AIDDATA
A Research Lab at William & Mary

# WORKING PAPER 127

January 2024

## Transport infrastructure and the spatial distribution of $CO_2$ emissions

**Martin Kleimann**
FernUniversität in Hagen

**Hans-Jörg Schmerera**
FernUniversität in Hagen
Institute for Employment Research (IAB)

## Abstract
This study presents an analysis of transport infrastructure projects and their effect on greenhouse gas emissions. Causal inter-regional spillover effects are estimated for the launch of the China Europe Railway Express using a spatial difference in differences estimator. The preferred Spatial Durbin Model reveals a small negative direct effect of approximately -3.5 percent and a more substantial positive indirect effect of around 12.7 percent. These results suggest a crowding out effect on pollutant heavy industries from the nodal regions into the periphery. We provide further evidence for this hypothesis by testing the effect on the respective regions' structural composition. A placebo exercise supports causality of our benchmark findings.

## Author Information

**Martin Kleimann**
FernUniversität in Hagen
martin.kleimann@fernuni-hagen.de

**Hans-Jörg Schmerera**
FernUniversität in Hagen

**Key words:** BRI, Trade, Rail Freight, Spatial Spillovers, CO2 emissions
**JEL classification:** C33, F18, R11, R12



## Acknowledgments
This research did not receive any specific grantfrom funding agencies in the public, commercial, or not-for-profit sectors.

# Transport infrastructure and the spatial distribution of $CO_2$ emissions

Martin Kleimann*[a] and Hans-Jörg Schmerer[a, b]

[a]FernUniversität in Hagen
[b]Institute for Employment Research (IAB)

This version:
November 28, 2023

**Abstract**

This study presents an analysis of transport infrastructure projects and their effect on greenhouse gas emissions. Causal inter-regional spillover effects are estimated for the launch of the China Europe Railway Express using a spatial difference in differences estimator. The preferred Spatial Durbin Model reveals a small negative direct effect of approximately -3.5 percent and a more substantial positive indirect effect of around 12.7 percent. These results suggest a crowding out effect on pollutant heavy industries from the nodal regions into the periphery. We provide further evidence for this hypothesis by testing the effect on the respective regions' structural composition. A placebo exercise supports causality of our benchmark findings.

**JEL Codes** *C33, F18, R11, R12*
**Keywords** *BRI, Trade, Rail Freight, Spatial Spillovers, $CO_2$ emissions*

*Corresponding author: martin.kleimann@fernuni-hagen.de . This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# 1 Introduction

Since 2011, the China Europe Railway Express (CER-Express) has been operating as a new mode of freight transport connecting the People's Republic of China (PRC) and the European Union (EU). It has been integrated into the overarching narrative of the Belt and Road Initiative (BRI) proclaimed by Chinese president Xi Jinping in 2013[1] and became one of the most tangible projects within the framework.

Much of the existing research on the economic impact of this program is focusing on trade creation without considering potential external effects caused by the various projects associated with the BRI. For instance, Mau and Seuren (2023) find that the CER-Express's establishment elevated European companies' propensity to export to China. Related to their study, Fang et al. (2021) investigate the effect of CER-Express on local development in EU regions connected by CER-Express. Their estimated impact on output is insignificant but the launch can be associated with increasing intermodal transport and, to a lesser extent, additional employment. These effects can be justified by theoretical arguments based upon the seminal works of Krugman (1991) and Baldwin and Forslid (2000). The former argues that declining transportation cost fuels agglomeration, while the latter show how agglomeration can spur endogenous growth.

Building on those insights and the theoretical extension of the Krugman (1991) model by Grazi et al. (2007), we are focusing on potential external effects on the environment due to infrastructure projects. Improvements in infrastructure may raise emissions by attracting new manufacturing firms with particular interest in accessing the Chinese market. Producers of more expensive goods may prefer the direct routes over time consuming transport by ship (supply side argument). It may also be relevant for offshoring firms relying on stable connections to their Chinese companions (demand side argument). The former channel may spur local production and emission, while the latter might even reduce pollution in the nodal regions through the so called pollution haven channel. Due to the better and more stable access to intermediates from China, incumbents and new entrants may offshore more dirty production stages. This is particularly relevant for goods sourced from firms in the Chinese hinterland where wages are still competitive but access to international markets by ship is also more cumbersome.

Thus, our analysis contributes to the existing empirical literature on the relationship between transport infrastructure and economic agglomeration (Ahlfeldt and Feddersen 2018; Behrens et al. 2018; Liu et al. 2022), which can also be linked to increased emission of greenhouse gasses as proposed by other research (Chen et al. 2018; Cheng 2016; Dong et al. 2020).

These contributions motivate the following hypotheses about the expected treatment effects associated with the establishment of a connection to the CER-Express system on pollution.

*Hypothesis 1* EU regions that serve as nodal points for the CER-Express are expected to attract more economic activity due to enhanced market access. Economic activity itself may spur emissions. This argument primarily revolves around the supply side. However, the demand side could also play an important role in amplifying this effect. Intermodal transport

[1] `https://www.gov.cn/govweb/ldhd/2013-09/08/content_2483565.htm` (in Chinese, visited on 10/16/2023)

is especially relevant for firms that offshore some fragments of their production chain when time matters. Offshoring opens another channel through which better access to Chinese suppliers by train may reduce emissions when more pollutant-laden production stages are offshored.

*Hypothesis 2* The exact spatial pattern of agglomeration is ambiguous. Since the nodal points are often located in rather densely populated areas, the effects might rather be identified within their proximal vicinity.[2]

We investigate the hypothesized impact of the newly established train connections on carbon dioxide ($CO_2$) emissions (*Hypothesis 1*) taking into consideration the spatial interdependence between the regions of interest (*Hypothesis 2*) as outlined in the methodology section.

The remainder of the paper is structured as follows: Section 2 describes the data employed in our study and its various sources. Section 3 presents the estimation strategy and the motivation behind the choice of the model used for identification. Section 4 reports the findings obtained from the empirical analysis. Section 5 concludes.

## 2 Data

The period of analysis has been restrained to the years between 2003 and 2018, which is seven years before and seven years after the initial launch of the CER-Express. Thus, the sample is symmetric around the initial treatment assignment, which corroborates consistency in difference in differences estimates (Chabé-Ferret 2015).

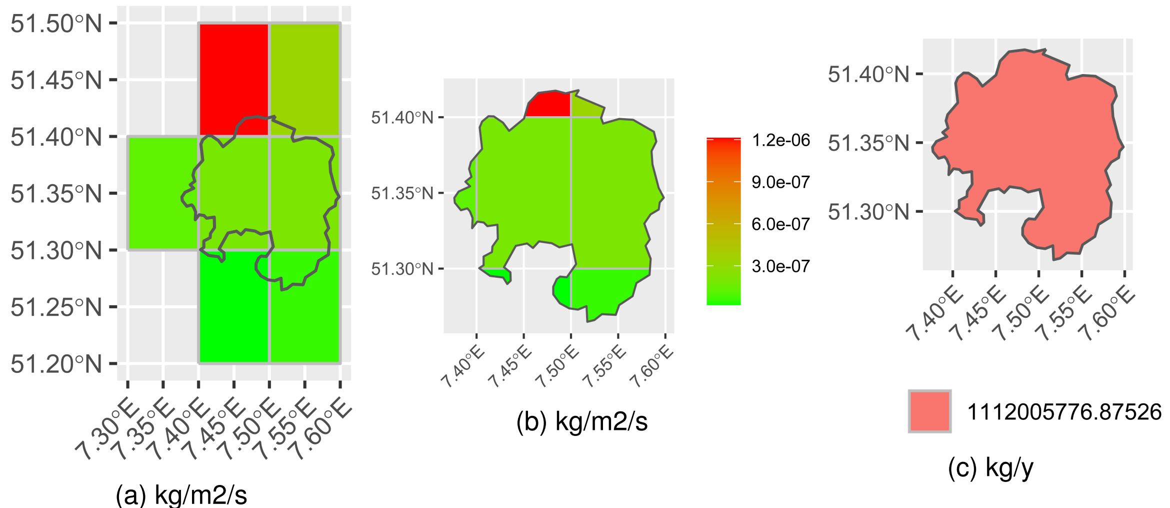
Data on $CO_2$ emissions is taken from the Emissions Database for Global Atmospheric Research (EDGAR, Crippa et al. 2019). We utilize yearly information provided at a 0.1*0.1 grid map excluding short-cycle carbon, i.e. biomass and biofuel combustion. Regional emissions are computed by overlaying this with *Nomenclature des unités territoriales statistiques* (NUTS, Eurostat 2020) level 3 boundaries from the 2021 iteration of adjustments. Figure 1 illustrates this process, see its description for more technical detail.

We create our primary variable of interest, the CER-Express treatment dummy, from information collected by Mau and Seuren (2023). Table 1 replicates the information from their paper. Similar to them, we account for the date of commencement within the initial period in each connected region by applying $\frac{13-month}{12}$. In all following periods the dummy takes a value of unity, irrespective of the number of connections, as we do not possess sufficient quantitative information on the factual degree of utilization.

