g'$ and/or $h = h'$ (2way in Table 4). Finally, dyadic clustering additionally picks up cases where $g = h'$ or $h = g'$ (dyad in Table 4), allowing to obtain a heteroskedastic robust variance estimator accounting at the same time for all potential correlations.

Table 4 – Correlations for the spatially differenced data.

$(g, h) / (g', h')$	(1,2)	(1,3)	(2,3)	(3,4)
(1,2)	up,c	2way,c	dyad	0
(1,3)	2way,c	up,c	2way	dyad
(2,3)	dyad	2way	up,c	dyad
(3,4)	0	dyad	dyad	up,c

The variance estimator has in all cases the classical sandwich form, that is $\hat{V} = aA^{-1}BA^{-1}$, with $A = \left[\sum_{i=1}^n \sum_{j \in S_i^d} \Delta_{ij}^d \mathbf{x}'_{ic} \Delta_{ij}^d \mathbf{x}_{ic} \right] = \Delta X' \Delta X$, and a the finite sample adjustment. In the case of the heteroskedasticity-robust White (1980) variance matrix, we have that $a = \frac{N}{N-k}$ and

$$B = \sum_i \sum_{j \in S_i^d} \Delta_{ij}^d \hat{u}_{ic}^2 \Delta_{ij}^d \mathbf{x}'_{ic} \Delta_{ij}^d \mathbf{x}_{ic}.$$

As for the one-way cluster-robust variance matrix, clustered at c -level, $a = \frac{C}{C-1} \frac{N-1}{N-k}$ and

$$B = \sum_c \sum_i \sum_{j \in S_i^d} \Delta_{ij}^d \mathbf{x}'_{ic} \Delta_{ij}^d \hat{u}_{ic} \Delta_{ij}^d \hat{u}'_{ic} \Delta_{ij}^d \mathbf{x}_{ic}.$$

The two-way cluster-robust variance matrix, clustered at g and h -level, is $\hat{V} = A^{-1}BA^{-1}$ with

$$\hat{B} = a_1 \sum_i \sum_{j \in S_i^d} 1(g_i = g'_j) P_{ij} + a_2 \sum_i \sum_{j \in S_i^d} 1(h_i = h'_j) P_{ij} - a_3 \sum_i \sum_{j \in S_i^d} 1[(g_i, h_i) = (g'_j, h'_j)] P_{ij}, \quad (4)$$

where $P_{ij} = \Delta_{ij}^d \hat{u}_{ic} \Delta_{ij}^d \hat{u}_{jc} \Delta_{ij}^d \mathbf{x}'_{ic} \Delta_{ij}^d \mathbf{x}_{jc}$, $a_1 = \frac{G-1}{G-2} \frac{N}{N-k}$, $a_2 = a_1$ and $a_3 = \frac{N}{N-k}$, with G the number of units effectively used in the estimation which, due to the spatial differencing at a given distance, is usually smaller than N . The dyadic-robust estimator has $a = a_1$ and

$$\hat{B} = \sum_i \sum_{j \in S_i^d} 1(g_i = g'_j | g_i = h'_j | h_i = g'_j | h_i = h'_j) P_{ij}.$$

Finally, the analytically-corrected variance matrix estimator proposed by Duranton et al. (2011) adapted for the cross-sectional case is

$$\hat{V} = \hat{\sigma}_2 A^{-1} B A^{-1}$$

with

$$B = \Delta X' \Delta \Delta' \Delta X$$

$$A = \Delta X' \Delta X.$$

and

$$\hat{\sigma}^2 = (tr(\Delta \Delta') - tr(AB))^{-1} \sum_i \sum_{j \in S_i^d} \Delta_{ij}^d \hat{u}'_i \Delta_{ij}^d \hat{u}_i$$

The main feature of this estimator is that it allows for an analytical correction of the variance matrix exploiting the spatial difference matrix Δ . In the cross-sectional case the latter has the following structure

$$\begin{pmatrix} \dots & 1 & 0 & 0 & -1 & 0 & 0 & 0 & \dots \\ \dots & 0 & 1 & 0 & 0 & -1 & 0 & 0 & \dots \\ \dots & 0 & 0 & 1 & 0 & 0 & -1 & 0 & \dots \\ \dots & 0 & 0 & 0 & 1 & 0 & 0 & -1 & \dots \end{pmatrix}$$

where each column can contain the values 1 and -1 several times depending on the number of times a unit has been matched with neighbours. The main drawback is that it is not robust to heteroskedasticity.

