Online Appendix

The persistent effects of place-based policy:
Evidence from the West-German Zonenrandgebiet

Maximilian v. Ehrlich   Tobias Seidel

A   Data & Descriptives

Table A1 displays the number of observations by state as well as the area and population shares of the ZRG.

Table A1: CHARACTERISTICS OF THE ZONENRANDGEBIET

	No. districts		No. municipalities		Area ZRG	Pop. ZRG
	Non-ZRG	ZRG	Non-ZRG	ZRG	in %	in %
West Germany	382	90	6,839	1,573	18.6	12.3
Schleswig-Holstein	6	14	414	710	53.3	81.3
Lower Saxony	54	22	711	282	28.6	33.4
North-Rhine Westfalia	80	0	386	0	0	0
Hesse	33	12	323	96	27.8	19.1
Bavaria	112	42	1,533	485	25.8	21.9
Other West-German states	97	0	3,472	0	0	0

Notes: The states (Länder) Schleswig-Holstein, Lower Saxony, Hesse, and Bavaria belonged to the ZRG. We add North Rhine-Westphalia as it borders with the ZRG, but drop the city states of Hamburg and Bremen. The districts correspond to the 1971 classification while we use data from 1986 for the number of municipalities, and the year 1961 for population shares.

All districts that accommodated at least 50 percent of their area or population within 40 kilometers to the inner-German or Czechoslovakian border on January 1, 1971 became part of the Zonenrandgebiet.[19] A list of all 1971 districts that belonged to the ZRG is

---

[19]See Deutscher Bundestag (1970), Drucksache VI/796 and Ziegler (1992, p.9). According to a statement by state secretary Sauerborn, the 40-kilometer rule also included less needy regions, but was

contained in the federal law 'Zonenrandförderungsgesetz', 1971. Data on population by municipality is available from the Federal Statistical Office for the years 1975-2010. For the years 1950-1970 we have acquired the data from the Statistical Offices of the five Länder we consider. Likewise, information on income, business taxes, and employment was provided by the statistical offices of the individual states. Data on municipal area shares covered by private, public, commercial, industrial, and residential capital was provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development. The data on land prices are provided by the states' expert committees and by $F\&B$ real estate consulting.

Georeferencing. To georeference municipality data, we use digital maps (shape files) from the Bundesamt für Kartographie und Geodäsie. As they are only available since 1997, we assign each municipality to a district in 1971 and drop all observations where the municipality cannot be linked to a district with at least 90 percent of its area (20 municipalities or about 0.4 percent of the sample).[20] Moreover, we drop 31 municipalities due to partial treatment (i.e. the ZRG border crosses the municipality based on the 1997 or 2010 classification). This occurs due municipality mergers between municipalities being located on opposite sides of the ZRG border. In these cases the municipalities that were located outside the ZRG and merged with municipalities in the treatment area could not become eligible for transfers but the treatment border passes through the municipalities. Hence, the jurisdictional boundaries as of January 1, 1971 remained relevant for treatment throughout the duration of the program. The *boundary sample* of municipalities contains all jurisdictions with a distance to the ZRG border of less than 100 kilometers. This includes all municipalities in the treated region and about 68 percent of the municipalities in the five states west of the ZRG border. For the boundary sample at the district level, we limit the observations to jurisdictions that are sufficiently close to the threshold determining transfer eligibility i.e. $M_d \leq 150$. This includes again all treated observations and about 50 percent of the districts outside the treated area and in the five states. Note that all our analyses are based on the 1971

---

appealing for practical reasons in the first place (see Protocol of the 39th session of the cabinet committee of economics). It was recognized in the parliamentary debate on June 17, 1971 that the treatment border must remain fixed over time in order to rule out strategic modifications of local district borders (see Protocol of the 128th session of the Bundestag).

[20]This may happen due to changes in administrative boundaries that were frequent especially in the 1970s. Note that all our results are robust to the exclusion of *all* municipalities that could not perfectly be assigned to a 1971 district.

district classification such that the number of districts remains constant over time.

Depending on the municipality classification of each data source we assign it either to the 1997 or 2010 shape file of municipal boundaries. Thereby, we link 1971 districts (and thus treatment status), coordinates, and distances from the ZRG boundary to the outcome variables. Geospatial data processes have been performed and documented in ArcGIS using shape files about administrative boundaries in 1971, 1997, and 2010. Table A2 provides information about the data sources of all variables we use and the respective level of spatial aggregation.

Administrative & satellite data. Table A3 displays summary statistics for all variables in the boundary sample used in our analysis. *income* and *business taxbase* are measured in 1,000 euros and in current terms. Note that the business tax base is defined homogeneously across all municipalities in Germany. The average income per square kilometer was about 1.5 million euros during the transfer program and it increased to about 2.8 million euros in 2010. Note that averages across municipalities deviate from the country average because cities receive the same weight as small municipalities. Municipalities are the smallest administrative units comprising on average about 33 square kilometers with a high standard deviation and a minimum of 0.45 square kilometers while the largest municipality stretches over 359 square kilometers. The average number of inhabitants increased from about 5,900 in 1986 to 6,500 in 2010. Accordingly, the municipal average of population density reaches about 160 (178 in 2010) individuals per square kilometer which is well below the German average of about 220. Likewise, per capita income and employment density average at relatively low levels of about 25,000 euros and 40 employees per square kilometer. Note also that 19 (7) municipalities exist in 1986 (2010) that have no employment and taxable business income such that they will be dropped when specifying these outcomes in logarithmic terms.

If not stated differently, all variables refer to municipalities. *public capital* and *private capital* measure the area share of a municipality covered by public infrastructure (area used for streets, railway, airports, seaports, public squares, public buildings etc.) and private structures (industry, business, and housing), respectively. Note that *private capital* is the sum of *residential capital* and *industrial capital*. *human capital* refers to the share of residents with tertiary education. Land prices correspond to current prices per m^2 and stem from evaluations of expert committees.

44

Private capital and radiance are also summarized for grid cells of $100m \times 100m$ and 30×30 arc-seconds (about $900m \times 600m$ in the center of Germany), respectively. *Radiance* of grids is computed according to digital integer numbers reported by satellite data of the Defence Meteorological Satellites Program – Operational Linescan System (DMSP-OLS).[21] These data measure night-time lights in the year 1992 and are widely used in research (see, e.g., Elvidge et al., 1997; Henderson, Storeygard, and Weil, 2012). The information on *PrivateCapital* is provided by the European Environmental Agencies CORINE project for the year 1990. The data contains a variable that indicates 44 different land cover classes. We set *PrivateCapital* = 1 if a place is covered by 'Continuous urban fabric', 'Discontinuous urban fabric', 'Industrial or commercial units', or 'Construction sites' and zero otherwise. Note that the area-weighted sum of *PrivateCapital* is highly correlated with our municipality-level variable for private capital which is described in section 3 (correlation coefficient of 0.84). Moreover, the averages for the area share of private capital measured at the grid-cell and municipality levels are relatively close to each other even though the data sources differ. The deviation is possibly again due to the identical weight attached to municipalities of different size when summarizing the municipality level data.

---

[21]The satellite data report digital integer numbers ranging from 0 to 63. These may be converted to radiance as a measure of night luminosity by using the formula $radiance = digitalnumber^{1.5}$ for a spatial unit which is denoted in terms of Watts/cm^2/sr/nm (in words: Watts per squared centimeter per steradian per nanometer of wave length).
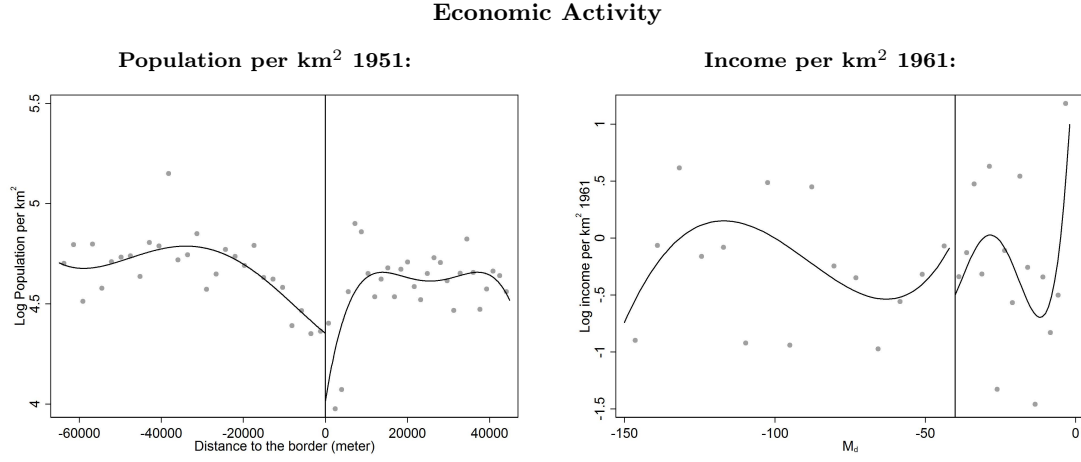
Table A2: Data sources