Moreover, information extracted from Eurostat include regional gross domestic product (GDP), primary and secondary sector shares, as well as total gross value added (GVA), size of the respective region and regional population counts. These are then used to construct the dependent variable, which is per capita (p.c.) $CO_2$ emissions (in kg), and a number of con-

[2]Formulation of the second hypothesis is complicated by the fact that the size of regions included in our analysis varies considerably.

Figure 1: Exemplary overlay procedure for Hagen municipality (Germany)




**$CO_2$ emissions (excluding short cycle) in 2018.** *The overlaying procedure starts by projecting both layers of information in a joint coordinate space (1a). It is then determined which areas of the 0.1*0.1 grid maps are covered by every distinct NUTS 3 entity (1b). Finally, emission intensity in each polygon is multiplied by its size and values are aggregated to produce a concrete value for each region-year pair in the sample (1c).*

trols; namely p.c. regional GDP (in Euros), regional primary and secondary sector shares (in percent) and population density (in inhabitants per km$^2$).

Table 1: Number of connections

| **Destination** | **Year** | **Count** |
|---|---|---|
| Amsterdam | 2018 | 1 |
| Antwerp | 2018 | 1 |
| Bratislava | 2017 | 1 |
| Budapest | 2017 | 2 |
| Budapest | 2018 | 3 |
| Duisburg | 2011 | 1 |
| Duisburg | 2013 | 2 |
| Duisburg | 2014 | 5 |
| Duisburg | 2016 | 6 |
| Duisburg | 2017 | 7 |
| Hamburg | 2013 | 1 |
| Hamburg | 2014 | 2 |
| Hamburg | 2015 | 4 |
| Hamburg | 2016 | 5 |
| Hamburg | 2017 | 7 |
| Kouvala | 2017 | 1 |
| Liege | 2018 | 1 |
| Lodz | 2013 | 1 |
| London | 2017 | 1 |
| Lyon | 2016 | 1 |
| Madrid | 2014 | 1 |
| Malaszewice | 2011 | 1 |
| Malaszewice | 2014 | 3 |
| Malaszewice | 2015 | 5 |
| Mannheim | 2018 | 1 |
| Milan | 2017 | 2 |
| Munich | 2017 | 1 |
| Nuremberg | 2015 | 1 |
| Prague | 2017 | 2 |
| Riga | 2016 | 1 |
| Riga | 2017 | 2 |
| Riga | 2018 | 3 |
| Rotterdam | 2017 | 1 |
| Tilburg | 2016 | 1 |
| Tilburg | 2018 | 2 |
| Vienna | 2018 | 1 |
| Vuosaari | 2018 | 1 |
| Warsaw | 2012 | 1 |
| Warsaw | 2013 | 2 |
| Warsaw | 2014 | 3 |
| Warsaw | 2017 | 4 |
| Zaragoza | 2017 | 1 |
| Zaragoza | 2018 | 2 |

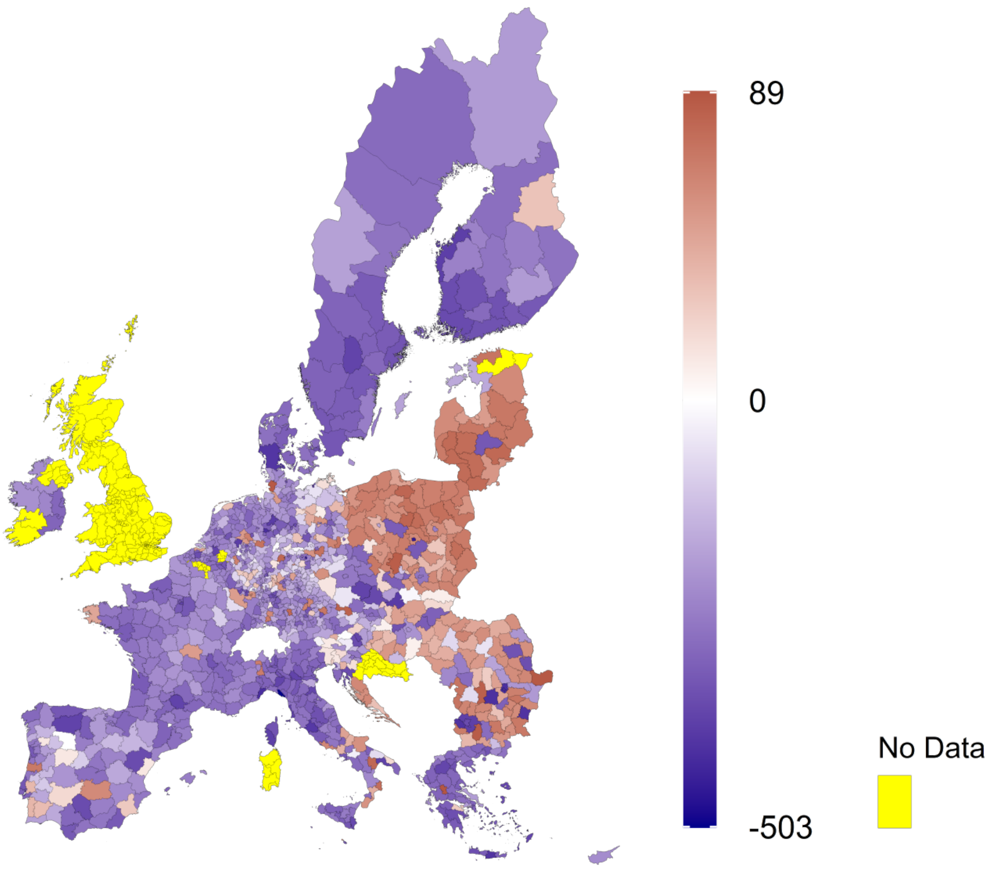
*Based on Mau and Seuren (2023)*

The final data set covers 1117 NUTS 3 regions in 27 EU countries, excluding the United Kingdom (UK) because of data availability issues related to its withdrawal from the EU. Since an agreement on statistical cooperation has not yet been signed, both entities use different primary and secondary sector definitions.[3] To allow for the UK's inclusion, a further robustness check excludes the two sector share variables from the regression. The analysis also

[3]The former includes mining and quarrying, while the latter does not.

excludes the EU's outermost regions[4] French Guiana, Guadeloupe, Martinique, Mayotte, Reunion Island and Saint-Martin (France), Azores and Madeira (Portugal), and the Canary Islands (Spain). Although officially belonging to the respective countries and thus the EU, the geographical distance from the mainland hinders interdependence of these areas in accordance with the channels formulated in the introduction.

Figure 2: Emissions over time




*Percentage change of p.c. $CO_2$ emissions (excluding short cycle)*
*2003 to 2018, logarithmic color scale*

Figure 2 gives a detailed picture of all regions covered by our data. The missing values outside of the UK are due to boundary recoding issues during the period, stemming from national administrative border adjustments. Regional borders within a country can change due to various reasons and Eurostat routinely incorporates these updates, resulting in the loss of 33 observations. Moreover, Figure 2 also traces the region specific change of the dependent variable over the investigated period.

[4] `https://ec.europa.eu/regional_policy/policy/themes/outermost-regions_en` (visited on 10/16/2023)

A first glimpse at Figure 2 reveals significant variation of p.c. $CO_2$ over time and space. P.c. rates have been reduced substantially in most parts of France and Scandinavia, but the picture is more mixed for Germany, Spain and Italy. Also, many Eastern European regions became more $CO_2$-intensive over time. For the 252 regions with increasing emissions over time, the average value is 19.79 percent, while the median is 16.58. The remaining 865 territorial units managed to reduce their emissions. The equivalents are 29.82 and 21.5 percent.

Extensive summary statistics can be found in Table A.1. The descriptive statistics reveal that the primary driver of the variation in all variables other than the treatment dummy is between- rather than within-region variation. This finding highlights the substantial degree of disparity between observations in the sample. Additionally, Table A.2 maps correlates between the variables included in our sample. The statistics do not indicate problematically large values that may cause multicollinearity issues in multivariate models.

# 3 Methodology

The channels inferred in *Hypotheses 1* and *2* stress the possibility of regional interdependence through agglomeration.[5] The spatial difference in differences (SDID) model described in Chagas et al. 2016 accounts for this issue by fitting

$$Y_t = (\alpha + W\beta)D_t + (\mu + W\nu)X_t + \phi + \theta_t + \Xi_t, \tag{1}$$

where $Y_t = (Y_{1t}, ..., Y_{nt})'$ is a $nt \times 1$ vector of p.c. $CO_2$ emissions in $\sum_1^n = N$ regions in period $t$. The primary variable of interest $D$ indicates binary treatment. The dummy takes the value 1 when the respective region is directly connected to the China Europe Railway Express system and/or indirectly treated through one of the neighboring regions. The direct effect is captured by parameter $\alpha$ and the indirect effect is measured by $\beta$.