Differently from the one-way cluster-robust method where asymptotic results require that the number of clusters $C \rightarrow \infty$, the dyadic-robust approach requires $G \rightarrow \infty$. Thus, the dyadic-robust approach can be very useful when the number of cluster C is small but the number of units G used in the estimation is not so small, as is common in empirical applications. It is well-known that the Wald tests based on one-way (and two-way) cluster-robust standard errors can over-reject when there are few clusters (say, less than 10, see Cameron and Miller, 2015, for a detailed discussion) and the same is true for the dyadic-robust standard errors when there are few units (say, less than 100, see Cameron and Miller, 2014, for a detailed discussion). In the next Section we shall consider finite-cluster and finite-unit issues in some detail.

3. Monte Carlo Evidence

3.1. Set-up

By theory, the coefficient estimate of a 95% confidence interval should contain the true coefficient value in 95 out of 100 cases. However, when a variance estimator is not perfectly calibrated, the related standard error estimates are biased and statistical tests (such as the t -test) lose their validity. In what follows, we use the coverage rate (CR) of the 95% confidence interval to assess the finite sample properties of the five variance estimators reported in Section 2.2.³

To simulate the data, we consider the cartographic boundary shapefile of Kansas at county level.⁴ We start by over-imposing a fishnet of 7×3 rectangular cells. Then, a buffer of 15km radius was placed around the centroid of each rectangular cell and n random points were created within each of these buffers, which in our simulated space represent the $Z = 21$ unobserved

³Clearly $(1 - CR^{95\%})$ gives the size of the statistical test, i.e. $\alpha = 5\%$.

⁴We have chosen the Kansas US states for the regularity of its county's polygons. The Kansas state (with its 105 counties) has been selected from the complete US shapefile at county level, the latter downloaded at http://www2.census.gov/geo/tiger/GENZ2016/shp/cb_2016_us_county_20m.zip.

locations.⁵ In this way, the spatial difference transformation at $d \leq 30km$ will pick up only units with the same θ_z , thus ensuring the consistency of the spatial difference estimator in (3). We used $n = (25; 50)$, which creates two samples with size $N = (525; 1050)$. Finally, we select only the random points completely within the cartographic boundaries (losing 3 and 16 points, respectively) and only the counties effectively containing the random points. The final spaces are rendered in Figure 1 and 2. They include, respectively, a total of 522 and 1034 units, 82 and 88 counties, and 21 “unobserved” locations.

We then consider the following data generating process (DGP) with just one regressor

$$y_{ic} = \theta_z + x_{ic}\beta + \exp(\delta x_{ic})\epsilon_{ic}, \quad \begin{cases} i = 1, \dots, N \\ c = 1, \dots, C \\ z = 1, \dots, Z \end{cases} \quad (5)$$

with $N = 522, 1034$, $C = 82, 88$, $Z = 21$. θ_z is assumed to be correlated with x_{ic} . For simplicity but without loss of generality, we assume that

$$x_{ic} = \phi\theta_z + (1 - \phi^2)^{1/2}z_{ic}, \quad (6)$$

with ϵ_{ic} and z_{ic} being standard Gaussian i.i.d. random variables, independently distributed from θ_z , the latter itself is simulated as a standard Gaussian i.i.d. random variable. Thus, the covariance of x_{ic} and θ_z is proportional to ϕ ; when $\phi = 0$, x_{ic} and θ_z are uncorrelated, and θ_z is a Standard Gaussian random variable. The DGP in equation (5) allows to consider homoskedastic errors ($\delta = 0$) as well as different degrees of heteroskedasticity according to the value of the parameter δ . For each space, that is for each sample size, we set $\beta = 1$ and $\phi = 0.5$. We then consider $d = 5km, 10km, 15km, 20km, 30km$ and $\delta = 0, 0.25, 0.5, 1$ for a total of 40 different experiments. Simulations are based on 1000 replications for each experiment.