Variable	Spatial aggregation	Source
area	municipality/district	Shapefiles from Federal Institute for Research on Building, Urban Affairs and Spatial Development; Municipality data from Federal Statistical Office
income	municipality	Statistical Offices of the Länder
population	municipality	Federal Statistical Office; Statistical Offices of the Länder for pre 1970 data
employment	municipality	Statistical Offices of the five Länder
human capital	municipality	Federal Employment Agency
public capital	municipality/grid cell	Federal Office for Building and Regional Planning; European Environmental Agency
local public budget: -public investments -federal invest subsidies -local tax revenues	municipality	Statistical Offices of the Länder
private capital	municipality/grid cell	Federal Office for Building and Regional Planning; European Environmental Agency
business taxbase	municipality	Statistical Offices of the Länder
land price	municipality	States' expert committees; F&B real estate consulting
radiance	grid cell	NOAA Defence Meteorological Satellites Program
roads	municipality/grid cell	Openstreetmap.org

Table A3: Descriptive statistics of outcome variables

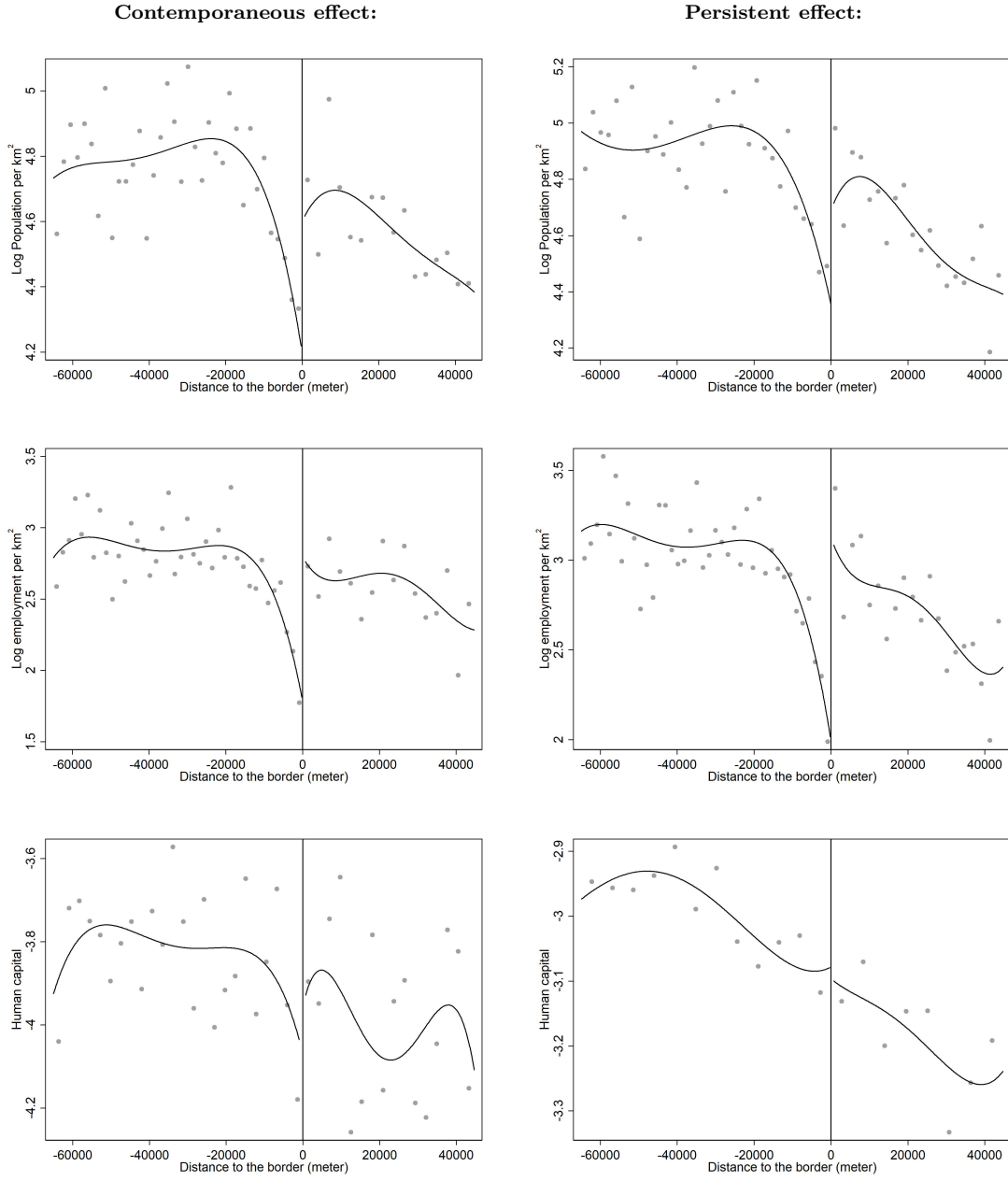|	Contemporaneous				Persistent			
	Mean	Std.dev.	Obs.	Year	Mean	Std.dev.	Obs.	Year
area in km^2	33.042	33.378	3,870	1986	33.167	33.572	3,881	2010
income/km^2	1,250.7	2,228.8	3,870	1986	2,848.0	4,701.9	3,881	2010
radiance (grid cell)	47.546	71.478	110,300	1992	112.420	107.019	110,300	2010
population	5,931.77	20,318.12	3,870	1986	6,530.91	21,848.77	3,881	2010
population/km^2	160.083	238.896	3,870	1986	178.333	258.123	3,881	2010
employment/km^2	39.268	99.807	3,826	1986	48.891	119.248	3,665	2010
income/capita	19.450	4.675	3,870	1986	31.461	7.247	3,881	2010
human capital	0.028	0.024	1,782	1985	0.061	0.048	2,576	2010
public capital	0.044	0.020	3,855	1984	0.051	0.023	3,865	2010
private capital	0.048	0.048	3,845	1984	0.067	0.057	3,851	2010
private capital (grid cell)	0.061	0.240	7,786,402	1990	0.074	0.262	7,786,402	2012
industrial capital	0.007	0.012	1,260	1988	0.008	0.010	1,234	2010
business taxbase/km^2	7.985	24.456	3,533	1986	18.568	77.828	3,792	2010
business taxbase/employee	0.201	0.403	3,667	1986	0.425	0.626	3,642	2010
land price	26.197	21.016	982	1988	74.188	70.601	3,635	2010
tax revenues/capita	305.459	135.204	2,650	1985	682.282	2492.996	3,169	2010
federal invest. subsidies/capita	67.998	102.889	2,650	1985	97.609	289.147	3,169	2010
public investment/capita	332.549	244.764	2,614	1985	571.427	567.357	3,171	2010

Notes: We consider the states (Länder) Schleswig-Holstein, Lower Saxony, North Rhine-Westphalia, Hesse, and Bavaria. We restrict the sample to observations located within 100km of the ZRG border. *income* and *business taxbase* are measured in current 1,000 euros, *human capital* refers to the share of residents with tertiary education, *public capital* and *private capital* measure the area share of a municipality covered by public infrastructure (area used for streets, railway, airports, seaports, public squares, public buildings etc.) and private structures (industry, business, and housing), respectively. *industrial capital* refers to the area share of a municipality covered by industry and business structures. Note that the latter is available for only three states. Land prices correspond to current prices per m^2.

