



**COLLEGIUM OF ECONOMIC ANALYSIS
WORKING PAPER SERIES**

Using geolocation data in spatial-
econometric construction of multiregion
input-output tables: a Bayesian approach

Andrzej Torój

Using geolocation data in spatial-econometric construction of multiregion input-output tables: a Bayesian approach

Andrzej Torój*

Interregional input-output tables for Poland at NUTS-3 level are built by using the Bayesian approach to spatial econometric analysis. I apply the multi-equation Durbin specification proposed by Torój (2021) to derive the sample density and Statistics Finland (2006) regional I–O tables to derive the prior hyperparameters. This prior aims to introduce additional information in the presence of noisy spatial data, but also to avoid the areas where the spatial decay profiles representing the supply geography become insensitive to the parameter values of the selected functional form. To measure the distance, the real-world driving distance between the most populated cities of the regions from Google Maps is used. Posterior distributions indicate that the agricultural commodities and advanced services are supplied to the most distant locations, whereas the simple services – to the least distant ones; the result for the former group of sectors is characterized with the highest uncertainty. The illustrative simulation indicates that 82.2% of the indirect effects occur in the home region, with a posterior-based confidence interval from 71.5% to 92.4%. The results do not change qualitatively when I use the driving time (averaged over 42 equidistant moments in a 7-day week) as the alternative measure of distance, but the hybrid time- and distance-based model is strongly preferred in the Bayes factor comparison, since for all sectors except industry (NACE sections B-E), the time-based metric turned out to be dominant. When commuting is taken into account in the induced effect calculation (measured with mobile geolocation data), 4.9% of the induced effects are relocated from the home region (central point in a big agglomeration) to the other regions, especially the surrounding “ring”.

JEL Classification: C31, C67, R12, R15.

Keywords: input-output, interregional input-output tables, spatial econometrics, Bayesian estimation, regional economic impact assessment.

1 Introduction and literature review

Inter-regional input–output tables (IRIO) on the subnational level remain a Holy Grail for researchers in regional economics. In the usual case of national I–O table availability from the statistical offices with

*SGH Warsaw School of Economics, Institute of Econometrics (andrzej.toroj@sgh.waw.pl), Al. Niepodległości 162, 02-554 Warszawa, orcid.org/0000-0003-3623-168X. The support of the National Science Centre in Poland is gratefully acknowledged (research project 2018/31/D/HS4/00316).

intermediate flow matrix denoted as $\mathbf{x}^{\mathbf{S} \times \mathbf{S}}$, the problem of generating an IRIO transaction matrix can be formulated as construction of the following $R \times R$ matrices for each sectoral pair (s, v) , $s, v \in 1, \dots, S$:

$$\hat{\mathbf{x}}^{s,v} \equiv \begin{bmatrix} \hat{x}_{1,1}^{s,v} & \hat{x}_{1,2}^{s,v} & \dots & \hat{x}_{1,R}^{s,v} \\ \hat{x}_{2,1}^{s,v} & \hat{x}_{2,2}^{s,v} & \dots & \hat{x}_{2,R}^{s,v} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{R,1}^{s,v} & \hat{x}_{R,2}^{s,v} & \dots & \hat{x}_{R,R}^{s,v} \end{bmatrix}, \quad (1)$$

in place of the known scalar element $x^{s,v}$ of $\mathbf{x}^{\mathbf{S} \times \mathbf{S}}$, so that $\sum_{r,p} \hat{x}_{r,p}^{s,v} = x^{s,v}$. Element $\hat{x}_{r,p}^{s,v}$ shall be understood as the use of products from supply-side sector s and region r in demand-side sector v and region p ($r, p \in 1, \dots, R$ – domestic regions). Such elements need to be estimated because they are not directly observable in subnational granularity as taking excessive effort to compile (unlike multinational sources like World Input-Output Database, Global Trade Analysis Project – MRIO, or OECD IOTs). Ever since Leontief and Strout (1963), the I–O literature has been looking for some form of regional tables. Most of these attempts have focused on single-region, or *intraregional* I–O tables, sized $S \times S$. In this strand of literature, probably the most widespread non-survey approach is the Location Quotient (LQ) method, used in multiple variations (see i.a. Flegg et al. 1995; Bonfiglio 2009; Flegg and Tohmo 2013b,a, 2016). An alternative approach is CHARM, developed by Kronenberg (2009). However, when simulating the local economic impact in the supply chain for a relatively small area (like NUTS-3), one has to bear in mind that single-region tables do not account for cross-regional feedback effects (Folmer and Nijkamp, 1985) and the case for using IRIO arises (Wiedmann et al., 2011; Miller and Blair, 2009, ch. 3, pp. 76-101).