The latter effect depends on the distance between the respective region and the potential access points, which is introduced by the matrix $W$. This weighting matrix introduces the spatial dependencies between all regions into the model. Regions without direct access to the China Europe Railway Express system may still be treated through indirect treatment when surrounding regions have access the railway system. Put differently, all regions somehow depend on all other regions, but the relevancy is declining in distance. $X$ is a matrix containing the additional covariates introduced in the data section above, while $\phi$ and $\theta$ represent region and time fixed effects (FE). $\Xi$ is an error term.

[5] The primary technical motivation for choosing such a model lies within the implausibility of the stable unit treatment values assumption (SUTVA, Rubin 1980) in the analyzed setting. The newly commissioned rail connections are expected to not only influence economic activity in the directly connected node regions, but also other, especially nearby localities.

The model in equation (1) corresponds to the spatial lag of X model (SLX, Halleck Vega and Elhorst 2015), which represents the baseline approach to our investigation. It does not control for spatial autocorrelation or spatial dependency of errors. However, these terms can be easily incorporated by including

$$\lambda W Y_t \tag{1a}$$

on the right hand side of equation (1) and rewriting

$$\Xi_t = \rho W u_t + \varepsilon_t \tag{1b}$$

Theoretically, both of these adjustments can be performed simultaneously, producing the general nesting spatial model (GNS). However, Burridge et al. (2016) show that this approach has significant drawbacks as it has only yet been identified using one specific form of $W$ and carries the potential of overparameterization. Therefore, the more common approach is to account separately for either spatial autocorrelation or spatial dependency of errors generating the Spatial Durbin Model (SDM) or the Spatial Durbin Error Model (SDEM). Put another way, not all potential features can be accounted for simultaneously but various steps can be taken to evaluate the alternatives.

As suggested by the literature on model selection within spatial econometrics (e.g. Elhorst 2014), we present the following statistics: Lagrange multiplier (LM) tests can justify the inclusion of spatial lag and error terms. However, Halleck Vega and Elhorst (2015) argue that these tests might be less suitable for settings where the SLX model is the point of departure. Also, these tests do not allow for discriminating between SDM and SDEM, which are not mutually nested. Therefore, we utilize the Bayesian information criterion (BIC, Schwarz 1978) as the primary benchmark for selecting the appropriate model.

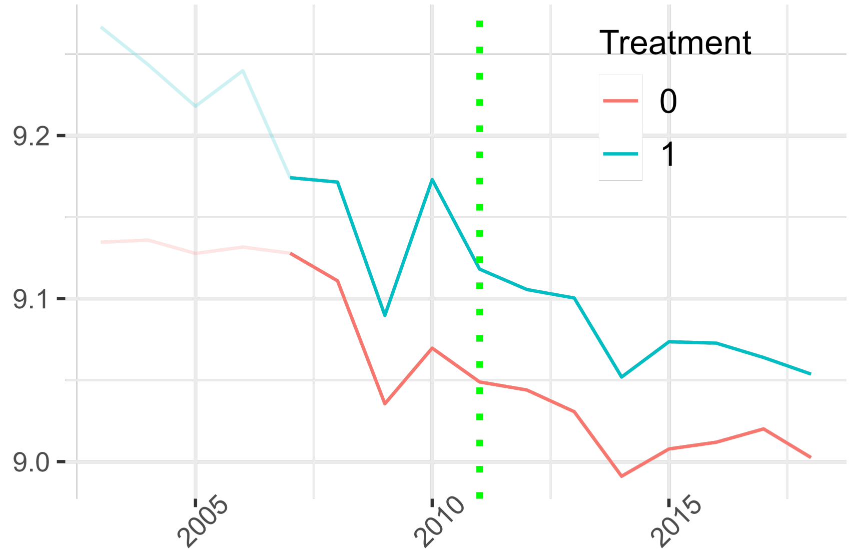
We also use BIC to choose between different weights matrices $W$. The perceived sensitivity of results to the choice of the right weighting matrix is one of the main points of criticism for spatial models (see Harris et al. 2011). Similar to, for instance, Chagas et al. 2016 we address this concern by presenting results obtained from alternative configurations. Since the theoretical channels outlined above are more suggestive of distance rather than contiguity as a relevant spill-over channel, we specify them through varying rates of distance based decay. Row-normalization of weights facilitates the estimated parameters' interpretation as average effects among the connected regions.

# 4 Results

Identification of the CER-Express's impact in our SDID setup, just like the original DID methodology, hinges on credible retention of the parallel trends assumption (PTA, Angrist and Pischke 2009). Since trends after treatment are, by definition, only observed for one manifestation of $D$ in the control and treatment group respectively, we base our discussion around visual investigation of trends in the pre-treatment period. Accordingly, Figure 3 tracks logged p.c. $CO_2$ emissions in both groups from 2003 to 2018.

The divergent pre-trends in the initial periods of our sample raise some concern about the validity of the PTA. Since this divergence seemingly ceases after 2007, we argue that it does

Figure 3: Pre- and post-treatment trends




*Average logged p.c. $CO_2$ emissions (excluding short cycle), 2003 to 2018*

not imply fundamental differences in both groups' development. Moreover, this potential source of bias is purged by curtailing the sample as we drop the first four periods in our preferred specification. Full sample results are reported in Tables A.3 and A.4. Notice that this piece of graphical evidence does not account for spatial dependency as we were suggesting in equation (1). All regions are connected to all access points through the spatial lags. Thus, the control group includes regions that are spatially treated with non-zero weights in $W$. The distance based definition of neighborhood utilized in our approach implies that there is no region that is not treated at all as all regions are modelled to be interconnected to some extent. Additionally, Figure 3 does not adequately portray the inherent staggered implementation within the treatment group. The treatment period 2011 only identifies the initial connection of Duisburg, Germany to the CER-railway system. Additional regions are included in later years as portrayed in Table 1.

## Main Findings

Table 2 presents basic results for the SDM and SDEM models and the four different decay rates[6]. The previously discussed tests and criteria used for model selection, both in terms of

[6] Additional controls displayed in Table A.5 exhibit the expected signs The direct effects estimated in rows one to four indicate a significantly positive relationship with p.c. GDP and the secondary sector share, while areas more densely populated and more active in the primary sector showcase significantly lower emission levels. For the spatially lagged values (rows five to eight) of the same four variables only the secondary sector share does not exhibit significant point estimates throughout all specifications. For p.c. GDP the sign remains unchanged, indicating that vicinity to wealthier regions is associated with heightened emissions in any of the $N$ regions. For population density and the primary sector share signs switch, suggesting that areas with relatively agricultural and populous neighboring regions are more likely to exhibit more carbon intensive economic activity. Another pattern that can be identified here, relevant also for the examination of the treatment dummy, is that the effect size tends to decrease in the exponent of the decay rate.

spatial component inclusion and distance decay rate used in $W$, have been displayed in the bottom panel.

The test statistics suggest using a $W$ with a $\frac{1}{x^2}$ decay rate as preferred specification. The simple LM-Tests unequivocally advise for inclusion of either spatial component. The locally robust LM-Tests allow identifying spatial lag and error dependence. However, each of both tests is assuming that the respectively other feature is present in the data. Put differently, we cannot test both features simultaneously. Given the previously described drawbacks of these tests, we still interpret the frequent rejections of the $H_0$ as sufficient evidence for considering the SDM and SDEM models as superior to the SLX approach. Computing BIC for both models leaves us with consistently lower values for the SDM model. The lowest one is computed in column 3 and values are increasing in both, a larger and a smaller exponent. This is why we focus on the results obtained from this model setup.

Table 2: Curtailed sample results

| *Dependent variable Logged p.c. $CO_2$ emissions (kg)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $W$ decay rate / Variable | $\frac{1}{x}$ (1) | (2) | $\frac{1}{x^2}$ (3) | (4) | $\frac{1}{x^3}$ (5) | (6) | $\frac{1}{x^4}$ (7) | (8) |
| Treatment | -0.029* | -0.027 | -0.035** | -0.024 | -0.033** | -0.028 | -0.032* | -0.031* |
| | (0.017) | (0.017) | (0.016) | (0.018) | (0.016) | (0.017) | (0.017) | (0.017) |
| Treatment *(slag)* | 1.306*** | 1.314*** | 0.12*** | 0.122*** | 0.053** | 0.045* | 0.028 | 0.018 |
| | (0.332) | (0.369) | (0.041) | (0.045) | (0.024) | (0.027) | (0.02) | (0.022) |
| $\lambda$ | 0.902*** | | 0.511*** | | 0.306*** | | 0.245*** | |
| | (0.014) | | (0.019) | | (0.012) | | (0.01) | |
| $\rho$ | | 0.926*** | | 0.529*** | | 0.308*** | | 0.244*** |
| | | (0.014) | | (0.019) | | (0.012) | | (0.01) |
| R-sqrt | 0.977 | 0.977 | 0.978 | 0.977 | 0.978 | 0.976 | 0.978 | 0.976 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |
| LM test for spatial lag dep. | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | |
| Locally robust LM test for spatial lag dep. sub spatial err. | ✗ (0.3281) | | ✗ (0.112) | | ✓ (0.005223) | | ✓ (0.0001158) | |
| LM test for spatial err. dep. | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | | ✓ (< 2.2e-16) | |
| Locally robust LM test for spatial err. dep. sub spatial lag | ✓ (< 2.2e-16) | | ✓ (1.336e-08) | | ✗ (0.7608) | | ✗ (0.05154) | |
| BIC | -13854.8 | 75445.75 | -14059.38 | 75256.67 | -14018.03 | 75330.21 | -13955.26 | 75397.85 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are *p<0.1; **p<0.05; ***p<0.01. ✓ indicates rejection of the $H_0$ at the 5% level in favor of spatial component inclusion, ✗ failure to do so. P-values are added in parentheses.