47

Figure A1: PRE-TREATMENT BALANCING

Economic Activity

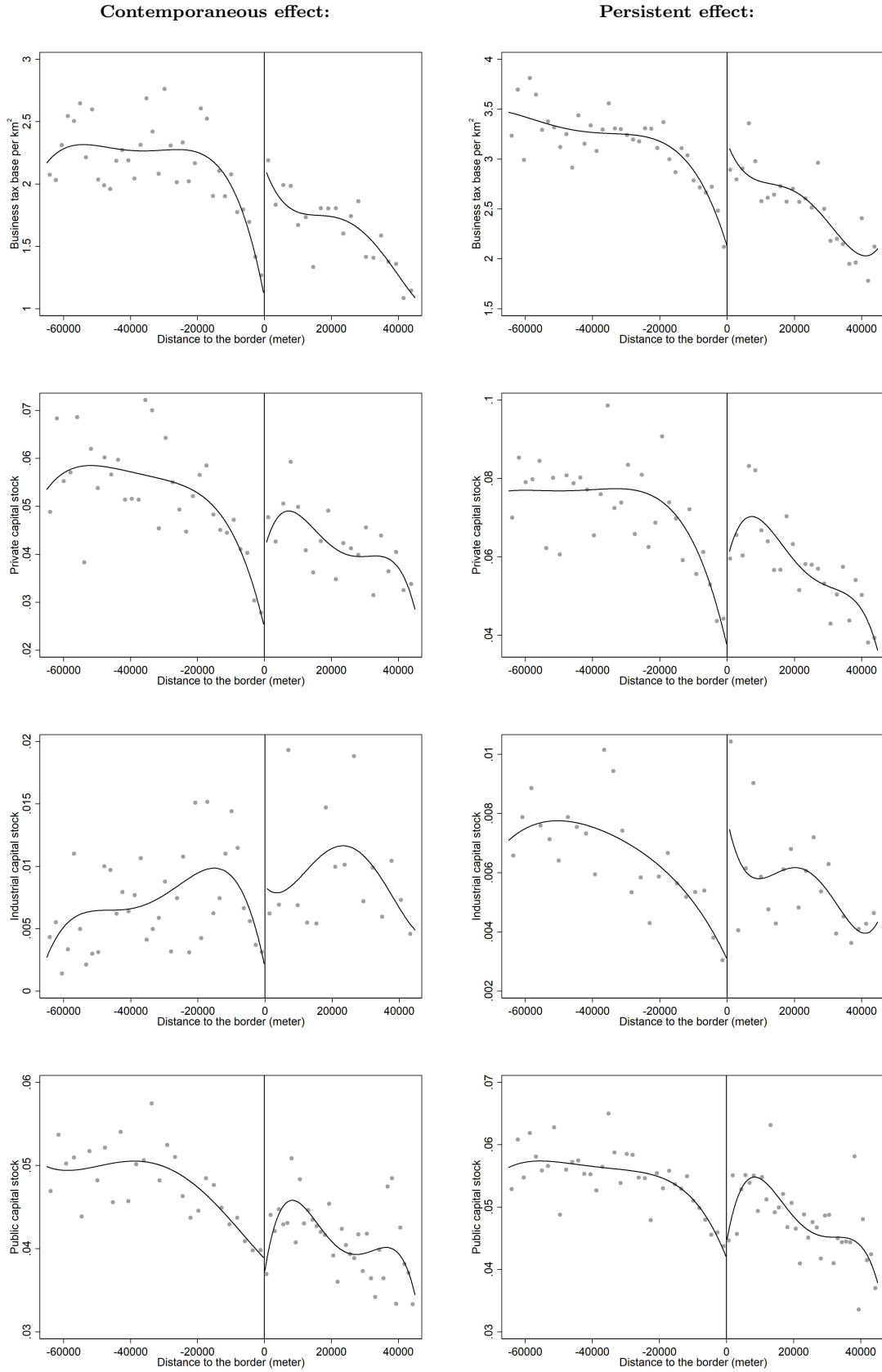**Population per km² 1951:**   **Income per km² 1961:**

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using non-overlapping evenly spaced bins on each side of the threshold following Calonico, Cattaneo, and Titiunik (2015). The left figure is based on municipality data with distance to the treatment border acting as the forcing variable in the spatial RDD. The right figure is based on district border and depicts income against the forcing variable in the fuzzy RDD, i.e. M_d (see chapter 4.2).

Figure A2: DISCONTINUITIES: LABOR

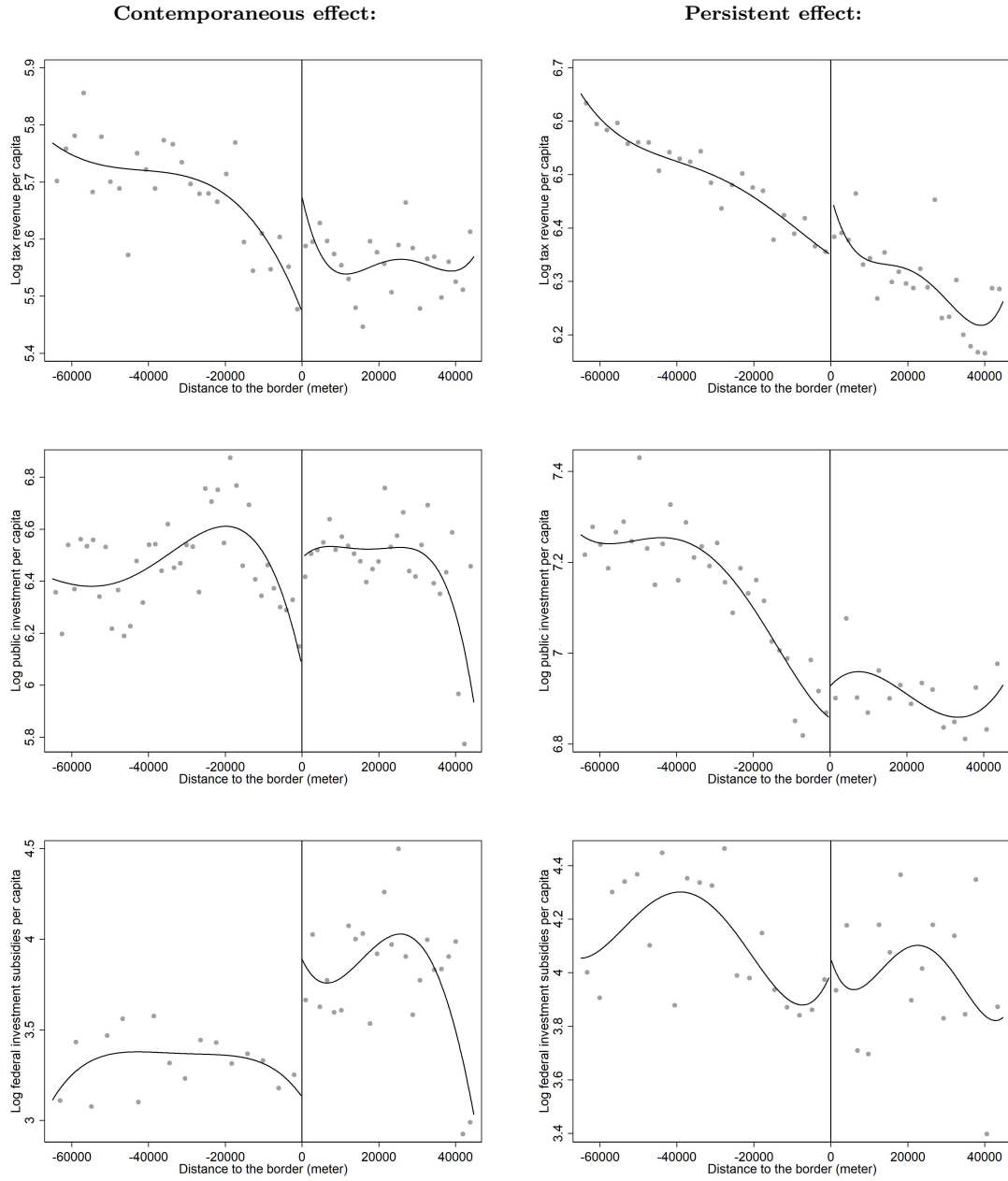**Contemporaneous effect:**   **Persistent effect:**