To summarize, the Monte Carlo simulations for each of the 40 parameter settings defined above proceed as follows:

1. Generation of 8 cross-sectional datasets with $N = 522, 1034$ and $\delta = 0, 0.25, 0.5, 1$ as specified above;
2. Estimation of the spatial difference estimator in (3) using different distance thresholds, i.e. $d = 5km, 10km, 15km, 20km, 30km$;
3. After having replicated steps (1) and (2) 1000 times, the CRs of the 95% confidence

⁵We used the `create fishnet` and `create random points` toolboxes of ARCGIS 10.3.1 for Desktop.

intervals for all five standard error estimates are gathered. This is achieved by obtaining the fraction of times the nominal 95%-confidence interval for $\hat{\beta}$ contains the true coefficient value of $\beta = 1$.

3.2. Results

Results of the simulations are presented graphically in Figures 3 and 4 when the sample size is $N = 522$, and in Figures 5 and 6 for $N = 1034$.⁶

Before going through the main findings of our analysis it is relevant to highlight that the distance threshold used to spatially differentiate the data dramatically affects the characteristics of the estimation sample. For instance, when the original sample size is $N = 522$ ($C = 81$), a distance threshold of $5km$ implies that only 38 units (located in $C = 16$ administrative areas) will be considered for spatial differencing, producing a sample of 23 pairs. On the other hand, when the threshold increases to $10km$, 153 units (located in $C = 52$ administrative areas) will be included in the estimation sample, increasing the sample size to 123 pairs. This has important implications for the interpretation of the results. Indeed, when the sample size grows, a larger number of correlated observations will be included in the sample affecting in different ways the properties of the considered variance matrix estimators.

As expected, when errors are homoskedastic, the Duranton et al. (2011) estimator, the benchmark in this case, delivers CRs in line with their nominal value, irrespective of the distance threshold used to spatially differentiate the data, while the other way around is true for the less “robust” White (1980) standard errors, with CRs that are increasingly below the nominal value as the threshold increases. On the other hand, both one-way clustering (on c) and two-way clustering (on g and h) exhibit CRs that are below their nominal value (around 90%), with the latter being much less sensitive to the distance threshold. Interestingly, the one-way clustering CRs significantly deteriorate as the distance threshold grows. This rather contradictory result is due to the fact that the one-way clustered standard errors are not robust to the “two-way” and/or “dyadic” correlations that are gradually included in the estimation sample. As expected, due to the fact that the number of units included in the estimation is far below the rule of thumb suggested by Cameron and Miller (2014), also the dyadic-robust standard errors show a not perfectly calibrated CR (around 91%) when the distance threshold is $5km$. However, and contrary to the one- and two-way clustering cases, the dyadic-robust CRs converge to the nominal value as the distance threshold grows.

⁶A Stata command implementing the methods described in this paper is available. It can be installed from within Stata by typing `net install sreg, from(http://www.econometrics.it/stata)`. Please, notice that the command cannot be downloaded from the website but just directly installed through the Stata software.

As shown in Figure 3, the results change significantly when heteroskedasticity is present. For Duranton et al. (2011) standard errors, heteroskedasticity leads to CRs that are far below their nominal value, much worse when the distance threshold increases. On the other hand, what is striking is that the dyadic-robust CRs are perfectly calibrated even in presence of extreme heteroskedasticity and, except for the aforementioned case where the distance threshold is $5km$, it always outperforms the competitors.

Figure 4 contains a complementary representation of the results presented in Figure 3. Here, for each variance estimator considered in the analysis, the average standard error estimate from the simulation is divided by the standard deviation of the coefficient estimates. Since the standard deviation of the Monte Carlo estimated coefficients approximates the true standard errors as the number of replications grows, for a variance estimator to be unbiased this ratio should be close to one. Consistently with the findings from above, Figure 4 shows that dyadic-robust standard errors are very well calibrated even in presence of heteroskedasticity. The only competitor who does best being the Duranton et al. (2011) estimator under homoskedasticity.

Finally, Figures 5 and 6 show very similar results as the original sample size grows ($N = 1034$), further confirming the properties of the dyadic-robust variance estimator.

Summarizing, under the fairly general data generating processes described in Section 3.1, the results show that the finite sample performances of the considered approaches improve the greater the “robustness” of the variance estimation method to the correlation induced by the spatial differencing transformation, with the dyadic-robust variance matrix estimator being always the best in presence of heteroskedasticity, followed in order by two-way and one-way clustering.