The literature on building sub-national IRIO remains scarce. As one of the few examples, Jahn (2017) shall be named. He combines the FLQ framework with a trade gravity model estimated for the EU-28 countries, proposing a method named IRIO LQ, and therefore takes into account the geographic proximity of regions. Distance is the most intuitive criterion for IRIO connectivity supplementing the regional and sectoral supply and demand criteria, traditionally postulated in the literature (Round, 1978). This may be derived from transport costs, imperfect information, or, generally, Tobler’s (1970) paradigm of physical proximity as determinant in socio-economic analyses. Gravity model estimation (Leontief and Strout, 1963; Polenske, 1970; Gordon, 1976; Lindall et al., 2006) can solve this problem, but it is feasible only when some flow matrix is already available (see also Jackson et al., 2006).

According to Rey (2000) and Loveridge (2004), the development of spatial econometrics can particularly set a promising direction for the so-called integrated econometric–I–O models, and they appear to be well-equipped to address the role of distance in specific small-area problems (judging by the example small-area problems provided by Morrison and Smith, 1974). Torój (2016; 2021) demonstrates a convenient specification embedded in the econometric analysis of the Spatial Durbin Model (SDM). His framework endogenises the spatial weight matrix as build upon parametric spatial decay schemes. They are inversely related to the gamma cumulative distribution functions (CDF). In comparison to the location quotient family methods, the spatial approach has turned out to be promising in replicating

the key summary statistic of the scarce available regional I-O tables: the fraction of local sourcing as compared to total domestic sourcing (i.e. to the sum of supplies from the home region and other domestic regions). For the purpose of benchmarking, the Finnish regional tables were used (Statistics Finland, 2006). The same source was previously heavily exploited by Flegg and Tohmo (2013b; 2016) in their explorations of LQ properties.

Regardless of the construction method applied to obtain the initial version of IRIO, some of the recent literature focuses on IRIO table balancing procedures. They are mostly extensions to the RAS algorithm, proposed by Stone (1961) to ensure compliance of the table with column and row sum constraints, while preserving as much as possible similarity to the initial version of the table. Torój (2021) adds an intermediate step to each RAS iteration to ensure the compliance of block sums with national I–O table entries. Temursho et al. (2021) propose a more general algorithm, Multi Regional Generalized RAS method, that allows for adding constraints in other forms than on row and column sums, in multiple balancing, updating or regionalizing applications. Lamonica et al. (2017; 2020) look into performance evaluation of multiple methods, including constrained matrix-balancing methods (and cross-entropy methods in particular).

Although the previous works on the spatial econometric method of building IRIO shed some light on the general performance and the underlying econometrics, a few problems related to that approach still remain unresolved. First, the econometric analysis of gamma-CDF-related functions can be tricky and misleading when at least some of their parameters – shapes or scales – fall into the region where the CDF becomes insensitive, up to working precision, to their values (see Figure 2 in Subsection 3.1 for details). As an obvious consequence, the related log-likelihood function can be difficult to maximize, and the apparent maximum can be ill-behaved in Hessian, preventing the researcher from due assessment of the statistical uncertainty.

Second, and perhaps more importantly, the multivariate SDM presented in the previous literature can be regarded as a sort of spatial filter, scanning the map for spatial autocorrelation patterns to attribute them to the known I–O relationships. However, the spatial autocorrelation can also originate elsewhere, introducing additional noise into the analyzed data. To avoid the impact of this noise, one could i.a. impose additional structure on top of the previously considered spatial econometric specification by using it jointly with other data sources (such as microdata or data from other countries).