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using nonoverlapping evenly spaced bins on each side of the threshold following the data-driven method for optimal choice of the number of bins described in Calonico, Cattaneo, and Titiunik (2015). The lines represent a 4th-order polynomial distance control function for the 65km window.
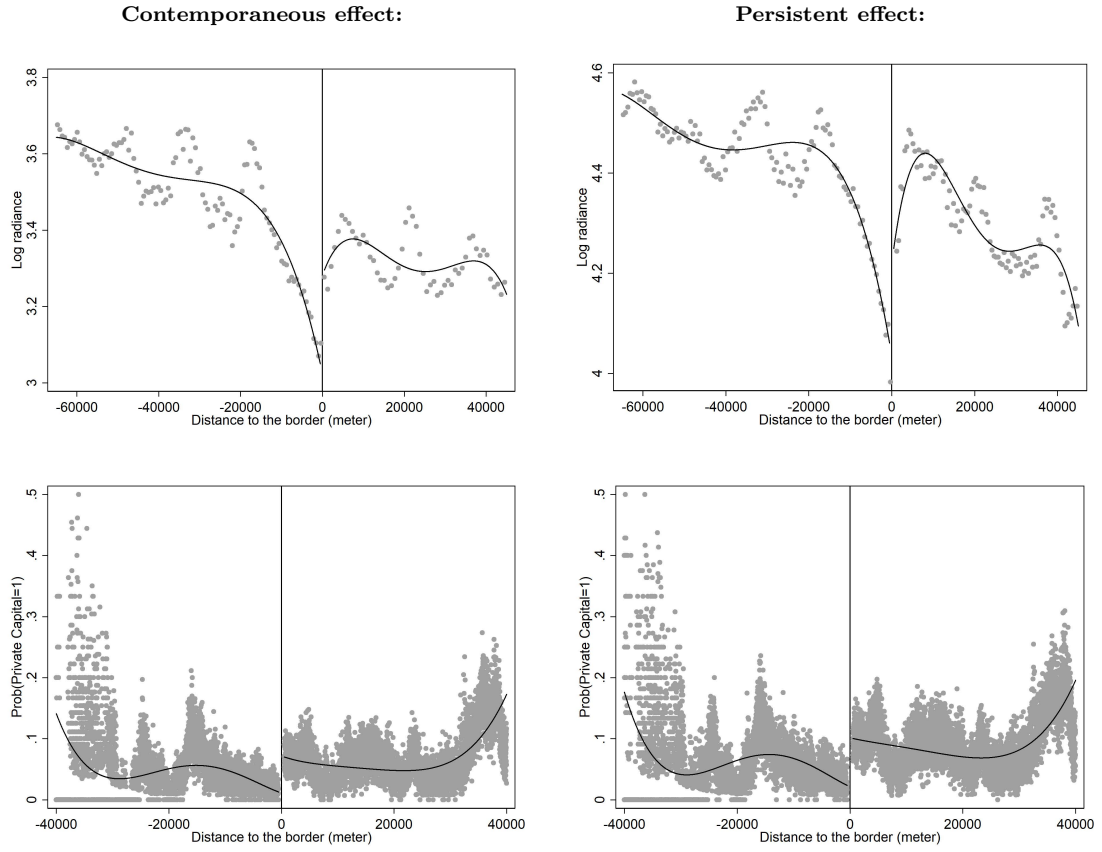
Figure A3: DISCONTINUITIES: CAPITAL

Contemporaneous effect: | **Persistent effect:**

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using nonoverlapping evenly spaced bins on each side of the threshold following the data-driven method for optimal choice of the number of bins described in Calonico, Cattaneo, and Titiunik (2015). The lines represent a 4th-order polynomial distance control function for the 65km window.

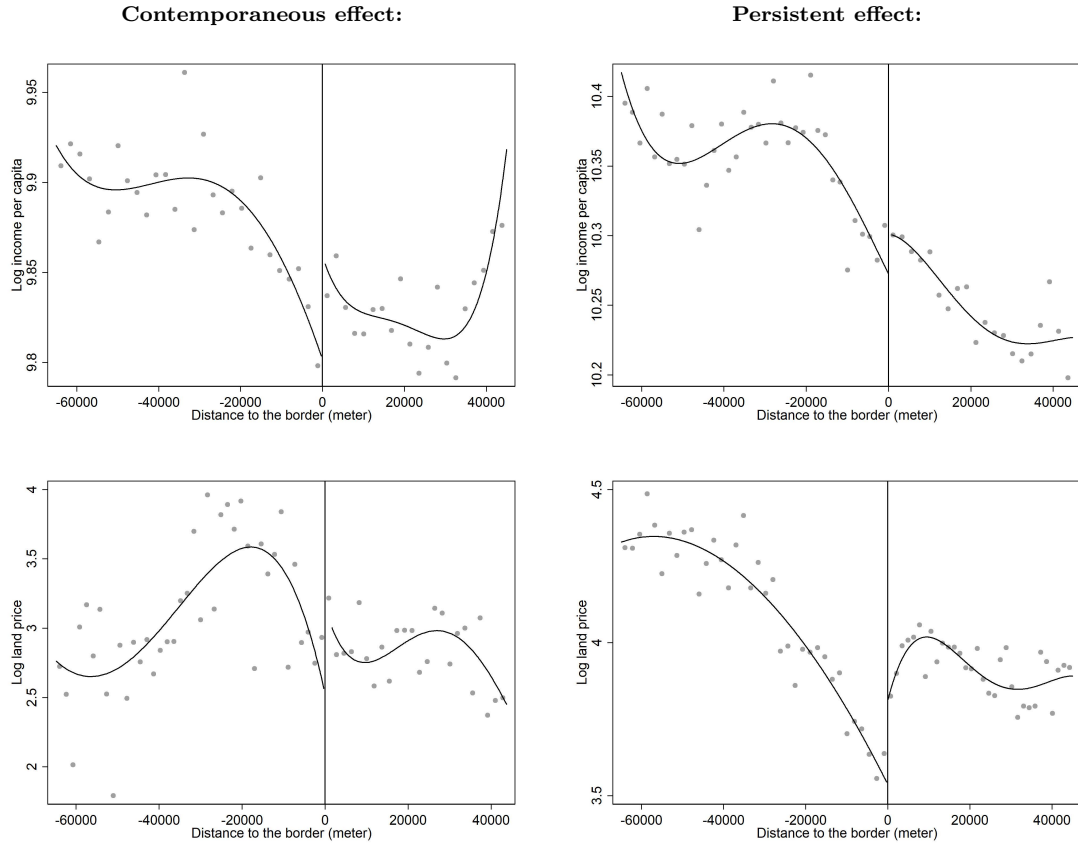Figure A4: DISCONTINUITIES: LOCAL PUBLIC INVESTMENT

Contemporaneous effect:

Persistent effect:

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using nonoverlapping evenly spaced bins on each side of the threshold following the data-driven method for optimal choice of the number of bins described in Calonico, Cattaneo, and Titiunik (2015). The lines represent a 4th-order polynomial distance control function for the 65km window.

Figure A5: Discontinuities: Grid Cell Data

Contemporaneous effect: **Persistent effect:**

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using nonoverlapping evenly spaced bins on each side of the threshold following the data-driven method for optimal choice of the number of bins described in Calonico, Cattaneo, and Titiunik (2015). The lines represent a 4th-order polynomial distance control function.

Figure A6: Discontinuities: Per-Capita Income and Land Prices

Contemporaneous effect: **Persistent effect:**

Notes: We run separate regressions on each side of the threshold. The plots represent local sample means using nonoverlapping evenly spaced bins on each side of the threshold following the data-driven method for optimal choice of the number of bins described in Calonico, Cattaneo, and Titiunik (2015). The lines represent a 4th-order polynomial distance control function.

B   Identification strategy

B.1   Spatial RDD

We denote by Y_{i0} and Y_{i1} the potential outcomes of a municipality i in the situations with and without transfers, respectively. Our aim is to identify the effect of a transfer T_i which corresponds to $\tau = Y_{i0} - Y_{i1}$. As counterfactual situations for individual units are unobservable, we aim at estimating an average treatment effect $E[\tau_i]$ for a group of comparable treated and control units. Our outset represents a special case of a two-dimensional RDD where the location of each municipality relative to the threshold is described by latitude and longitude, $\mathbf{L}_i = (L_{ix}, L_{iy})$. Similarly, the boundary between the treatment area $\mathcal{A}^+$ and the control area $\mathcal{A}^-$ consists of an infinite number of border points $\mathbf{b} = (b_x, b_y) \in \mathbf{B}$.

Due to the geographic nature of the policy measure, assignment to treatment is a discontinuous function of location, $T = 1\{\mathbf{L}_i \in \mathcal{A}^+\}$, where units east of $\mathbf{B}$ receive treatment while those to the west do not. In the spatial discontinuity design, location acts as the so-called forcing variable and we focus on the discontinuity of expected outcome at the geographical border:

$$\tau(\mathbf{b}) \equiv E[Y_{i1} - Y_{i0}|\mathbf{l} = \mathbf{b}] = \lim_{\mathbf{l}^+ \to \mathbf{b}} E[Y_i|\mathbf{L_i} = \mathbf{l}^+] - \lim_{\mathbf{l}^- \to \mathbf{b}} E[Y_i|\mathbf{L}_i = \mathbf{l}^-], \qquad \text{(B.1)}$$

where $\mathbf{l}^+ \in \mathcal{A}^+$ and $\mathbf{l}^- \in \mathcal{A}^-$ refer to locations in treated and control areas, respectively. Accordingly, $\tau(\mathbf{b})$ identifies the average treatment effect at the border point $\mathbf{b}$. In contrast to a one-dimensional regression discontinuity design, our approach yields a function of treatment effects evaluated at each border point $\mathbf{b} \in \mathbf{B}$. For most of our analysis we consider the average treatment effect along the whole border while we explore variations in the treatment effects across locations for sensitivity checks and to analyze the role of agglomeration externalities for persistence.[22]

We implement the spatial RDD both in a parametric and in a nonparametric way. In the former case we state the conditional expectations in (B.1) as $E[Y_{i0}|\mathbf{L}_i] = \alpha + f(\mathbf{L}_i) + g_0(D_i)$ and $E[Y_{i1}|\mathbf{L}_i] = \alpha + \tau + f(\mathbf{L_i}) + g_1(D_i)$ where $f(\mathbf{L_i})$ represents flexible polynomials of geographic location and D_i refers to the shortest distance from i's centroid to the treatment border ($\mathbf{B}$), i.e. the perpendicular to the closest border point. The inclusion of asymmetric distance control functions accounts for the possibility that proximity to the treatment border influences outcomes differently for transfer recipients and non-recipients.[23] Controlling for $\mathbf{L}_i$ may be important as units with the same distance to $\mathbf{B}$ may in fact be quite different if they are located in different parts of Germany (e.g.