4. Concluding Remarks

In this paper, we provide a practitioners’ corner on spatial differencing estimation, with a special emphasis on the issues affecting inference. Failure to properly control for the error correlation induced by spatial differencing can lead to greatly under-estimated standard errors and over-stated t -statistics. Differently from previous literature we argue that, since the spatial differencing transformation produces a special form of dyadic data, the dyadic-robust variance matrix estimator proposed by Cameron and Miller (2014) is, in general, a better solution compared to the commonly used approaches. In particular, the Monte Carlo analysis conducted in this paper suggest that the choice of the covariance matrix estimator is crucial for inference’s validity. The results confirm that, under homoskedasticity, the Duranton et al. (2011) “analytically corrected” variance estimator is always the right choice, even though the dyadic-

robust estimator shows comparable performance when there are enough (> 100) underlying units forming the dyads. The presence of heteroskedasticity changes significantly the ranking of the considered estimators and, with the exception represented by the case in which too few units are used to forming the dyads, dyadic-robust standard errors are always the best choice. By comparison, the other methods generally under-estimate the standard errors leading to coverage rates far below the nominal size.

References

- Belotti, F., di Porto, E., and Santoni, G. (2016). The effect of local taxes on firm performance: evidence from geo referenced data. Working Papers 2016-03, CEPII research center.
- Black, S. E. (1999). Do Better Schools Matter? Parental Valuation Of Elementary Education. *The Quarterly Journal of Economics*, 114(2):577–599.
- Cameron, A. C. and Miller, D. L. (2014). Robust inference for dyadic data. Technical report, University of California at Davis.
- Cameron, A. C. and Miller, D. L. (2015). A Practitioner’s Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–373.
- Chirinko, R. S. and Wilson, D. J. (2008). State investment tax incentives: A zero-sum game? *Journal of Public Economics*, 92(12):2362 – 2384. New Directions in Fiscal Federalism.
- Duranton, G., Gobillon, L., and Overman, H. G. (2011). Assessing the effects of local taxation using microgeographic data. *Economic Journal*, 121(555):1017–1046.
- Einio, E. and Overman, H. G. (2016). The (Displacement) Effects of Spatially Targeted Enterprise Initiatives: Evidence from UK LEGL. SERC Discussion Papers 0191, Spatial Economics Research Centre, LSE.
- Fack, G. and Grenet, J. (2010). When do better schools raise housing prices? evidence from paris public and private schools. *Journal of Public Economics*, 94:59 – 77.
- Gibbons, S., Machin, S., and Silva, O. (2013). Valuing school quality using boundary discontinuities. *Journal of Urban Economics*, 75:15 – 28.
- Holmes, T. J. (1998). The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders. *Journal of Political Economy*, University of Chicago Press, 106(4):667–705.
- Kahn, M. E. (2004). Domestic pollution havens: evidence from cancer deaths in border counties. *Journal of Urban Economics*, 56(1):51 – 69.
- Kortelainen, M., Saarimaa, T., and Harjunen, O. (2014). Best education money can buy? capitalization of school quality in finland. Working Papers 58, Government Institute for Economic Research Finland (VATT).
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

Figure 1 – Kansas counties with simulated random points ($N=522$, $Z=21$, $C=82$)