Those two “birds” can be killed by one stone by using the Bayesian methods, popular elsewhere in spatial econometrics (cf. LeSage and Pace, 2009; Haining and Li, 2020). On the one hand, one could solve the problem of spatial decay profile insensitivity to parameters by prior exclusion restrictions imposed on the subset of the parameter domain where this phenomenon occurs. On the other hand, one can exploit the scarce instances of published IRIO to derive prior knowledge on the feasible part of the domain. For the latter purpose, I revisit the well-known dataset compiled by Statistics Finland (2006) and build prior distributions for shape and scale in all 7 groups of NACE sections in Poland, where data on value added is available in NUTS-3 regional breakdown.

This is accomplished in Section 3, after the summary of the problem specification (Section 2) leading to a formulation of the likelihood function, here used as the sampling density. In Section 4, I present the

posterior distributions of all shape and scale parameters, as well as the spatial decay profiles, including the credible sets around them, derived by sampling from the posterior density. Section 5 presents an illustrative regional simulation results in Poland. I focus on the key parameter – the fraction of indirect effects falling on the region of origin of the impulse – and provide a related confidence interval, based on sampling from the posterior density. Section 7 concludes.

In the IRIO discussion, one must not forget that the ongoing boom in big data analytics, including spatial big data (cf. Yamagata and Seya, 2020) might facilitate future refocusing from non-survey to survey methods, like commodity flow surveys, or – more likely – some sensor-based or GPS-tracking version thereof. As a small step in that direction, in this research I also build on currently available spatial big datasets. First, to account for the distance between NUTS-3 regions in Section 4, unlike the previous literature, I do not use straight-line distances between centroids, but real-world road distances between the most populated cities of each region derived from Google Maps. Second, I compare this to another measure of economic distance between these points – driving time (also derived from Google Maps as an average over various points in the course of day and week). Third, in the illustrative simulation in Section 5, I compute the regional induced effects on the basis of geolocation data for mobile devices. This is, hopefully, a more realistic treatment than assuming that all the income earned in a given region is spent in the same region.

2 IRIO problem as multi-equation Durbin model

Following Torój (2021), I specify the problem of obtaining IRIO as corresponding to the econometric problem of estimating the multi-equation version of the SDM. The main underlying assumption is that the sectoral domestic cost structures do not vary by region and so the ratios of domestic costs, foreign costs and value added to output for a given v remain constant across p . On the above assumption, I decompose $x^{s,v}$ into the column sums of $\hat{\mathbf{x}}^{s,v}$ in proportion to the value added of the recipient sector–region pair (v, p) :

$$\sum_{r=1}^R \hat{x}_{r;p}^{s,v} \equiv \hat{z}_p^v \cdot x^{s,v}, \quad (2)$$

$$\hat{\mathbf{z}}^v \equiv \begin{bmatrix} \hat{z}_1^v & \dots & \hat{z}_R^v \end{bmatrix} \equiv \begin{bmatrix} \frac{va_1^v}{\sum_{p=1}^R va_p^v} & \dots & \frac{va_R^v}{\sum_{p=1}^R va_p^v} \end{bmatrix}. \quad (3)$$

where va_r^v denotes gross value added in sector v in region r . Column sums $\sum_r \hat{x}_{r;p}^{s,v}$ need to be allocated to the supply-side regions r (rows). This allocation is described by the vertical non-negative vector, $\hat{\mathbf{w}}_p^s = \left[w_{1,p}^s \quad w_{2,p}^s \quad \dots \quad w_{R,p}^s \right]^T$, specific for each pair (s, p) and satisfying $\forall_{s,p} \sum_{r=1}^R \hat{w}_{r,p}^s = 1$. These vectors form columnwise $R \times R$ -sized matrices $\hat{\mathbf{W}}^s \equiv \left[\hat{\mathbf{w}}_1^s \quad \hat{\mathbf{w}}_2^s \quad \dots \quad \hat{\mathbf{w}}_R^s \right]$. In line with Tobler’s law, for a given pair (s, v) , elements $\hat{w}_{r,p}^s$ shall increase as the regions (r, p) become closer to one another.