---

[22]See Papay, Willett, and Murnane (2011) for treatment effect heterogeneity in a two-dimensional RDD and non-geographic context. Importantly, this design allows us to limit the estimation to border segments where frictions at municipality borders are likely to be absent.

[23]By presuming symmetric functions on both sides of the RDD threshold, a kink may be misinterpreted as a discontinuity (see Lee and Lemieux, 2010).

Table B1: DISTANCE & ASSIGNMENT VARIABLE

	ZRG				Non-ZRG			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Distance to **B** (D_i)	22,970	15,925	88	97,603	39,811	28,465	183	99,953
M_d	20.194	11.262	3	45	87.560	29.702	42	149

Notes: Distances are in meters and refer to municipality centroids. The assignment variable M_d is defined as the minimum distance (in km) from the Iron Curtain that includes the majority share of the district area. It is determined at the district level according to the 1971 classification. Each municipality is uniquely assigned to a district. Three districts received treatment although not being eligible according to the rule and thus generate fuzzyness. Of those districts being eligible according to the treatment rule all received treatment. We dropped all observations with a distance of more than 150km to the ZRG border and districts with $M_d > 150$.

north versus south or distance to the sea, state/country borders). Thus, the regression model is given by:

$$Y_i = \alpha + g_0(D_i) + f(\mathbf{L_i}) + T_i[\tau + g_1(D_i) - g_0(D_i)] + \varepsilon_i. \qquad (B.2)$$

Since $g_1(D_i) - g_0(D_i)$ converges to zero for observations close to the border, the average treatment effect is captured by $\hat{\tau}$. Since the credibility of the results rest on the correct specification of the control functions, we run alternative regressions with different functional forms (e.g. order of the polynomials), with and without coordinate control functions ($f(\mathbf{L_i})$), for different windows around the ZRG border, and we include border-segment fixed effects as well as state fixed effects.

The assumptions about the form of the geographic control functions can be further relaxed by estimating the treatment effect in a nonparametric way. To do so, we employ local linear regressions and estimate the conditional expectations at the border as stated in (B.1). Notice that we base our estimates for $E[Y_{i0}|\mathbf{L}_i = \mathbf{b}]$ and $E[Y_{i1}|\mathbf{L}_i = \mathbf{b}]$ only on observations in $\mathcal{A}^+$ and $\mathcal{A}^-$, respectively. As in the parametric approach, we condition on the forcing variable $D_{i\mathbf{b}}$ and estimate univariate local linear regressions for a set of 20 border points $\mathbf{b}^1, ..., \mathbf{b}^{20}$ which are allocated at equal distances along the border.[24] The corresponding results crucially depend on the choice of bandwidth.

Table B1 reports descriptive statistics on the distance of observations from **B**. Although the treated area corresponds mostly to a narrow band of 40 kilometers there are treated observations in the north-east (in particular on the island Fehmarn) located at a

---

[24]We check the sensitivity of our results with 10 and 30 border points. As an alternative approach we followed Papay, Willett, and Murnane (2011) using a bivariate nonparametric regression with the arguments L_{ix} and L_{iy}. Due to the "curse of dimensionality" bivariate local linear regressions require a much higher density of data. For this reason we favor the univariate nonparametric approach. However, all our results are robust to the bivariate nonparametric regression approach. See Appendix, Figure B1 for a more detailed description of the nonparametric specifications.

distance of up to 100 kilometers from the ZRG border. The closest municipal centroids lie at about 88 and 183 meters from **B** for the treatment and control groups, respectively.[25] Due to the nature of the transfer program the distance to the ZRG border is positively correlated with the distances to the Iron Curtain. However, as the location of the ZRG border is determined by the districts' shape, size, and location the correlation between distance to **B** and distance to the Iron Curtain is only about 0.6 for the boundary sample and reduces to less than 0.05 when we limit the sample to a 20-kilometer window from **B**. This points to an important advantage of our setting, namely the clear geographic criterion that defined the Zonenrandgebiet.

Non-parameteric identification. For the nonparametric identification strategy we resort to local linear regressions as these are particular well-suited for inference in the RDD (see Fan and Gijbels, 1996; Imbens and Lemieux, 2008). We employ a triangular kernel function and choose different bandwidths according to optimality criteria. We compute the distance of each municipality's centroid to 20 (or 30) border points that are allocated at equal distances along **B** as shown in Figure B1. Then, we assign municipalities to the closest border point, add border-point fixed effects, and use the distance to the respective border point in the local linear regressions to estimate $E[Y_{i0}|\mathbf{L}_i = \mathbf{b}]$ and $E[Y_{i1}|\mathbf{L}_i = \mathbf{b}]$. All our results are insensitive to choosing 20 or 30 border points (red dots and blue triangles in Figure B1, respectively).
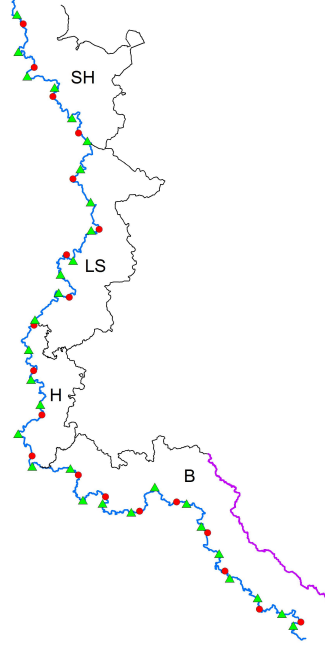
As a further robustness check, we refrain from allocating border points and estimate bivariate local linear regressions based on Cartesian coordinates. In this approach we use a product kernel $K_{h_x}(L_{ix} - L_{0x})K_{h_y}(L_{iy} - L_{0y})$ and minimize:

$$\sum_{i=1}^{n} \left\{ Y_i - \alpha - (L_{ix} - L_{0x})\beta_1 - (L_{iy} - L_{0y})\beta_1 \right\}^2 K_{h_x}(L_{ix} - L_{0x})K_{h_y}(L_{iy} - L_{0y}). \quad \text{(B.3)}$$

Again, this is done separately for units west and east of **B** to obtain $\hat{\alpha}$ for both sides. The pair of bandwidths is chosen according to a cross-validation criterion. In practise this approach is less efficient than the univariate approach based on border points because the additional dimension requires disproportionately more observations. Therefore, we present our nonparametric results generally for the border point approach and note that results are robust to employing bivariate local linear regressions.

---

[25]As an alternative to the centroids' distances from **B** – which can be very small with narrow municipalities – we approximate the location of a municipality by the average over a sufficiently large number of grid cells within the municipal boundaries. All our results are robust to this alternative. In this case we split each municipality into $100m \times 100m$ grids, determine longitude, latitude as well as distances from **B** for each grid cell and take the municipal averages across grid cells to obtain $g(D)$ and $f(\mathbf{L})$.

Figure B1: Border points

Notes: The red dots (green triangles) mark the 20 (30) border points we employ in our analysis. These are allocated at equal distances along the ZRG border. The black lines mark the inner-German border and the state borders. The treated states were Bavaria (B), Hesse (H), Lower Saxony (LS) and Schleswig-Holstein (SH). The Czechoslovakian border is marked in purple. When splitting up the sample into treated units closer to the former GDR or to Czechoslovakia we use the perpendicular distances of municipal centroids or pixels to the respective borders.