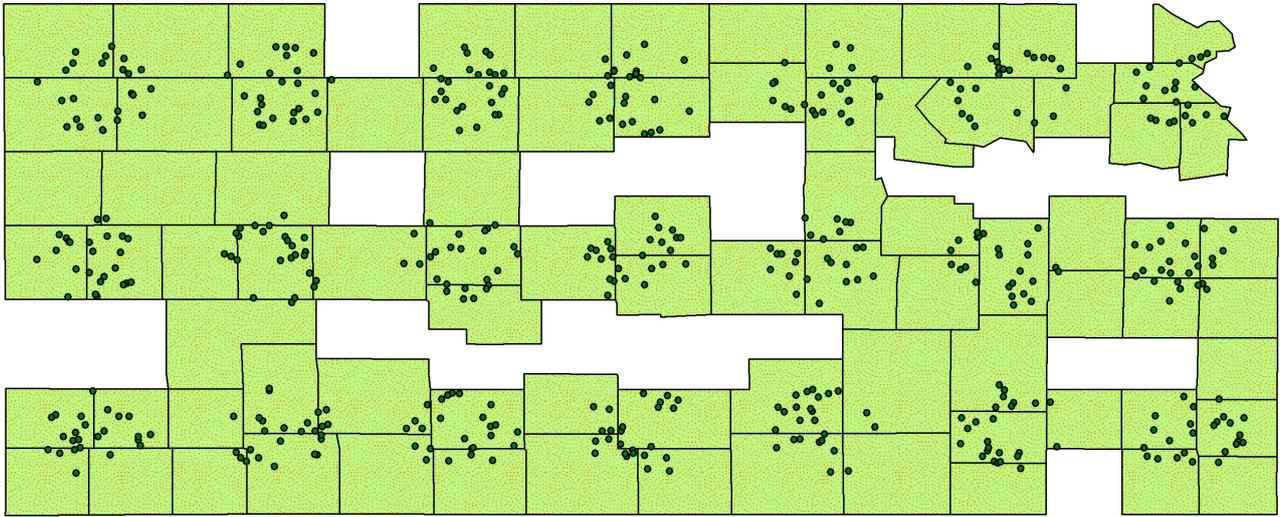
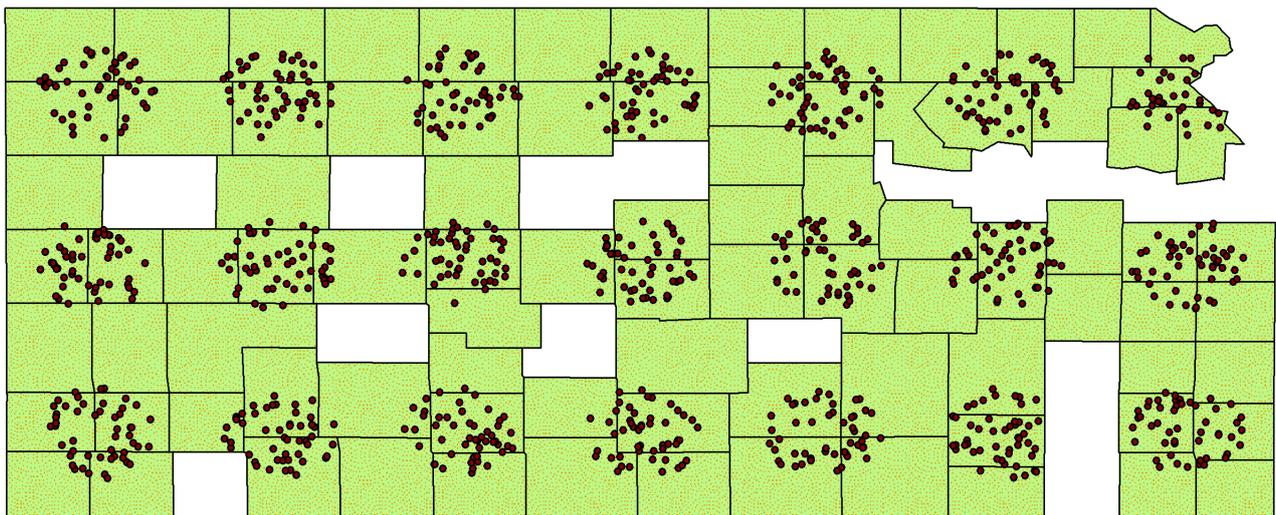


Figure 2 – Kansas counties with simulated random points ($N=1034$, $Z=21$, $C=88$)