Evaluating the CER-Express treatment dummy, the direct impact estimated in row one is always negative and significant for the SDM models, varying between approximately -2.9 percent in column 1 and -3.5 percent in column 3. For the SDEM models only the estimate in column 8 is significant at the 10 percent level and quantitatively similar to SDM alternatives.

For the lagged treatment, the impact, as for the controls, decreases with the exponent in the decay rate and is no longer significant for the models in columns 7 and 8. In column 3 this entails a significant average increase of approximately 12.7 percent. For the next best specification in column 5 it is ≈5.4 percent. Spatial components included ($\lambda$ and $\rho$) are consistently positive and strongly significant in all models. This serves as further affirmation for their inclusion.

Substantially, this result can be interpreted as meaningful redeployment of carbon intensive economic activity (*Hypothesis 1*). Within the connected node regions, more pollution intensive businesses are driven out by other, potentially service-oriented entities. They relocate to nearby areas, whose vicinity to the node regions presents an incentive for additional industries to shift their production here (*Hypothesis 2*).

## Sample adjustments

Although effect size deviates to some extent from previously identified values for the uncurtailed sample also including years before 2007 (Table A.4), the results are qualitatively similar. For the preferred iteration, which according to Table A.3 is still the SDM model with a $W$-Matrix characterized through a $\frac{1}{x^2}$ decay rate, it is retained at the 1 percent level for the direct and lagged treatment effect. The associated point estimates at ≈-8.2 and ≈33.1 percent are substantially larger than their counterparts identified in Table 2. However, given the lack of parallelism in the uncurtailed sample's initial periods, we are less convinced about these results unbiasedness.

Also estimates obtained from an expanded sample model including regions in the UK, which, consequentially, includes another region into the treatment group, while dropping primary and secondary sector shares from $X$ in equation (1), do not allow for qualitatively essentially different conclusions. While test outcomes presented in Table A.6 advice for use of the same SDM specification as before, treatment effects presented in Table A.7 are slightly inflated from the models excluding the UK. However, it could be argued that these differences are driven by the remainder of regressors picking up the now unobserved structural information.

Ultimately, adjusting the analysed sample in terms of either $T$, $N$ or $X$ does serve to uphold the initially drawn conclusions. Quantitatively though, point estimates are arguably exposed to more sources of bias in Tables A.4 and A.7, which is why alternatives in Table 2 are considered more reliable.

## Mechanisms

As we have established a measurable relationship between the roll-out of CER-Express and spatial patterns of human-induced $CO_2$ emissions, the underlying mechanisms deserve some elucidation.

The most straightforward way to do this is an investigation of the structural characteristics of the affected regions. This concerns primarily the trajectory of the secondary sector, which, presumably, is responsible for the detected environmental burden associated with CER-Express. Since the originally facilitated share variable is bounded, linear estimation might fit values outside of its support (Migliorati et al. 2018). Therefore, we replicate the previous steps of testing (see Table A.8) and estimation (see Table 3[7]) for logged values of secondary sector output. The preferred model in column 3 identifies a pattern similar to the previously discussed coefficients of Table 2. Manufacturing output is decreasing in the nodal regions, while it increases in their vicinity, although the latter effect is only significant at the ten percent level and measurably smaller than the increase in $CO_2$ emissions.

Table 3: Curtailed sample results, secondary sector output

| *Dependent variable Logged Secondary sector output (1000 Euros)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable \ $W$ decay rate | $\frac{1}{x}$ | | $\frac{1}{x^2}$ | | $\frac{1}{x^3}$ | | $\frac{1}{x^4}$ | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment | -0.03** | -0.031*** | -0.027** | -0.022* | -0.029*** | -0.028** | -0.03*** | -0.03** |
| | (0.012) | (0.011) | (0.011) | (0.012) | (0.011) | (0.012) | (0.011) | (0.012) |
| Treatment *(slag)* | -0.198 | -0.196 | 0.046* | 0.023 | 0.022 | 0.005 | 0.013 | -0.001 |
| | (0.229) | (0.253) | (0.027) | (0.031) | (0.016) | (0.019) | (0.013) | (0.016) |
| $\lambda$ | 0.896*** | | 0.783*** | | 0.518*** | | 0.425*** | |
| | (0.006) | | (0.01) | | (0.01) | | (0.009) | |
| $\rho$ | | 0.978*** | | 0.826*** | | 0.521*** | | 0.425*** |
| | | (0.004) | | (0.012) | | (0.01) | | (0.009) |
| R-sqrt | 0.996 | 0.996 | 0.997 | 0.996 | 0.997 | 0.996 | 0.997 | 0.996 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01.

Table 4: Curtailed sample results, GDP p.c.

| *Dependent variable Logged GDP p.c. (Euros)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable \ $W$ decay rate | $\frac{1}{x}$ | | $\frac{1}{x^2}$ | | $\frac{1}{x^3}$ | | $\frac{1}{x^4}$ | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment | 0.037*** | 0.041*** | 0.021*** | 0.021*** | 0.021*** | 0.02** | 0.023*** | 0.019** |
| | (0.009) | (0.009) | (0.007) | (0.008) | (0.006) | (0.009) | (0.007) | (0.009) |
| Treatment *(slag)* | 2.209*** | 2.107*** | 0.045*** | 0.039* | 0.013 | 0.017 | 0.008 | 0.008 |
| | (0.176) | (0.199) | (0.017) | (0.02) | (0.009) | (0.013) | (0.008) | (0.012) |
| $\lambda$ | 0.992*** | | 0.994*** | | 0.825*** | | 0.738*** | |
| | (0.001) | | (0.001) | | (0.005) | | (0.005) | |
| $\rho$ | | 0.993*** | | 0.994*** | | 0.835*** | | 0.746*** |
| | | (0.001) | | (0.001) | | (0.005) | | (0.005) |
| R-sqrt | 0.994 | 0.991 | 0.996 | 0.99 | 0.997 | 0.99 | 0.997 | 0.99 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01.

Trying to answer the question if and how this result translates into aggregate growth dynam-

[7] Additional controls displayed in Table A.9

ics, Table 4[8] presents results for p.c. GDP as the dependent variable. Despite the previously identified increase in emissions p.c. and secondary sector output, no identifiable effect can be discerned for the preferred model specification (for test outcomes see Table A.10). For the nodal regions however, the effect is positive (≈2.3 percent) and strongly significant.

## Dynamics

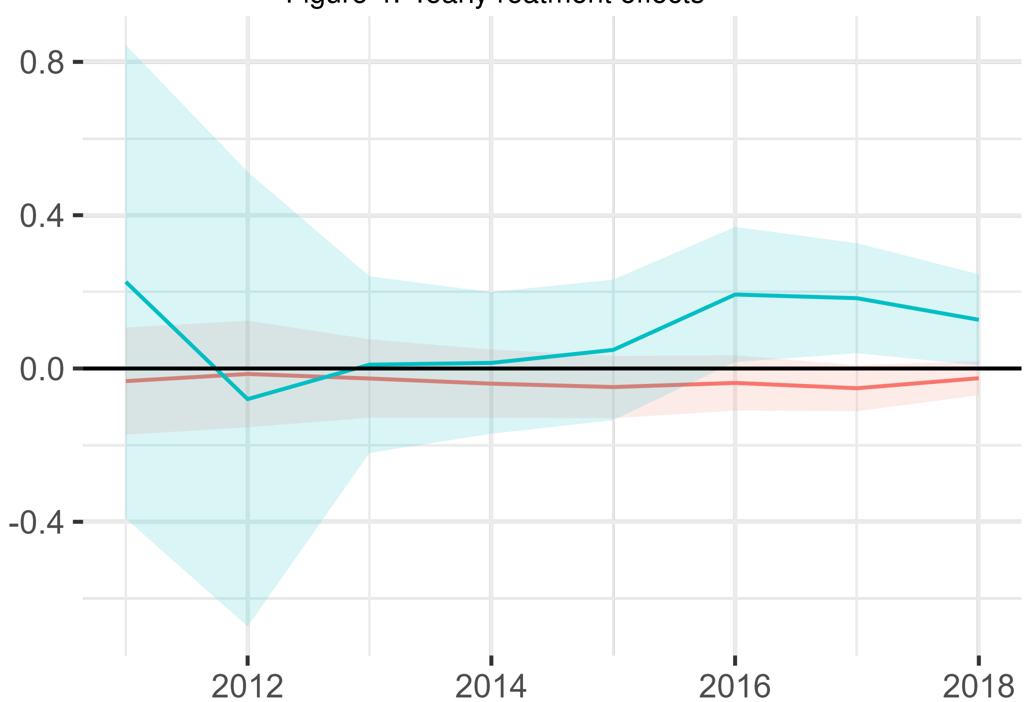
Given the dynamic nature of CER-Express's roll-out, one reasonable assumption is that effect size varies over time. More precisely, it can be expected that, as the number of trips and average facilitation increases, the impact on regional economic activity grows, causing $CO_2$ emissions to rise.