The elements of $\hat{\mathbf{W}}^s$ can be found through an econometric analysis of the equations:

$$\begin{bmatrix} va_1^s \\ \vdots \\ va_R^s \end{bmatrix} = \begin{bmatrix} \hat{w}_{1,1}^s & \dots & \hat{w}_{1,R}^s \\ \vdots & \ddots & \vdots \\ \hat{w}_{R,1}^s & \dots & \hat{w}_{R,R}^s \end{bmatrix} \begin{bmatrix} va_1^s \\ \vdots \\ va_R^s \end{bmatrix} \beta_s^s + \begin{bmatrix} \hat{w}_{1,1}^s & \dots & \hat{w}_{1,R}^s \\ \vdots & \ddots & \vdots \\ \hat{w}_{R,1}^s & \dots & \hat{w}_{R,R}^s \end{bmatrix} \begin{bmatrix} va_1^{v_1} & \dots & va_1^{v_{S-1}} & \hat{y}_1^s \\ \vdots & \ddots & \vdots & \vdots \\ va_R^{v_1} & \dots & va_R^{v_{S-1}} & \hat{y}_R^s \end{bmatrix} \begin{bmatrix} \beta_{v_1}^s \\ \vdots \\ \beta_{v_{S-1}}^s \\ \beta_0^s \end{bmatrix} + \begin{bmatrix} \varepsilon_1^s \\ \vdots \\ \varepsilon_R^s \end{bmatrix}, \quad (4)$$

where $\beta_v^s \equiv \frac{va^s}{X^s} a^{s,v} \frac{X^v}{va^v}$ and $\beta_0^s \equiv \frac{va^s}{X^s}$ denote known coefficients, calibrated on the basis of country-wide input-output data ($a^{s,v}$ – cost structure matrix element at (s, v) , X^s – global output in sector s), $v_1, \dots, v_{S-1} \neq s$, \hat{y}_r^s – data on final output in sector s and region r and $\hat{w}_{r,p}^s$ is estimated for all r, p, s (see Torój (2021; 2016) for more discussion). Note that this is a counterpart to a single-equation Durbin model, except the fact that the causal coefficients are known, while the spatial weights – estimated.

Collecting equations for each $s = 1, \dots, S$ and defining $\beta \equiv \begin{bmatrix} \beta_1^1 & \dots & \beta_S^1 \\ \vdots & \ddots & \vdots \\ \beta_1^S & \dots & \beta_S^S \end{bmatrix}$ and $\beta_0 \equiv \begin{bmatrix} \beta_{0,1} \\ \vdots \\ \beta_{0,S} \end{bmatrix}$

yields:

$$\underbrace{\left(\mathbf{I}_{S \cdot R} - \beta \otimes \begin{bmatrix} \hat{\mathbf{W}}^1 & \dots & \hat{\mathbf{W}}^1 \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{W}}^S & \dots & \hat{\mathbf{W}}^S \end{bmatrix} \right)}_{\equiv \hat{\mathbf{A}} = \mathbf{A}[\beta, \mathbf{W}^1(\hat{\theta}^1), \dots, \mathbf{W}^S(\hat{\theta}^S)]} \mathbf{va} = \underbrace{\beta_0 \otimes \begin{bmatrix} \hat{\mathbf{W}}^1 \\ \vdots \\ \hat{\mathbf{W}}^S \end{bmatrix}}_{\equiv \hat{\mathbf{B}} = \mathbf{B}[\beta_0, \mathbf{W}^1(\hat{\theta}^1), \dots, \mathbf{W}^S(\theta^S)]} \mathbf{y} + \boldsymbol{\varepsilon}, \quad (5)$$