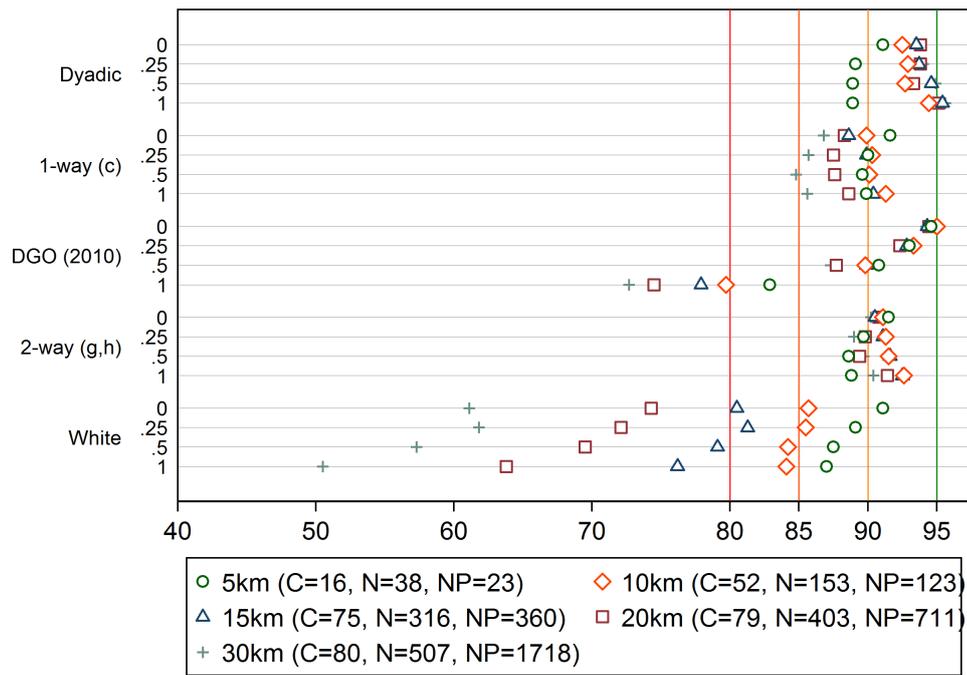


Figure 3 – Coverage rates of 95% confidence intervals: Comparison of different techniques for estimating standard errors. Monte Carlo simulation with 1,000 runs per parameter setting for a cross-section of N=522 units. The y-axis labels denote the values of the heteroskedasticity parameter δ .

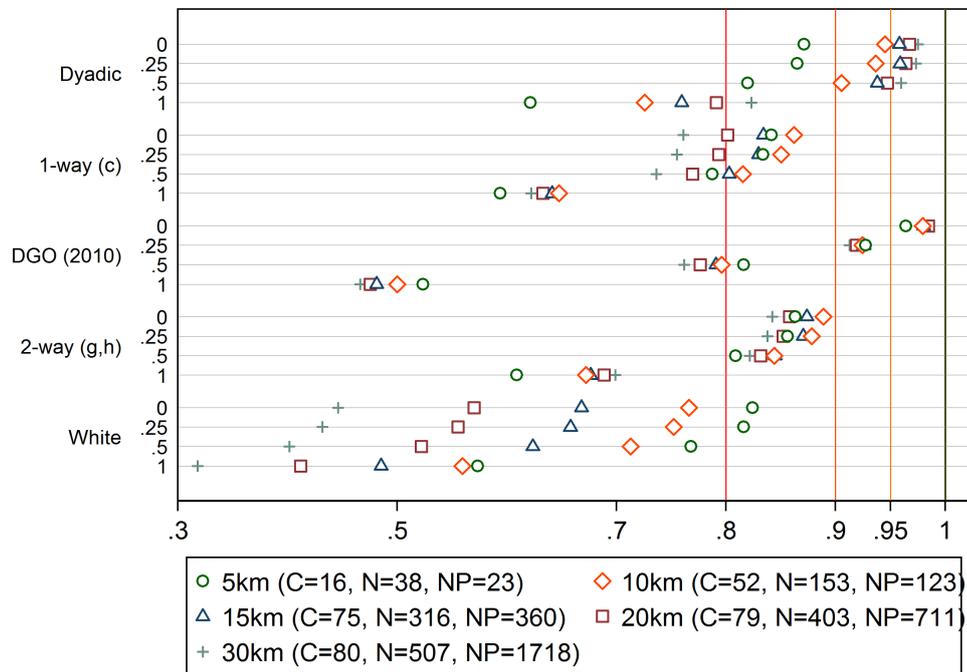


Figure 4 – Ratio of estimated to true standard errors: Monte Carlo simulation with 1,000 runs per parameter setting for a cross-section of N=522 units. The y-axis labels denote the values of the heteroskedasticity parameter δ .