Interacting both treatment indicators with the matrix of time dummies $\theta_t$ produces the modified regression equation for our preferred SDM

$$Y_t = (\alpha_t + W\beta_t)D_t\theta_t + \lambda WY_t + (\mu + W\nu)X_t + \phi + \theta_t + \Xi_t, \tag{2}$$

The new sets of parameters $\alpha_t$ and $\beta_t$, which are accordingly estimated as the average yearly direct and spatially lagged treatment effects, have been plotted in Figure 4. The graph reveals several interesting findings refining previous results

Figure 4: Yearly reatment effects




*Point estimates (solid line) and 95% confidence intervals (shaded area) for the direct (red) and spatially lagged (blue) treatment. Adjustments corresponding to the model represented in column (3) of Table 2*

[8]Additional controls displayed in Table A.11

A common feature, which is however much more pronounced for the spatially lagged effects, is that precision, depicted here by the width of the confidence band, improves over time. This is most likely caused by the increase in observations with $D = 1$ in later periods.

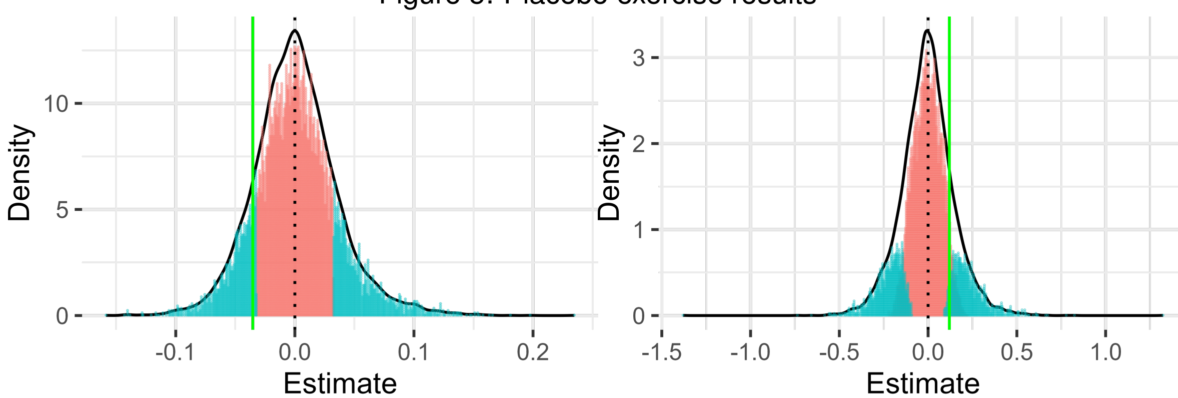
For the direct treatment, i.e. the effects observed within the nodal points listed in Table 1, yearly estimates are in fact non-discernible form zero at the 5-percent level. This undermines credibility of the previously identified negative average effect. Potentially, this mismatch is driven by excessive weights on some of the periods with slightly significant estimates as discussed by Goodman-Bacon (2021).

For the spatially lagged treatment, yearly effects only become significant and positive starting from 2016. This realization is in line with the continued growth of CER-Express. Moreover, this finding also aligns with recent research identifying the lagged realization of infrastructural policies' outcomes (Lindgren et al. 2021).

## Placebo exercise

To further scrutinise previous results, a placebo exercise randomly assigns treatment to an equally large group of regions from the original control group following an allocation pattern identical to the one depicted in Table 1.

Figure 5: Placebo exercise results




(a) Direct (b) Spatially lagged

**Estimate distribution for 10,000 iterations of randomly assigned treatment.** *Treatment assigned following the same allocation pattern depicted in Table 1 and using the same specification as in column (3) of Table 2. The density function maps over all estimates, while estimates are grouped by significance at the 5% level (red above, blue below). Green lines shows originally estimated coefficients.*

The resulting patterns of both distributions are centered around zero, as their mean clearly deviates from coefficients identified in column (3) of Table 2. Also, most placebo assignments produce insignificant point estimates.

For Figures 5a and 5b respectively, 485 and 33 significant placebo coefficients are located within a one percent window around the actually identified estimates. This supports the original expectation that there is no systematically unobserved information driving the effect, especially for the spatially lagged treatment.

# 5 Conclusion

Reflecting on our findings, one can reasonably attest that there is sufficient evidence for an identifiable environmental impact of the transport embodied within CER-Express. These potentially detrimental effects should be figured in when considering political decisions on further and increased local participation. However, it can also be concluded that there is substantial room for trade-offs, both between those regions prompting participation decisions and its indirectly affected neighbors, as well as between economic and environmental goals.

Projecting the potential future trajectory of the identified environmental repercussions, numbers published at the height of the Covid-19 pandemic are indicating that, in terms of utilization, CER-Express seemingly profited from increased rates cited for competing modes of transport, especially maritime freight[9]. Reportedly, the number of trips between the PRC and the EU grew from 8225 in 2019[10] to 12406 in 2020[11] and 15000 in 2021[12], suggesting ample room for continued growth.

More recently though, a number of challenges are threatening to derail the project, rendering its continuation infeasible. These include the potential unreliability of routes running through Russian territory following the country's war with Ukraine and ensuing sanctions, potential supply chain issues, and, from a long term perspective, the looming scenario of deglobalistaion induced by altercations between China on the one hand and the United States of America (USA) and its allies on the other. However, despite these challenges, most recent numbers published for the year 2022 indicate a substantial degree of robustness in operations (16000 trips[13]).

# References

Ahlfeldt, Gabriel M. and Arne Feddersen (2018). "From periphery to core: measuring agglomeration effects using high-speed rail". In: *Journal of Economic Geography* 18.2, pp. 355–390. DOI: 10.1093/jeg/lbx005.

Angrist, Joshua David and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, N.J. and Oxford: Princeton University Press. DOI: 10.1515/9781400829828.

Baldwin, Richard E. and Rikard Forslid (2000). "The Core-Periphery Model and Endogenous Growth: Stabilizing and Destabilizing Integration". In: *Economica* 67.267, pp. 307–324. DOI: 10.1111/1468-0335.00211.

Behrens, Kristian, W. Mark Brown, and Théophile Bougna (2018). "The World Is Not Yet Flat: Transport Costs Matter!" In: *The Review of Economics and Statistics* 100.4, pp. 712–724. DOI: 10.1162/rest_a_00729.

[9] https://unctad.org/news/shipping-during-covid-19-why-container-freight-rates%2Dhave-surged (visited on 08/16/2022)

[10] http://kz.mofcom.gov.cn/article/jmxw/202001/20200102929219.shtml (in Chinese, visited on 08/16/2022)

[11] http://www.gov.cn/xinwen/2021-01/19/content_5581186.htm (in Chinese, visited on 08/16/2022)

[12] http://www.gov.cn/xinwen/2022-01/05/content_5666468.htm (in Chinese, visited on 08/16/2022)

[13] http://www.gov.cn/xinwen/2023-01/04/content_5734924.htm (in Chinese, visited on 03/15/2023)

Burridge, Peter, J. Paul Elhorst, and Katarina Zigova (2016). “Group Interaction in Research and the Use of General Nesting Spatial Models”. In: *Spatial econometrics*. Ed. by Badi H. Baltagi, James P. Lesage, and R. Kelley Pace. Vol. 37. Bingley, UK: Emerald Group Publishing Limited, pp. 223–258. DOI: 10.1108/S0731-905320160000037016.

Chabé-Ferret, Sylvain (2015). “Analysis of the bias of Matching and Difference-in-Difference under alternative earnings and selection processes”. In: *Journal of Econometrics* 185.1, pp. 110–123. DOI: 10.1016/j.jeconom.2014.09.013.

Chagas, André L.S., Carlos R. Azzoni, and Alexandre N. Almeida (2016). “A spatial difference-in-differences analysis of the impact of sugarcane production on respiratory diseases”. In: *Regional Science and Urban Economics* 59, pp. 24–36. DOI: 10.1016/j.regsciurbeco.2016.04.002.

Chen, Dengke, Shiyi Chen, and Hao Jin (2018). “Industrial agglomeration and CO2 emissions: Evidence from 187 Chinese prefecture-level cities over 2005–2013”. In: *Journal of Cleaner Production* 172, pp. 993–1003. DOI: 10.1016/j.jclepro.2017.10.068.