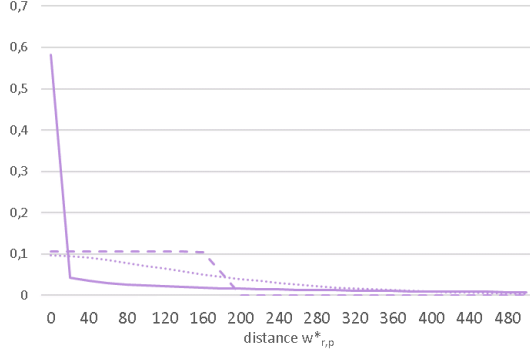
with $\mathbf{va} = \begin{bmatrix} \mathbf{va}^1 \\ \vdots \\ \mathbf{va}^S \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} \mathbf{y}^1 \\ \vdots \\ \mathbf{y}^S \end{bmatrix}$ and $\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}^1 \\ \vdots \\ \boldsymbol{\varepsilon}^S \end{bmatrix}$. The last one is the stochastic error term

distributed as multivariate normal with zero mean and a certain structure of its variance-covariance matrix $\boldsymbol{\Sigma}$. For each sector s , the variance of the error is homogeneous and equals σ_s^2 . Errors can be correlated across sectors s, v in the same region with covariance $\hat{\sigma}_{s,v}$. It is therefore useful to define the following symmetric, semi-positive-definite $S \times S$ matrix of cross-sectoral error correlations in the

same region: $\boldsymbol{\Omega} \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,S} \\ \cdot & \sigma_2^2 & \dots & \sigma_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \dots & \sigma_S^2 \end{bmatrix}$, and further define $\boldsymbol{\Sigma} = \boldsymbol{\Omega} \otimes \mathbf{I}_R$.

Obviously, since there are too many elements in the matrix $\hat{\mathbf{W}}^s$, their unconstrained estimation is not feasible. As demonstrated by Torój (2021), the use of gamma cumulative distribution function $P_\Gamma(\cdot)$

Figure 1: Three illustrative spatial decay profiles ($\hat{w}_{r,p}^s$) for regional supply sourcing



Source: author.

parametrised for each sector s with shape $\hat{\theta}_1^s > 0$ and scale $\hat{\theta}_2^s > 0$ to define

$$\hat{w}_{r,p}^s \equiv \frac{1 - P_{\Gamma}(\theta_1^s, \theta_2^s, w_{r,p}^*)}{\sum_{r=1}^R [1 - P_{\Gamma}(\theta_1^s, \theta_2^s, w_{r,p}^*)]} \quad (6)$$

can be considered both flexible and parsimonious in building spatial weights representing Tobler-like pattern of spatial decay across supply-side sectors (commodities) S . For example, it can be fitted to three different situations: (i) local suppliers are strongly preferred (see solid line in Fig. 1), (ii) a threshold of tolerance exists (dashed line in Fig. 1), and (iii) the supply intensity decreases mildly with distance (dotted line in Fig. 1).

Collecting the unknown parameters into the vector $\boldsymbol{\theta}^T = \left[\boldsymbol{\theta}^{1T} \quad \dots \quad \boldsymbol{\theta}^{ST} \quad \text{vec}(\boldsymbol{\Omega}) \right]$ (with $\boldsymbol{\theta}^{sT} = \left[\theta_1^s \quad \theta_2^s \right]^T$), the econometric problem can be represented with the following likelihood function:

$$\begin{aligned} \ln L(\boldsymbol{\theta} | \boldsymbol{\beta}, \boldsymbol{\beta}_0, \mathbf{W}^*, \mathbf{va}, \mathbf{y}) &= \\ &= -\frac{S \cdot R}{2} \ln(2\pi) + \ln |\mathbf{A}(\boldsymbol{\theta})| - \frac{R}{2} \ln |\boldsymbol{\Omega}(\boldsymbol{\theta})| - \frac{1}{2} \boldsymbol{\varepsilon}(\boldsymbol{\theta})^T \left(\boldsymbol{\Omega}(\boldsymbol{\theta})^{-1} \otimes \mathbf{I}_{\mathbf{R}} \right) \boldsymbol{\varepsilon}(\boldsymbol{\theta}). \end{aligned} \quad (7)$$