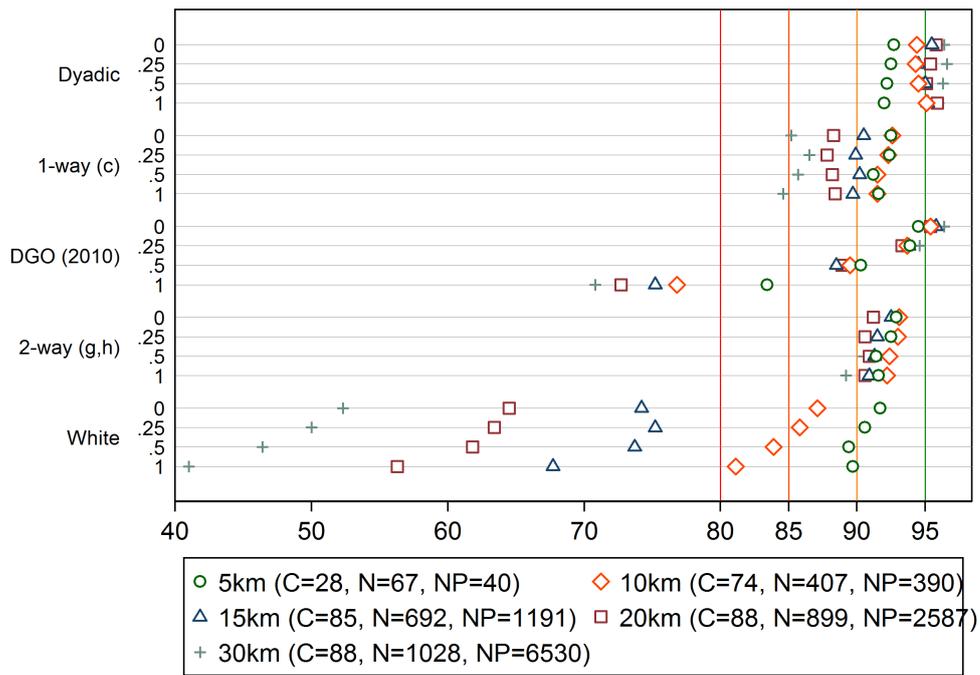


Figure 5 – Coverage rates of 95% confidence intervals: Comparison of different techniques for estimating standard errors. Monte Carlo simulation with 1,000 runs per parameter setting for a cross-section of $N=1034$ units. The y-axis labels denote the values of the heteroskedasticity parameter δ .

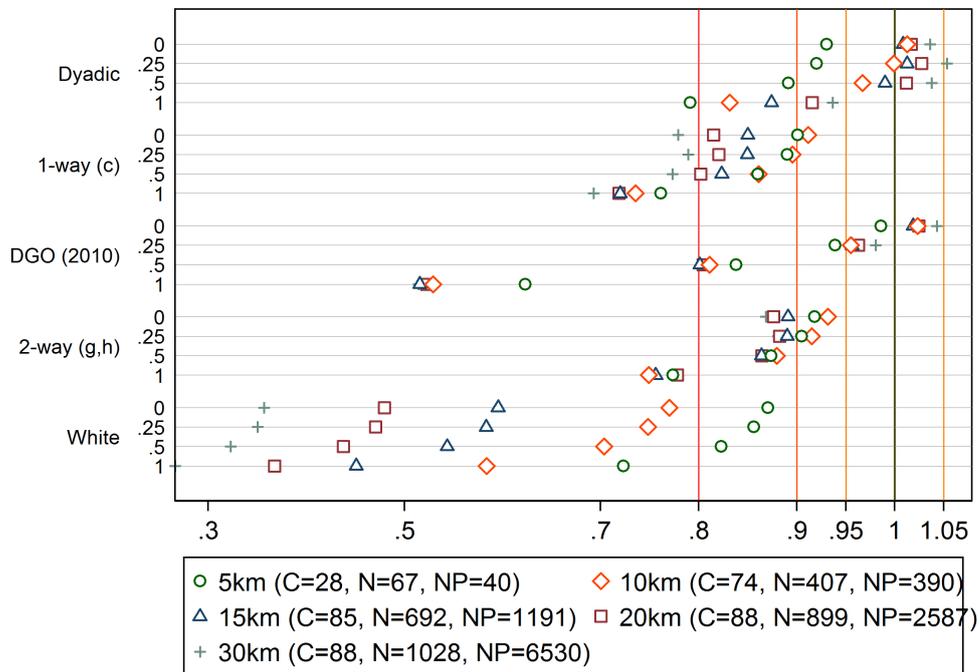


Figure 6 – Ratio of estimated to true standard errors: Monte Carlo simulation with 1,000 runs per parameter setting for a cross-section of $N=1034$ units. The y-axis labels denote the values of the heteroskedasticity parameter δ .