Cheng, Zhonghua (2016). “The spatial correlation and interaction between manufacturing agglomeration and environmental pollution”. In: *Ecological Indicators* 61, pp. 1024–1032. DOI: 10.1016/j.ecolind.2015.10.060.

Crippa, Monica, Gabriel Oreggioni, Diego Guizzardi, Marilena Muntean, Edwin Schaaf, Eleonora Lo Vullo, Efisio Solazzo, Fabio Monforti-Ferrario, Jos Olivier, and Elisabetta Vignati (2019). *Fossil CO2 and GHG emissions of all world countries: 2019 Report*. Vol. JRC117610. Luxembourg: Publications Office of the European Union. DOI: 10.2760/687800.

Dong, Feng, Yue Wang, Lu Zheng, Jingyun Li, and Shouxiang Xie (2020). “Can industrial agglomeration promote pollution agglomeration? Evidence from China”. In: *Journal of Cleaner Production* 246, p. 118960. DOI: 10.1016/j.jclepro.2019.118960.

Elhorst, J. Paul (2014). “Matlab Software for Spatial Panels”. In: *International Regional Science Review* 37.3, pp. 389–405. DOI: 10.1177/0160017612452429.

Eurostat (2020). *Statistical regions in the European Union and partner countries: NUTS and statistical regions 2021 : 2020 edition*. 2020 edition. Luxembourg: Publications Office of the European Union. DOI: 10.2785/850262.

Fang, Ling, Martin Kleimann, Yuan Li, and Hans-Jörg Schmerer (2021). “The implications of the New Silk Road Railways on local development”. In: *Journal of Asian Economics* 75, p. 101326. DOI: 10.1016/j.asieco.2021.101326.

Goodman-Bacon, Andrew (2021). “Difference-in-differences with variation in treatment timing”. In: *Journal of Econometrics* 225.2, pp. 254–277. DOI: 10.1016/j.jeconom.2021.03.014.

Grazi, Fabio, Jeroen C. J. M. van den Bergh, and Piet Rietveld (2007). “Spatial welfare economics versus ecological footprint: modeling agglomeration, externalities and trade”. In: *Environmental and Resource Economics* 38.1, pp. 135–153. DOI: 10.1007/s10640-006-9067-2.

Halleck Vega, Solmaria and J. Paul Elhorst (2015). “The SLX Model”. In: *Journal of regional science* 55.3, pp. 339–363. DOI: 10.1111/jors.12188.

Harris, Richard, John Moffat, and Victoria Kravtsova (2011). “In Search of ‘ W ’”. In: *Spatial Economic Analysis* 6.3, pp. 249–270. DOI: 10.1080/17421772.2011.586721.

Krugman, Paul (1991). “Increasing Returns and Economic Geography”. In: *Journal of Political Economy* 99.3, pp. 483–499. DOI: 10.1086/261763.

Lindgren, Erik, Per Pettersson-Lidbom, and Björn Tyrefors (2021). *The Causal Effect of Transport Infrastructure: Evidence from a New Historical Database*. Working Paper 1407. Research Institute of Industrial Economics (IFN). URL: https://www.ifn.se/en/publications/working-papers/2021/1407/ (visited on 12/13/2022).

Liu, Zheming, Saixing Zeng, Zhizhou Jin, and Jonathan Jingsheng Shi (2022). "Transport infrastructure and industrial agglomeration: Evidence from manufacturing industries in China". In: *Transport Policy* 121, pp. 100–112. DOI: 10.1016/j.tranpol.2022.04.001.

Mau, Karsten and Rosalie Seuren (2023). "One belt, one road, one way? Where European exporters benefit from the new silkroad". In: *Review of World Economics*. DOI: 10.1007/s10290-022-00467-0.

Migliorati, Sonia, Agnese Maria Di Brisco, and Andrea Ongaro (2018). "A New Regression Model for Bounded Responses". In: *Bayesian Analysis* 13.3. DOI: 10.1214/17-BA1079.

Rubin, Donald B. (1980). "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment". In: *Journal of the American Statistical Association* 75.371, p. 591. DOI: 10.2307/2287653.

Schwarz, Gideon (1978). "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2. DOI: 10.1214/aos/1176344136.

Silver, M. S. and P. Golder (1981). "Negative Value Added and the Measurement of Production Changes". In: *Journal of Economic Studies* 8.1, pp. 3–15. DOI: 10.1108/eb002519.

# Appendix

Table A.1: Summary Statistics

| Variable | Unit | Comparison | Mean | sd | min | max |
|---|---|---|---|---|---|---|
| $CO_2$ emissions p.c. | kg | Overall | 8668.7995 | 11261.43 | 1157.13 | 232332.39 |
| | | Between | | 11002.20 | 1227.21 | 197368.96 |
| | | Within | | 2423.42 | -34856.55 | 55991.33 |
| GDP p.c. | Euros | Overall | 24373.7385 | 14048.03 | 1237.18 | 183778.44 |
| | | Between | | 13592.29 | 2653.23 | 118400.94 |
| | | Within | | 3571.01 | -17517.68 | 89751.23 |
| Density | Inhabitans per km$^2$ | Overall | 404.0980 | 1061.03 | 1.81 | 21411.24 |
| | | Between | | 1060.64 | 1.85 | 21008.27 |
| | | Within | | 42.14 | -1087.51 | 1666.13 |
| Primary sector share | % | Overall | 3.5923 | 3.91 | -3.61 | 44.01 |
| | | Between | | 3.74 | 0.00 | 25.75 |
| | | Within | | 1.14 | -8.39 | 24.17 |
| Secondary sector share | % | Overall | 30.1275 | 10.35 | 3.39 | 78.51 |
| | | Between | | 10.04 | 5.63 | 73.27 |
| | | Within | | 2.53 | 11.33 | 51.98 |
| Treatment | Dummy | Overall | 0.0035 | 0.06 | 0.00 | 1.00 |
| | | Between | | 0.03 | 0.00 | 0.50 |
| | | Within | | 0.05 | -0.50 | 0.94 |
| | | N = 17872 | n = 1117 | T-bar = 16 | | |

*Note* The displayed negative minimum value of the overall primary sector share stems from the Swedish Kronoberg County, which, alongside the adjacent Jönköping County, recorded a negative gross value added in agricultural production for 2005. This indicates that for this year sales were less valuable than bought-in goods and services (Silver and Golder 1981).

Table A.2: Correlation matrix

| | $CO_2$ emissions p.c. | GDP p.c. | Density | Primary sector share | Secondary sector share | Treatment |
|---|---|---|---|---|---|---|
| $CO_2$ emissions p.c. | 1.0000 | 0.0220 | -0.0509 | -0.0350 | 0.2116 | 0.0046 |
| GDP p.c. | 0.0220 | 1.0000 | 0.3605 | -0.5299 | -0.0498 | 0.0503 |
| Density | -0.0509 | 0.3605 | 1.0000 | -0.2552 | -0.2149 | 0.0808 |
| Primary sector share | -0.0350 | -0.5299 | -0.2552 | 1.0000 | -0.0390 | -0.0286 |
| Secondary sector share | 0.2116 | -0.0498 | -0.2149 | -0.0390 | 1.0000 | -0.0505 |
| Treatment | 0.0046 | 0.0503 | 0.0808 | -0.0286 | -0.0505 | 1.0000 |

Table A.3: Spatial dependency testing outcomes, uncurtailed sample

| | $\frac{1}{x}$ | $\frac{1}{x^2}$ | $\frac{1}{x^3}$ | $\frac{1}{x^4}$ |
|---|---|---|---|---|
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial lag dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✗ | ✓ | ✗ | ✗ |
| spatial lag dependence sub spatial error | (0.1409) | (0.01885) | (0.251) | (0.1219) |
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial error dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✓ | ✓ | ✗ | ✗ |
| spatial error dependence sub spatial lag | (< 2.2e-16) | (3.085e-10) | (0.1112) | (0.5576) |
| BIC SDM | -13828.35 | -14119.26 | -14105.99 | -14023.1 |
| BIC SDEM | 110400.5 | 110105.2 | 110156.1 | 110244.9 |

*Note* ✓ indicates rejection of the $H_0$ at the 5% level in favor of spatial component inclusion, ✗ failure to do so. P-values are added in parentheses.