The econometric analysis leads to estimated matrices $\hat{\mathbf{x}}^{s \cdot \mathbf{v}}$ for every s (each of them is the same for all v). Their complete set forms the matrix $\hat{\mathbf{x}}$ sized $(S \cdot R) \times (S \cdot R)$, further subject to an extended RAS balancing procedure, ensuring the simultaneous fulfillment of 3 conditions: (i) on the row sums (which is not fulfilled in the initial version, and the balancing can be viewed as scaling up or down to a regional supply-side potential), (ii) on the column sums (fulfilled in the initial version) and (iii) on all $R \times R$ block sums (fulfilled in the initial version). The balanced matrix $\tilde{\hat{\mathbf{x}}}$ is further transformed into cost structure matrix $\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} = \tilde{\hat{\mathbf{x}}} \otimes \left[\mathbf{1}^{(S \cdot R) \times 1} \otimes \mathbf{X}^{1 \times (S \cdot R)} \right], \quad (8)$$

where $\mathbf{X}^{1 \times (\mathbf{S} \cdot \mathbf{R})}$ is a vector of global output across the defined sectors and regions, estimated under the assumption of constant value-added-to-global-output ratios across regions for a given sector.

3 Prior elicitation

The elicitation of the prior distribution of θ is based on two criteria:

- attributing zero prior density to the subsets of the domain where the log-likelihood is indifferent to shape and scale changes, due to the asymptotic behaviour of the gamma CDF in shape and scale;
- using the Statistics Finland (2006) regional I–O tables to indicate prior differences in the supply geography of individual commodities.

3.1 Domain restriction

The distances between Polish NUTS-3 regions range from 0 to approximately 800 km. Figure 2 depicts the spatial decay profiles, corresponding to Figure 1, for specific shape and scale values. Notice the existence of 3 regions in this 2-dimensional domain, defined for each individual commodity, where the profile becomes virtually indifferent to changes in parameter values:

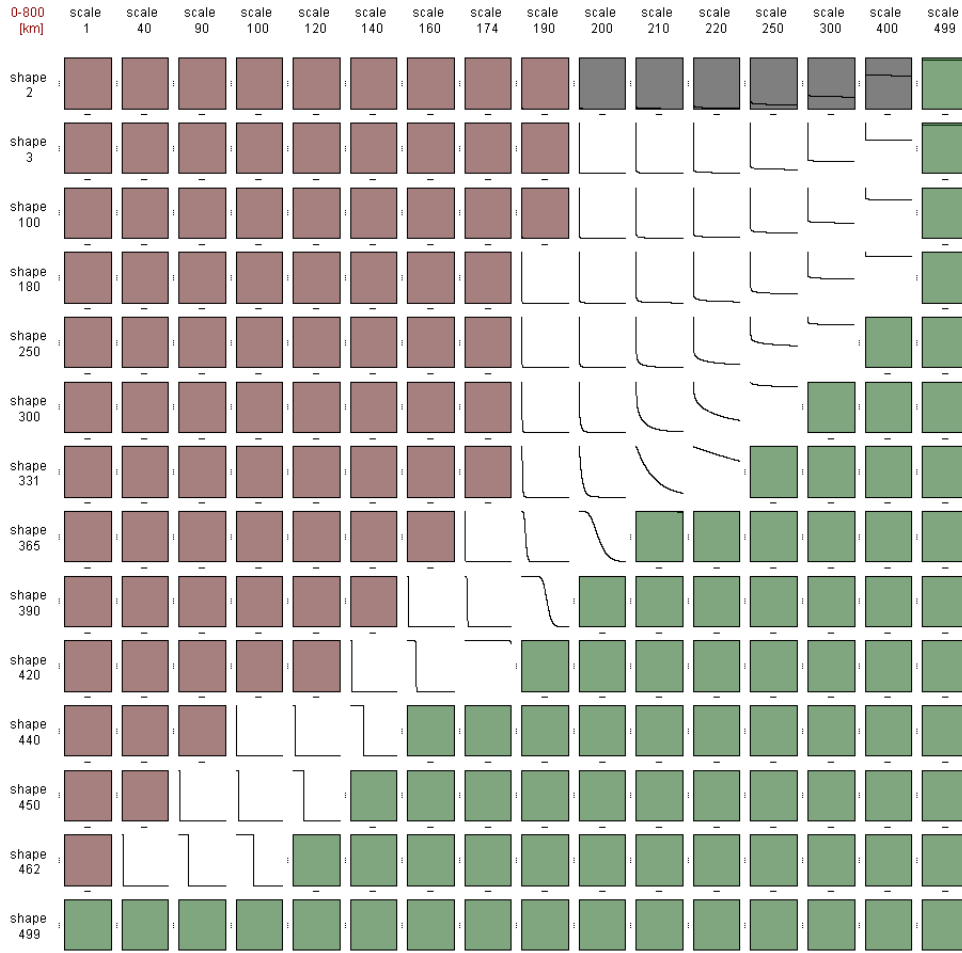
1. left-hand area (red) below a given scale with L-shaped profile, under which all supply is sourced in the same region;
2. right-hand area (green) above a given scale with flat profile, under which the distance does not matter in sourcing the supply;
3. top area (gray) in which the profile changes horizontally (L-shaped with a horizontal ramp on different levels), but exhibits no more material changes when moving towards the top.