B.2 Fuzzy RDD: Exploiting the political treatment rule

In the fuzzy RDD we exploit the fact that eligibility was governed by a clear rule: districts that accommodated either 50 percent of its area or population within a band of 40 kilometers to the Iron Curtain became part of the ZRG. The blue-shaded area in Panel A of Figure 2 illustrates the 40-kilometer buffer. It is evident that the ZRG border roughly follows the buffer. Hence, we compute an assignment variable, denoted by M_d, indicating a district's minimum distance from the Iron Curtain that includes the majority share of the district's area. If exemptions from the 40-kilometer rule were not too frequent, we should observe a jump in the probability of treatment at the threshold $M_0 = 40$:

$$P(T_{id}|\widetilde{M}_d) = \begin{cases} h_1(\widetilde{M}_d) & \text{if } \widetilde{M}_d \leq 0 \\ h_0(\widetilde{M}_d) & \text{if } \widetilde{M}_d > 0, \end{cases} \tag{B.4}$$

where $\widetilde{M}_d = M_d - M_0$ denotes the centered version of the assignment variable.

We estimate the fuzzy RDD in a parametric as well as in a nonparametric fashion. In the latter approach we estimate the conditional expectations of outcome and treatment probability by means of local linear regressions separately for observations with $\widetilde{M_d} > 0$ and those with $\widetilde{M_d} \leqslant 0$. We employ a triangular kernel and follow Imbens and Kalyanaraman (2012) in choosing an optimal bandwidth h^* that minimizes the mean squared error of the average treatment effect.[26] The parametric approach follows a 2SLS approach where the regression equations are given by:

$$Y_{id} = \alpha + f_0(\widetilde{M_d}) + T_{id}[\tau + f_1(\widetilde{M_d}) - f_0(\widetilde{M_d})] + \varepsilon_{id}, \tag{B.5}$$

$$T_{id} = \gamma + h_0(\widetilde{M_d}) + R_d[\delta + h_1(\widetilde{M_d}) - h_0(\widetilde{M_d})] + \nu_{id},$$

where $R_d = 1[M_d \leqslant M_0]$ indicates eligibility.[27] Since the political rule is applied at the district level d, we correct the estimated variance-covariance matrix for clustering at the level of districts and for heteroskedasticity of arbitrary form. We limit the sample to observations belonging to districts characterized by $M_d \leq 150$.

B.3 Difference-in-discontinuities specification

Our differences-in-discontinuities specification can be directly derived from (B.2), so we have

$$Y_{it} = \alpha + \gamma_t + g_0(D_i) + f(\mathbf{L_i}) + T_i[\tau + g_1(D_i) - g_0(D_i) + \beta S_t] + f(\mathbf{L_i}S_t) + \varepsilon_{it}. \tag{B.6}$$

Note that T_i indicates whether a municipality is located in the Zonenrandgebiet and S_t is a dummy variable equal to one after the shock has occurred (post-1990 and post-2004, respectively) and zero otherwise. Based on the insights from Redding and Sturm (2008) that the benefits of market access are declining in distance, we include a term that interacts the shock with location, $f(\mathbf{L_i}S_t)$. Moreover, time fixed effects γ_t absorb all common per-period effects of the shocks. Table 6 shows the estimates for the coefficients τ and β denoted by ZRG and $S \times ZRG$ in the table, respectively. The contribution of the respective shocks for the overall discontinuity is given by $\beta/(\beta + \tau)$.

---

[26]As noted by Imbens and Kalyanaraman (2012) this procedure often leads to bandwidth choices that are similar to those based on the optimal bandwidth for estimation of only the differences in expected outcomes (and applying the same bandwidth to the expectations of treatment probabilities). This holds also true in our case.

[27]In what follows, we will generally use linear probability models in the first stage, but the results are very similar to those obtained with a nonlinear probability model in the first stage.

Table C1: LOCAL TAX RATES (2010)

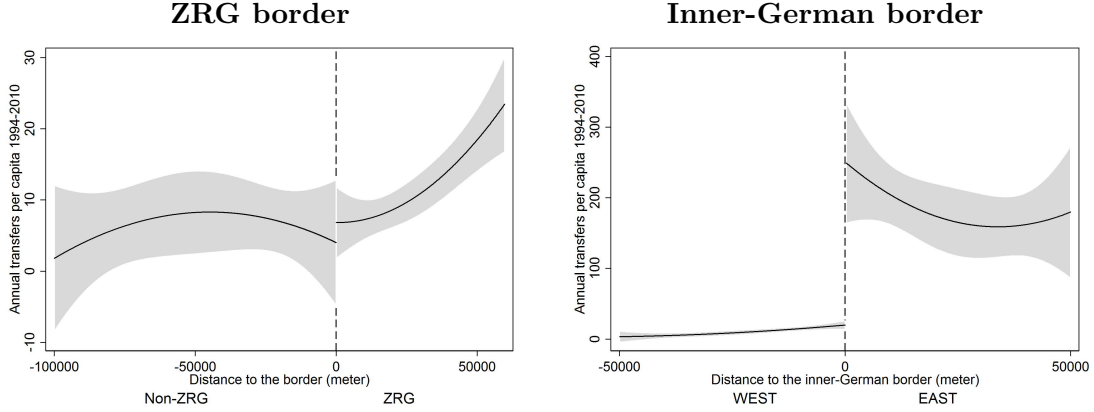|	Business tax rates			Property tax rates		
	Coordinate control		Nonparametric	Coordinate control		Nonparametric
	2nd	3rd	h^*	2nd	3rd	h^*
ZRG transfers	0.012*	0.000	0.014	0.032**	-0.006	-0.021
	(0.007)	(0.008)	(0.015)	(0.016)	(0.020)	(0.029)
Adj. R^2	0.30	0.30	-	0.41	0.42	-
Obs.	3,881	3,881	1,318	3,878	3,878	1,420

Notes: ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively. Robust standard errors in parenthesis. We drop all observations outside a 100km window of the ZRG border in the parametric specifications. Columns (1)-(2) and (4)-(5) refer to parametric specifications and include state indicators. Columns (3) and (6) refer to nonparametric specifications where h^* denotes the optimal bandwidth computed according to Imbens and Kalyanaraman (2012). We apply a logit transformation to *business tax rates* as well as *property tax rates* which are bounded between zero and unity.

C Substitutive policies after 1990

We have shown that transfers to the Zonenrandgebiet led to a *persistent* increase in income per square kilometer in the target region, channelled through higher capital stock, population and employment. This outcome can be interpreted as causal if the former ZRG border does not exhibit any other discontinuity ex post. For example, one can think about policy makers trying to compensate households and firms in the former treatment area in various ways. However, German municipalities do not have control over many policy instruments. Important tax rates like income taxes are chosen at the federal level and have to be approved by the states. And those taxes that municipalities can set themselves are mostly too small to be relevant for location decisions. The business tax rate is an important exception. One could hypothesize that municipalities in the Zonenrandgebiet lowered their business tax rates after 1994 to compensate firms for the loss in subsidies. Based on data from the German Statistical Office, Table C1 shows that there is no discontinuity in business tax rates in 2010. Our preferred specifications show insignificant effects. The one being significant rather points towards higher tax rates in the Zonenrandgebiet. Property tax rates are set at the municipality level as well but they correspond to a much lower revenue than business tax rates. Moroever, our estimates in Table C1 suggests no significant difference between the property taxes in formerly treated and control regions.

Further, policy makers could have decided to compensate former ZRG municipalities by an alternative transfer scheme that substituted the old program, at least to some extent. Table 10 and Figure C1 document that there is no discontinuity in transfer recipience for municipalities at the former ZRG border in 2010. Taking into account that the program was terminated in 1994 and that the former ZRG border does not correspond to district borders while transfer intensities are determined by districts, this

Figure C1: TRANSFERS 1994-2010

is not surprising. Instead, as Figure C1 shows we observe a discontinuity at the former inner-German border. This result fits into the general picture that regional transfers moved to the new Länder after German reunification. In addition to federal transfers, Figure C1 accounts for resources redistributed within states according to municipal fiscal equalization schemes (in the case of Bavaria). These within state transfers 'Bedarfszuweisungen' are at much lower scale (about 0.3 percent of the federal transfers) and display no discontinuity at **B**.

D   Grid-cell data

For the analysis at the fines spatial scale we use data on *PrivateCapital* and *Radiance* for grids of $100m \times 100m$ and 30×30 arc-seconds (about $900m \times 600m$ in the center of Germany), respectively (see details about the data in Appendix A). For a sensitivity check about the role of frictions at municipality borders we consider three subsamples that are arguably less prone to inter-municipality frictions. Table D1 reports in columns (1)-(3) the contemporaneous and in columns (4)-(6) the persistent average treatment effects of the ZRG transfers on the probability of a grid cell being covered by private capital and on log radiance.