Table A.4: Uncurtailed sample results

| *Dependent variable Logged p.c. $CO_2$ emissions (kg)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $W$ decay rate / Variable | $\frac{1}{x}$ (1) | (2) | $\frac{1}{x^2}$ (3) | (4) | $\frac{1}{x^3}$ (5) | (6) | $\frac{1}{x^4}$ (7) | (8) |
| Treatment | -0.075561*** | -0.07071*** | -0.085632*** | -0.053729*** | -0.081296*** | -0.063108*** | -0.078613*** | -0.068103*** |
| | (0.019178) | (0.019248) | (0.018939) | (0.02009) | (0.018935) | (0.019999) | (0.018984) | (0.019825) |
| Treatment *(slag)* | 2.392641*** | 2.524796*** | 0.285932*** | 0.325097*** | 0.138167*** | 0.133606*** | 0.094013*** | 0.082138*** |
| | (0.383295) | (0.425659) | (0.047501) | (0.051807) | (0.027871) | (0.030444) | (0.023514) | (0.025303) |
| ln(GDP p.c.) | 0.147251*** | 0.141052*** | 0.121875*** | 0.122159*** | 0.088414*** | 0.108799*** | 0.088735*** | 0.113129*** |
| | (0.016341) | (0.016951) | (0.018642) | (0.018073) | (0.019579) | (0.01835) | (0.019556) | (0.018173) |
| ln(Density) | -1.042822*** | -1.051717*** | -0.971371*** | -0.966645*** | -0.937699*** | -0.936759*** | -0.924913*** | -0.926627*** |
| | (0.035957) | (0.036545) | (0.038254) | (0.03642) | (0.039218) | (0.036699) | (0.039286) | (0.036775) |
| Primary sector share | -0.002914** | -0.003638*** | -0.005638*** | -0.004811*** | -0.004917*** | -0.003407*** | -0.003771*** | -0.002311* |
| | (0.001165) | (0.001186) | (0.001243) | (0.001216) | (0.001266) | (0.001206) | (0.001252) | (0.001185) |
| Secondary sector share | 0.002757*** | 0.002693*** | 0.002748*** | 0.00294*** | 0.003356*** | 0.003469*** | 0.003569*** | |
| | (0.000464) | (0.000467) | (0.000473) | (0.00046) | (0.000478) | (0.000461) | (0.000477) | (0.00046) |
| ln(GDP p.c.) *(slag)* | 0.565894*** | 1.007001*** | 0.180127*** | 0.381622*** | 0.19626*** | 0.277641*** | 0.193379*** | 0.246148*** |
| | (0.091864) | (0.110376) | (0.032954) | (0.036964) | (0.024055) | (0.024158) | (0.021863) | (0.021158) |
| ln(Density) *(slag)* | 1.858943*** | 1.384948*** | 0.532578*** | 0.013349 | 0.314227*** | 0.030125 | 0.249196*** | 0.037948 |
| | (0.212411) | (0.276771) | (0.077878) | (0.087678) | (0.05436) | (0.05439) | (0.048379) | (0.04701) |
| Primary sector share *(slag)* | 0.064599*** | 0.092515*** | 0.027342*** | 0.036368*** | 0.019643*** | 0.021976*** | 0.016756*** | 0.018053*** |
| | (0.006587) | (0.008333) | (0.002513) | (0.003095) | (0.001816) | (0.001938) | (0.00161) | (0.001633) |
| Secondary sector share *(slag)* | -0.003366 | 0.002281 | 4.2e-05 | 0.001556 | 7.5e-05 | 0.000998 | -9.3e-05 | 0.000658 |
| | (0.003747) | (0.004661) | (0.001124) | (0.000736) | (0.000784) | (0.000636) | (0.000652) | |
| $\lambda$ | 0.903648*** | | 0.489851*** | | 0.295385*** | | 0.233044*** | |
| | (0.013832) | | (0.016861) | | (0.010626) | | (0.008866) | |
| $\rho$ | | 0.919592*** | | 0.510796*** | | 0.298006*** | | 0.233433*** |
| | | (0.013555) | | (0.016724) | | (0.010652) | | (0.008891) |
| R-sqrt | 0.9649 | 0.9639 | 0.9658 | 0.9637 | 0.9658 | 0.9637 | 0.9656 | 0.9637 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 17872 | 17872 | 17872 | 17872 | 17872 | 17872 | 17872 | 17872 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01.

## Table A.5: Curtailed sample results (cont.)

| Dependent variable Logged p.c. $CO_2$ emissions (kg) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $W$ decay rate / Variable | $\frac{1}{x}$ (1) | (2) | $\frac{1}{x^2}$ (3) | (4) | $\frac{1}{x^3}$ (5) | (6) | $\frac{1}{x^4}$ (7) | (8) |
| ln(GDP p.c.) | 0.206075*** | 0.190592*** | 0.146875*** | 0.150454*** | 0.114297*** | 0.144698*** | 0.125487*** | 0.160772*** |
| | (0.0201) | (0.020704) | (0.022386) | (0.021589) | (0.023055) | (0.021572) | (0.022775) | (0.021159) |
| ln(Density) | -1.242377*** | -1.256577*** | -1.228574*** | -1.227452*** | -1.227026*** | -1.205845*** | -1.217567*** | -1.193098*** |
| | (0.043312) | (0.043993) | (0.046006) | (0.043708) | (0.047048) | (0.043867) | (0.047044) | (0.043854) |
| Primary sector share | -0.002398* | -0.002524* | -0.003604** | -0.003405** | -0.003581** | -0.002905** | -0.003378** | -0.002628* |
| | (0.00145) | (0.00146) | (0.001498) | (0.001462) | (0.001513) | (0.001458) | (0.001504) | (0.001444) |
| Secondary sector share | 0.001021** | 0.001136** | 0.001551*** | 0.001918*** | 0.002326*** | 0.002472*** | 0.00262*** | 0.002644*** |
| | (0.000501) | (0.000505) | (0.000516) | (5e-04) | (0.00052) | (5e-04) | (0.000517) | (0.000497) |
| ln(GDP p.c.) *(slag)* | 0.583395*** | 1.227689*** | 0.256318*** | 0.557665*** | 0.279974*** | 0.400569*** | 0.265463*** | 0.344715*** |
| | (0.10728) | (0.131554) | (0.040022) | (0.045492) | (0.029031) | (0.029253) | (0.026022) | (0.025284) |
| ln(Density) *(slag)* | 1.658691*** | 1.056118*** | 0.797692*** | 0.30172*** | 0.586196*** | 0.255697*** | 0.512683*** | 0.243838*** |
| | (0.236056) | (0.310725) | (0.088179) | (0.100318) | (0.06334) | (0.062904) | (0.056852) | (0.054749) |
| Primary sector share *(slag)* | 0.041411*** | 0.057536*** | 0.016966*** | 0.025529*** | 0.01211*** | 0.01399*** | 0.010362*** | 0.011228*** |
| | (0.012187) | (0.01562) | (0.004015) | (0.005028) | (0.00264) | (0.002894) | (0.002242) | (0.002332) |
| Secondary sector share *(slag)* | 0.00996** | 0.015261*** | 0.001845 | 0.002354 | -0.000121 | 7.7e-05 | -0.000805 | -0.000629 |
| | (0.004461) | (0.005515) | (0.001259) | (0.001492) | (0.000812) | (0.000872) | (0.000698) | (0.000721) |
| R-sqrt | 0.9773 | 0.9766 | 0.9779 | 0.9765 | 0.9779 | 0.9765 | 0.9777 | 0.9764 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

## Table A.6: Spatial dependency testing outcomes, curtailed sample including UK

| | $\frac{1}{x}$ | $\frac{1}{x^2}$ | $\frac{1}{x^3}$ | $\frac{1}{x^4}$ |
|---|---|---|---|---|
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial lag dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✗ | ✗ | ✗ | ✓ |
| spatial lag dependence sub spatial error | (0.09307) | (0.1554) | (0.4406) | (0.01966) |
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial error dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✓ | ✓ | ✗ | ✗ |
| spatial error dependence sub spatial lag | (< 2.2e-16) | (1.091e-07) | (0.2181) | (0.3943) |
| BIC SDM | -14158.45 | -14449.12 | -14280.7 | -14172.82 |
| BIC SDEM | 90952.32 | 90650.54 | 90844.56 | 90958.98 |

*Note* ✓ indicates rejection of the $H_0$ at the 5% level in favor of spatial component inclusion, ✗ failure to do so. P-values are added in parentheses.