Following a high-precision grid search, a number of points have been found at which the change in shape or scale in the above-mentioned directions is no more noticeable in terms of the spatial decay profile, given the software’s numerical precision. This implied a maximum and minimum value of the shape: $\bar{\theta}_{max}$ and $\bar{\theta}_{min}$ (for each s). Polynomials in shape of sufficiently high order (i.e. until the next added term was no more significant) were fitted to the sequences of limiting points on both sides. This yield the upper and lower border of the coloured area in Figure 3, as polynomials in shape of order 12 and 5, respectively, denoted as $T_{max}(\theta_1^s)$ and $T_{min}(\theta_1^s)$ (again, the same for each s). The colours in Figure suggest that, for a given shape, the increase in scale implies sourcing the supply of a given commodity from more distant locations.

Based on the observations above, the prior distributions for each scale parameter is set independently as uniform:

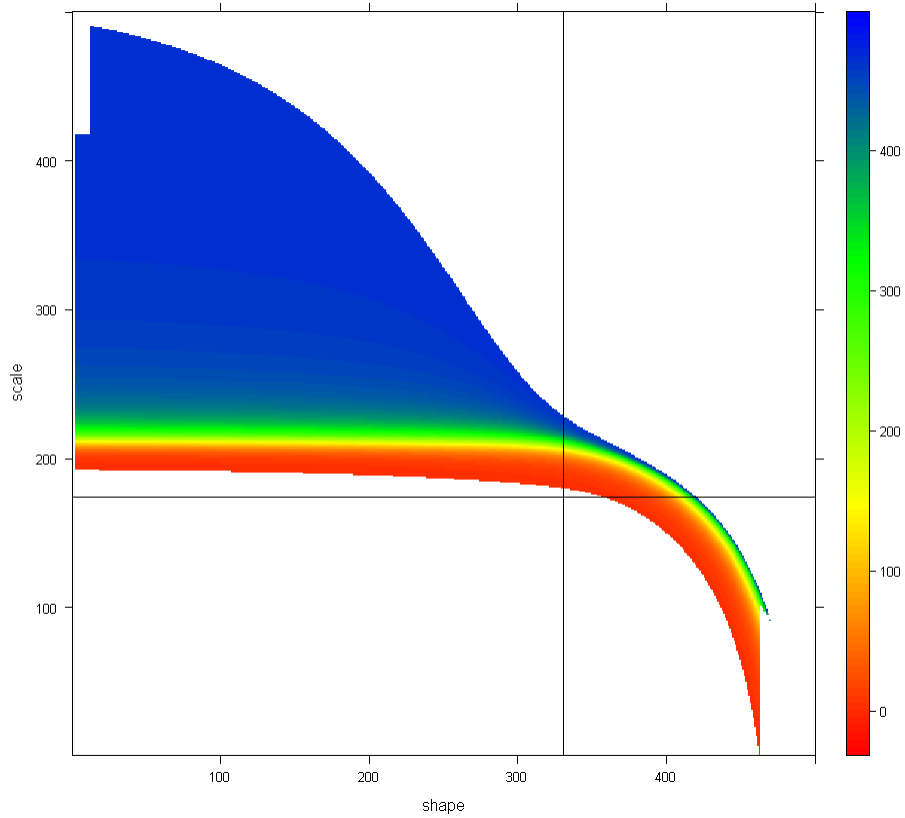
$$\theta_1^s \sim U(\bar{\theta}_1^{min}, \bar{\theta}_1^{max}), \quad s = 1, \dots, S \text{ i.i.d.} \quad (9)$$

Figure 2: Spatial decay profiles for different shape and scale values



Source: author.