The first sensitivity check limits the sample to municipality pairs around the treatment border that are highly connected. For this we employ data from OpenStreetMap provided by ThinkGeo which includes all streets and roads in Germany. We drop all small pathways and trails and use GIS techniques to extract roads that cross the ZRG border **B**. Then, we compute for each municipality around **B** the number of roads which cross both the ZRG border as well as the respective municipality. To adjust for

Table D1: Robustness: Grid Cell Data

	Contemporaneous effects			Persistent effects		
	Connectedness	Undeveloped	Polycentric	Connectedness	Undeveloped	Polycentric
Prob(PrivateCapital=1)						
ZRG transfers	0.024***	0.019***	0.007***	0.023***	0.022***	0.014***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Adj. R^2	0.01	0.02	0.01	0.01	0.02	0.02
Obs.	776,639	147,424	758,968	776,639	147,424	758,968
Log(Radiance)						
ZRG transfers	0.222***	0.376***	0.203***	0.217***	0.241***	0.214***
	(0.022)	(0.049)	(0.035)	(0.020)	(0.044)	(0.030)
Adj. R^2	0.12	0.17	0.11	0.10	0.24	0.09
Obs.	10,626	2,790	11,579	13,037	3,296	12,665

Notes: ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively. Robust standard errors in parenthesis. We drop all observations outside a 100km window of the ZRG border in the parametric specifications. Columns (1)-(3) and (4)-(6) refer to the contemporaneous and persistent estimates, respectively. All estimates correspond to the parametric specifications and include asymmetric control functions of distance to the ZRG border as well as coordinate controls. Columns (1) and (4) base on municipality pairs at the ZRG border and in the upper 10th percentile of connectedness. Columns (2) and (5) restrict the sample to municipality pairs at the ZRG border that are not separated by undeveloped land such as rocks, water bodies, or forests. Columns (3) and (6) restrict the sample to polycentric municipality pairs at the ZRG border. Observations refer to grid cells of $100m \times 100m$ and 30×30 arc-seconds for private capital and radiance, respectively.

municipality size this variable is weighted by municipality area. We keep only grid cells belonging to municipalities in the upper 10th percentile of the distribution of roads crossing **B** and the municipality. Thus the sample size drops by about 90 percent compared to Table 8 while the point estimates (see columns (1) and (4)) remain highly significant and well in line with those presented in Table 8.

For the second experiment we draw a buffer at 1km distance from both sides of **B**. Employing land-use data from the CORINE project we calculate the share of non-cultivable land within this buffer for all municipalities. We define as non-cultivable land the categories 23-44 ("Forest and and semi natural areas", "Wetlands", "Water bodies"). All municipalities with more then 5 percent of their land within the 1km buffer being non-cultivable are dropped from the sample. Thus, we estimate the effect on grid cells belonging to municipality pairs that feature homogeneous characteristics in the neighborhood of **B** as the land could be developed at relatively low costs. Again the estimates in columns (2) and (5) are not increasing substantially compared to the benchmark in Table 8 which supports the assumption that frictions are not decisive for our results.

The third experiment focuses on polycentric municipalities arguing that social networks are more dispersed in scattered neighborhoods. For each grid cell in the data set

we calculate the number of grids in a 200, 300, and 400 meter radius that are covered by capital structures. That is we construct a spatial lag for $capital_i^{spatlag} = \mathbf{W} \times capital_j$ for $j \neq i$ and $\mathbf{W}$ referring to a weight matrix with elements being unity if the distance between i and j is less than 200, 300, or 400 meters depending on the specification and zero otherwise. Using the municipality distributions of $capital_i^{spatlag}$ we refer to a point i as a municipality center if it features the highest number of covered grid cells in the surrounding neighborhood i.e. if $capital_i^{spatlag}$ corresponds to the municipality maximum. If this binary definition of a center is unity for multiple adjacent grid cells we aggregate these cells using ArcGIS. In case this definition yields several but *non-adjacent* municipality centers (grid-cells that feature the municipality maximum of $capital_i^{spatlag}$) we refer to the municipality as a polycentric municipality. Independent of the chosen radius, the majority of municipalities turns out monocentric according to the above definition. For the estimates presented in Table D1 we use the 200 meter radius and keep only municipalities with at least two non-adjacent centers. The treatment effects remain highly significant and the magnitude is very similar to the benchmark. Note that this holds also true for using the 300 and 400 meter radii.

E A simple model

In this section, we present a simple model to show that estimated discontinuities in a spatial RD design do not include externalities if these dissipate continuously in space. The argument has been made recently by Turner, Haughwout and van der Klaauw (2014) in the context of land regulation. We apply this idea to regional transfers.

Consider two initially symmetric regions accommodating locations of unit measure x between the boundaries $-\bar{x}$ and $\bar{x}$. The common (treatment) border is referred to by $x = 0$. One region with locations $x \in [0; \bar{x}]$ receives treatment T (referred to by $+$) while the other region with locations $x \in [-\bar{x}; 0]$ does not (referred to by $-$).[28] We define economic activity as output in location x as

$$Q(x) = A(x, I)I(x, A, T). \tag{E.7}$$

Output is increasing in physical inputs $I(\cdot)$ (for example, labor, capital, public goods or a combination thereof) and in a production amenity $A(\cdot)$. We assume the production amenity to be composed of two parts such that $A(x, I) = A_{ext}(x, I)A_{own}(x, I)$. While the first part dissipates in space, the second component is confined to the respective location. To ensure that there is always some economic activity in both regions, we impose that $A(\cdot)$ is hump-shaped in I capturing the idea that dispersion forces dominate agglomeration forces at a certain level of economic density or economic frictions (like trade costs). For example, $A_{ext}(\cdot)$ can be understood as agglomeration economies

---

[28]A region can be understood as an area composed of municipalities or as a municipality composed of small grid cells – depending on the level of aggregation in our empirical analysis.

stemming from knowledge spillovers, labor market pooling, supply linkages or home-market effects, among others (Marshall, 1920, and Duranton and Puga, 2004). $A_{own}(\cdot)$, in contrast, captures all effects on productivity that occur only under the condition of being located in that location (e.g. better public facilities, capital structures, tax benefits, etc.). Our specification of economic activity allows for hysteresis as inputs depend on production amenities and vice versa (as e.g. in Krugman, 1991). With respect to the second part in (E.7), we assume that (i) treatment T raises inputs and (ii) more productive locations have an incentive to invest more, so $\partial I(\cdot)/\partial A(\cdot) > 0$.[29]

To make both the contemporaneous and persistent effects of treatment transparent, we evaluate the model at three distinct points in time: 1. *before* treatment, 2. *during* treatment, and 3. *after* treatment.

1. Before treatment. As both regions are initially identical and $T^+ = T^- = 0$, it is immediate that input and production amenity levels are identical. Relating (E.7) to our empirical specification by taking logs and comparing economic activity at the treatment border $x = 0$ implies: $lnQ(0^+) - lnQ(0^-) = 0$, where $x = 0^+$ and $x = 0^-$ refer to locations converging to $x = 0$ from the treatment and the control side, respectively.

2. During treatment. Now suppose that one region receives treatment $T^+ > 0$ while $T^- = 0$ (e.g. capital subsidies or provision of public infrastructure). This leads to an increase in I^+ in a first step while I^- remains unchanged. We refer to this as the *direct input effect* of treatment. In a second step, the difference in $I(\cdot)$ raises the production amenity stimulating more input investments in consecutive steps until $\partial A(\cdot)/\partial I(\cdot) = 0$. We call this the *self-reinforcing effect.* Importantly, however, externalities do not stop at local borders so locations close to $x = 0$ benefit from input intensities on both sides of the threshold. We follow recent work by Turner, Haughwout, and van der Klaauw (2014) by assuming that $A_{ext}(x)$ is a weighted average of input intensity of neighboring locations according to