Table A.7: Curtailed sample results including UK

| Variable \ $W$ decay rate | $\frac{1}{x}$ (1) | (2) | $\frac{1}{x^2}$ (3) | (4) | $\frac{1}{x^3}$ (5) | (6) | $\frac{1}{x^4}$ (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| *Dependent variable Logged p.c. $CO_2$ emissions (kg)* | | | | | | | | |
| Treatment | -0.038223** | -0.033894** | -0.045575*** | -0.036468** | -0.04025** | -0.035552* | -0.038172** | -0.036749** |
| | (0.017188) | (0.017249) | (0.016866) | (0.018197) | (0.016964) | (0.018193) | (0.017041) | (0.018007) |
| Treatment *(slag)* | 2.314109*** | 2.255325*** | 0.162394*** | 0.124738*** | 0.081992*** | 0.055359* | 0.051214** | 0.029109 |
| | (0.358591) | (0.40343) | (0.043263) | (0.047999) | (0.02539) | (0.028247) | (0.021481) | (0.023472) |
| ln(GDP p.c.) | 0.243871*** | 0.217278*** | 0.18236*** | 0.174204*** | 0.147001*** | 0.165967*** | 0.1528*** | 0.176584*** |
| | (0.017179) | (0.017956) | (0.019029) | (0.018306) | (0.019613) | (0.018196) | (0.019372) | (0.017845) |
| ln(Density) | -1.167864*** | -1.189619*** | -1.178118*** | -1.212558*** | -1.226482*** | -1.241348*** | -1.250164*** | -1.255552*** |
| | (0.043217) | (0.044361) | (0.04625) | (0.043936) | (0.047673) | (0.043921) | (0.04765) | (0.043855) |
| ln(GDP p.c.) *(slag)* | 0.073241 | 0.680292*** | 0.061056* | 0.348664*** | 0.124944*** | 0.234689*** | 0.126528*** | 0.197499*** |
| | (0.073126) | (0.096642) | (0.031263) | (0.038102) | (0.023976) | (0.024486) | (0.021842) | (0.021244) |
| ln(Density) *(slag)* | 0.756854*** | -0.305421 | 0.550087*** | -0.103845 | 0.368092*** | -0.052823 | 0.32372*** | -0.010878 |
| | (0.17408) | (0.251194) | (0.079556) | (0.095629) | (0.061063) | (0.060122) | (0.055745) | (0.052859) |
| $\lambda$ | 0.958077*** | | 0.587178*** | | 0.344476*** | | 0.270044*** | |
| | (0.007327) | | (0.016377) | | (0.011014) | | (0.009307) | |
| $\rho$ | | 0.961587*** | | 0.59819*** | | 0.346262*** | | 0.270184*** |
| | | (0.007542) | | (0.016327) | | (0.011024) | | (0.009321) |
| R-sqrt | 0.9754 | 0.9739 | 0.9762 | 0.9739 | 0.976 | 0.974 | 0.9757 | 0.9739 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 15444 | 15444 | 15444 | 15444 | 15444 | 15444 | 15444 | 15444 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01.

Table A.8: Spatial dependency testing outcomes, curtailed sample, secondary sector output

| | $\frac{1}{x}$ | $\frac{1}{x^2}$ | $\frac{1}{x^3}$ | $\frac{1}{x^4}$ |
|---|---|---|---|---|
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial lag dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✗ | ✓ | ✓ | ✓ |
| spatial lag dependence sub spatial error | (0.6425) | (0.01683) | (2.057e-10) | (7.112e-15) |
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial error dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✓ | ✓ | ✗ | ✓ |
| spatial error dependence sub spatial lag | (< 2.2e-16) | (< 2.2e-16) | (0.1178) | (0.01246) |
| BIC SDM | -23718.17 | -25463.37 | -25462.87 | -25225.03 |
| BIC SDEM | 65163.47 | 63739.48 | 63907.29 | 64158.25 |

*Note* ✓ indicates rejection of the $H_0$ at the 5% level in favor of spatial component inclusion, ✗ failure to do so. P-values are added in parentheses.

Table A.9: Curtailed sample results, secondary sector output (cont.)

| *Dependent variable Logged Secondary sector output (1000 Euros)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $W$ decay rate | $\frac{1}{x}$ | | $\frac{1}{x^2}$ | | $\frac{1}{x^3}$ | | $\frac{1}{x^4}$ | |
| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ln(GDP p.c.) | 1.438691*** | 1.454884*** | 1.56034*** | 1.556312*** | 1.587329*** | 1.571526*** | 1.586689*** | 1.567751*** |
| | (0.012546) | (0.012619) | (0.012593) | (0.011786) | (0.012661) | (0.011558) | (0.012536) | (0.011358) |
| ln(Density) | 1.015987*** | 0.993606*** | 1.089481*** | 1.159804*** | 1.104656*** | 1.190267*** | 1.112396*** | 1.200394*** |
| | (0.029957) | (0.029952) | (0.029574) | (0.028024) | (0.030068) | (0.027992) | (0.030272) | (0.027949) |
| Primary sector share | -0.010389*** | -0.011032*** | -0.01276*** | -0.012348*** | -0.012941*** | -0.011989*** | -0.012533*** | -0.011469*** |
| | (0.000997) | (0.000987) | (0.000954) | (0.00092) | (0.000956) | (0.00092) | (0.000956) | (0.00092) |
| ln(GDP p.c.) *(slag)* | -1.253647*** | -0.002425 | -1.318022*** | -0.237292*** | -0.913224*** | -0.162195*** | -0.766648*** | -0.139778*** |
| | (0.058783) | (0.072777) | (0.025925) | (0.031533) | (0.021559) | (0.017943) | (0.019506) | (0.014817) |
| ln(Density) *(slag)* | 1.030805*** | 3.053548*** | -0.520465*** | 0.11052 | -0.333866*** | 0.142017*** | -0.235439*** | 0.188125*** |
| | (0.153423) | (0.201542) | (0.057046) | (0.073796) | (0.041471) | (0.042652) | (0.037431) | (0.036189) |
| Primary sector share *(slag)* | 0.046754*** | 0.060891*** | 0.021096*** | 0.023236*** | 0.014064*** | 0.01083*** | 0.011742*** | 0.007461*** |
| | (0.008403) | (0.010764) | (0.002565) | (0.00381) | (0.001678) | (0.002052) | (0.001435) | (0.001601) |
| R-sqrt | 0.9961 | 0.9957 | 0.9966 | 0.9957 | 0.9966 | 0.9957 | 0.9966 | 0.9957 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01.

Table A.10: Spatial dependency testing outcomes, curtailed sample, GDP p.c.

| | $\frac{1}{x}$ | $\frac{1}{x^2}$ | $\frac{1}{x^3}$ | $\frac{1}{x^4}$ |
|---|---|---|---|---|
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial lag dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial lag dependence sub spatial error | (< 2.2e-16) | (0.02146) | (< 2.2e-16) | (< 2.2e-16) |
| LM test for | ✓ | ✓ | ✓ | ✓ |
| spatial error dependence | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) | (< 2.2e-16) |
| Locally robust LM test for | ✓ | ✓ | ✗ | ✓ |
| spatial error dependence sub spatial lag | (< 2.2e-16) | (< 2.2e-16) | (0.07417) | (0.009465) |
| BIC SDM | -30341.07 | -36780.49 | -37238.26 | -36430.7 |
| BIC SDEM | 58982.18 | 52642.01 | 52193.73 | 53015.74 |

*Note* ✓ indicates rejection of the $H_0$ at the 5% level in favor of spatial component inclusion, ✗ failure to do so. P-values are added in parentheses.

Table A.11: Curtailed sample results, GDP p.c. (cont.)

| *Dependent variable Logged GDP p.c. (Euros)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $W$ decay rate | $\frac{1}{x}$ | | $\frac{1}{x^2}$ | | $\frac{1}{x^3}$ | | $\frac{1}{x^4}$ | |
| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ln(Density) | -0.658102*** | -0.617591*** | -0.582523*** | -0.59779*** | -0.557544*** | -0.542618*** | -0.558102*** | -0.548718*** |
| | (0.022612) | (0.023085) | (0.018275) | (0.017983) | (0.017577) | (0.01909) | (0.017874) | (0.019616) |
| Primary sector share | -0.009711*** | -0.008333*** | -0.004738*** | -0.006569*** | -0.004199*** | -0.006305*** | -0.004661*** | -0.007312*** |
| | (0.000778) | (0.000786) | (0.000616) | (0.000603) | (0.000585) | (0.000598) | (0.000591) | (0.000623) |
| Secondary sector share | 0.008836*** | 0.009356*** | 0.010981*** | 0.010737*** | 0.011437*** | 0.010851*** | 0.01159*** | 0.011041*** |
| | (0.000245) | (0.000246) | (0.000187) | (0.00019) | (0.000174) | (0.000197) | (0.000176) | (0.000205) |
| ln(Density) *(slag)* | -0.249936** | -1.405925*** | 0.460763*** | -0.082365* | 0.443031*** | 0.078865** | 0.399259*** | 0.049074* |
| | (0.124737) | (0.166571) | (0.03517) | (0.048549) | (0.023605) | (0.030661) | (0.021514) | (0.02683) |
| Primary sector share *(slag)* | -0.123588*** | -0.2079*** | -0.017572*** | -0.03702*** | -0.006861*** | -0.012663*** | -0.006476*** | -0.011523*** |
| | (0.006338) | (0.008157) | (0.001572) | (0.002664) | (0.000995) | (0.001531) | (0.000865) | (0.001194) |
| Secondary sector share *(slag)* | 0.025069*** | 0.039951*** | -0.011534*** | -0.005377*** | -0.010094*** | -0.002513*** | -0.008998*** | -0.00165*** |
| | (0.001846) | (0.002416) | (0.000453) | (0.000685) | (0.000282) | (0.000414) | (0.000248) | (0.000345) |
| R-sqrt | 0.9935 | 0.9915 | 0.9963 | 0.9905 | 0.9967 | 0.9902 | 0.9966 | 0.9903 |
| TWFE | YES | YES | YES | YES | YES | YES | YES | YES |
| Obs. | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 | 13404 |

*Note* All specifications use row standardized weight matrices. The term *slag* denotes spatially lagged variables. Significance levels as indicated through p-values are $^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01.