$$A_{ext}(x) = \frac{1-\gamma(x)}{2}\int_{-\bar{x}}^{0^-} I(x)dx + \frac{1+\gamma(x)}{2}\int_{0^+}^{\bar{x}} I(x)dx. \qquad (E.8)$$
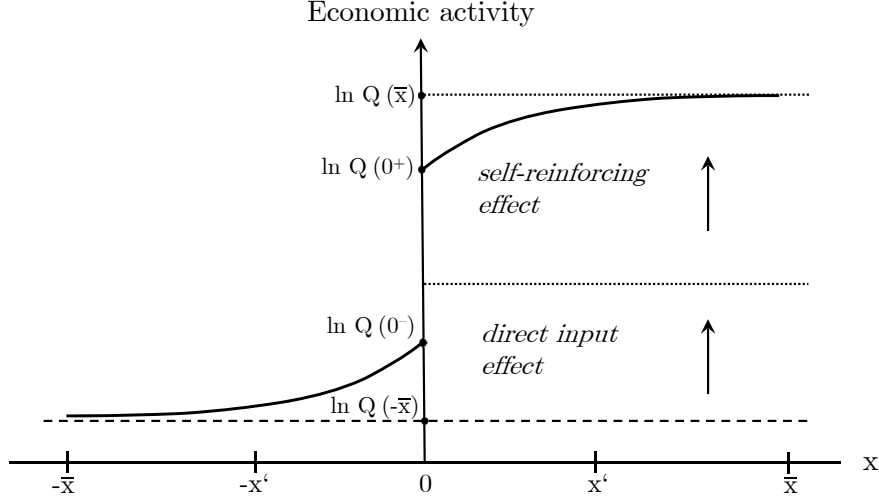
We define $\gamma(x)$ to be a weakly increasing continuous function with $\gamma(x) = -1$ if $x \leq -x' \wedge -x' > -\bar{x}$, $\gamma(0^-) = \gamma(0^+) = 0$ and $\gamma(x) = 1$ if $x \geq x' \wedge x' < \bar{x}$. Hence, according to (E.8) locations $x = 0^-$ and $x = 0^+$ experience the same level of externality, that is $A_{ext}(0^+) = A_{ext}(0^-)$. With respect to the location-specific production amenity, we observe a discontinuity at $x = 0$ due to differences in input intensities at this location. Using these insights and taking logs of (E.7), we obtain at $x = 0$ a discontinuity

$$\Delta Q(0) = \Delta I(0) + \Delta A_{own}(0),$$

where Δ refers to the difference in the logs of the respective variables. Importantly, the

---

[29]This feature is in line with heterogeneous firms models, e.g. Melitz (2003).

Figure E1: Direct effect versus productivity effect

Notes: The dashed line refers to the situation without transfers. The solid line refers to the output with transfers and the dotted lines illustrate the "direct input effect" and the "productivity effect". The area left of $x = 0$ belongs to the control group whereas observations right of $x = 0$ receive transfers.

observed discontinuity does not contain A_{ext}. Note that agglomeration externalities do raise economic activity, but it is in the nature of spatial RD identification that all continuous effects are not part of the discontinuity. Intuitively, self-reinforcing externalities exert the same incentive for investment at $x = 0^+$ and $x = 0^-$. Figure E1 provides a graphical illustration of this argument.

3. After treatment. After the end of the program, locations in the former treatment region no longer receive transfers and $T^+ = T^- = 0$. Hence, for the gap in economic activity to persist, our framework suggests two explanations: either inputs I stay at the same level even in the absence of treatment (e.g. capital does not depreciate) or location-specific productivity A_{own} remains at a higher level in the former treatment area persistently.

The model reveals that (agglomeration) externalities cannot be part of the estimated discontinuity if they dissipate continuously in space. More formally, $\gamma(x)$ has to be continuous function. In the main part of the paper, we undertake a number of exercises to check the plausibility of this assumption in our context.

Finally, it is noteworthy that the implications of the model do not change if we allow for local relocation. This would shift economic activity downwards in the control region and upwards on the other side of the treatment border, but the estimated discontinuity would still not include spatially continuous externalities.

F   Polynomial order & bandwidth

Table F1 presents the results for our main outcome, *income per km²*, across numerous different specifications. In panel (I) we report specifications with distance control functions using asymmetric 1st- to 5th-order polynomials with segment and state fixed effects. Panel (II) displays the coordinate control specifications analogously to Table 2 but adding a version with linear control function. Panel (III) contains nonparametric specifications without border point fixed effects but with the data being residualized on latitude instead. Panel (IV) shows the benchmark nonparametric specifications with 30 instead of 20 borderpoint fixed effects as used in Table 2. Panel (V) performs robustness checks for the fuzzy RDD results in Table 3 by choosing different bandwidth.

Table F1: ROBUSTNESS: POLYNOMIAL ORDER & BANDWIDTH

Log income per km^2	Contemporaneous effect				Persistent effect			
	Coef.	Std.err	Adj. R^2	AIC	Coef.	Std.err	Adj. R^2	AIC
(I) Distance control								
1st order	0.149***	(0.058)	0.21	10,798	0.161***	(0.055)	0.20	10,508
2nd order	0.271***	(0.079)	0.21	10,787	0.270***	(0.075)	0.21	10,500
3rd order	0.487***	(0.098)	0.22	10,761	0.495***	(0.094)	0.21	10,472
4th order	0.619***	(0.120)	0.22	10,756	0.572***	(0.115)	0.21	10,468
5th order	0.560***	(0.144)	0.22	10,742	0.560***	(0.144)	0.22	10,742
(II) Coordinate control								
1st order	0.206***	(0.059)	0.17	10,972	0.209***	(0.057)	0.17	10,673
2nd order	0.296***	(0.080)	0.19	10,869	0.296***	(0.077)	0.19	10,541
3rd order	0.528***	(0.099)	0.22	10,741	0.535***	(0.096)	0.22	10,404
(III) Nonparametric, controlling for latitude								
h^*	0.572 ***	(0.108)	-	-	0.583***	(0.117)	-	-
$0.8 \times h^*$	0.550***	(0.122)	-	-	0.555***	(0.133)	-	-
$1.2 \times h^*$	0.575 ***	(0.099)	-	-	0.581***	(0.105)	-	-
(IV) Nonparametric with 30 border point fixed effects								
h^*	0.410***	(0.083)	-	-	0.392***	(0.085)	-	-
$0.8 \times h^*$	0.380***	(0.089)	-	-	0.357***	(0.092)	-	-
$1.2 \times h^*$	0.420***	(0.078)	-	-	0.401***	(0.079)	-	-
(V) Fuzzy RDD, nonparametric specifications								
h^*	0.545***	(0.143)	-	-	0.404***	(0.132)	-	-
$0.8 \times h^*$	0.636***	(0.179)	-	-	0.409**	(0.161)	-	-
$1.2 \times h^*$	0.601***	(0.128)	-	-	0.495***	(0.119)	-	-

Notes: ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively. The dependent variable is log income per km^2 in all specifications. For the nonparametric specifications the bandwidth h^* is computed according the algorithm introduced by Imbens and Kalyanaraman (2012). Panel (I) reports the results for parametric specifications including a flexible control function of distance from the treatment boundary (D_i) using asymmetric 1st- to 5th-order polynomials with segment and state fixed effects. According to the AIC the 5th order polynomial is the preferred specification. Panel (II) reports the coordinate control function specifications as displayed in Table 2 but adding a linear version. According to the AIC the 3rd order polynomial is the preferred specification for the coordinate control. Note that we choose lower order polynomials for the coordinate control function $f(.)$ than for the distance control function $g(.)$ because the bivariate control function requires more parameters to be estimated than the corresponding univariate control function. Panel (III) reports nonparametric estimates controlling for a continuous measure of latitude instead of 20 border point fixed effects as in the nonparametric specifications in Table 2. Panel (IV) reports nonparametric estimates using 30 instead of 20 different border point fixed effects. Panel (V) displays the results of the nonparametric version of the fuzzy RDD specifications as in Table 3 for different bandwidths i.e. 0.8 and 1.2 times the optimal bandwidth.

Table F2: GROWTH OF OUTCOMES RELATIVE TO THE PRE-TREATMENT PERIOD

	Income growth			Employment growth			Population growth		
	Parametric M_d		Nonparametric	Parametric M_d		Nonparametric	Coordinate control		Nonparametric
	2nd	3rd	h^*	2nd	3rd	h^*	2nd	3rd	h^*
ZRG transfers	0.364	0.363	0.655	0.444***	0.445***	0.597***	0.226***	0.299***	0.264***
	(0.241)	(0.242)	(0.449)	(0.161)	(0.163)	(0.219)	(0.032)	(0.040)	(0.034)
R^2	0.01	0.01	-	0.03	0.03	-	0.20	0.22	-
AIC	629	631	-	379	381	-	3,631	3,548	.
Obs.	297	297	168	294	294	189	3,751	3,751	3,088

Notes: ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively. Robust standard errors in parenthesis. We measure growth by log difference between 1986 and 1961 for income and employment and by log differences between 1951 and 1986 for population. Regressions for income growth and employment growth are based on district level data and fuzzy RDD specifications whereas regressions for population growth are based on municipality level data and spatial RDD specifications. Observations with $M_d > 150$ are dropped from the sample in the columns (1) to (6). Columns (3), (6) and (9) refer to the corresponding nonparametric estimates where h^* denotes the optimal bandwidth computed according to Imbens and Kalyanaraman (2012). For the nonparametric specifications we also computed the Calonico et al. (2017) robust bias-corrected confidence bounds which confirm the conventional estimates: The corresponding p-values are 0.105 for income growth, 0.052 for employment growth, and 0.000 for population growth.