Figure 3: Area of non-zero prior shape and scale density for every sector



The colours indicate the weighted average distance under a spatial decay profile with a given shape and scale.

Source: author.

3.2 Statistics Finland regional I–O data

The prior densities of scales are, in turn, set differently for each s , to reflect sectoral specificity, and conditionally upon the shape θ_1^s , to reflect the preference for some feasible supply geographies. This elicitation builds on the data published by Statistics Finland (2006) – one of the few countries in the world where the statistical office published regional I–O tables. The Finnish tables have been derived from various sources, including commodity flow surveys, and reflect the flows as of 2002. Like in the analysed case of Poland, the regional breakdown of these tables corresponds to the NUTS-3 statistical division (defined, as a rule of thumb, as counting 150000-800000 inhabitants and, in the case of Finland, corresponding to the administrative division into 19 regions). The sectoral breakdown involves 27 sections, aggregated in this study to 7 groups of sections to correspond to the sectoral granularity of the Polish data on gross value added for NUTS-3 regions.

Note, however, that this set of tables is not an instance of IRIO, but intraregional tables. Therefore, one cannot infer directly the quantities desired in our model, and hence I propose the following indirect scheme. For each region ($r = 1, \dots, 19$), the separate table contains the row of imports split into two: foreign and domestic (from other regions in Finland). The final demand column, however, contains an aggregate of both intermediate exports to other domestic regions and the regional contributions to the national final demand (see Table 1a). According to the assumption about supply-sector-specific (or commodity-specific) $\hat{\mathbf{W}}^s$ matrices, my purpose is to evaluate the fraction of sector's s intermediate supply used in the home region (in relation to its intermediate supply used in all regions). To accomplish this, I proceed in the following steps:

1. Disaggregation of domestic imports row, xm_r^s , into sectors. To this aim, I:
 - (a) compute national cost structures in Finland in the same year: $\mathbf{A} = \mathbf{x} \oslash (\mathbf{1}^{S \times 1} \otimes \mathbf{X}^{1 \times S})$, where \mathbf{X} – vector of global output in Finland, \mathbf{x} – intermediate transaction matrix in Finland.
 - (b) approximate the total domestic intermediate demand matrix (i.e. intraregional transactions plus domestic imports) of region r , using nation-wide cost structures and global output in region r : $(\mathbf{x}_r + \hat{\mathbf{x}}\mathbf{m}_r) = \mathbf{A} \oslash (\mathbf{1}^{S \times 1} \otimes \mathbf{X}^r)$.
 - (c) approximate the initial version of region's r $S \times S$ domestic intermediate imports matrix (cf. Subtable 1b), $\mathbf{x}\tilde{\mathbf{m}}_r$, as a pairwise maximum across the respective elements of matrices $\left\{ (\mathbf{x}_r + \hat{\mathbf{x}}\mathbf{m}_r) - \mathbf{x}_r; \mathbf{0}^{S \times S} \right\}$ (to avoid negative values).
 - (d) Rescale columnwise $\mathbf{x}\tilde{\mathbf{m}}_r$ so that column sums match xm_r^A, xm_r^B etc. from Subtable 1a, obtaining $\mathbf{x}\hat{\mathbf{m}}_r = [xm_r^{s,v}]$ from Table 1b.
2. Approximation of mean domestic exports with mean domestic imports. Note that the total intermediate domestic exports of commodity s , $\sum_r IDE_r^s$, equals the total intermediate domestic imports of the same commodity s , and so does the regional mean (see Subtable 1c):

$$I\bar{D}E^s = \frac{1}{R} \sum_r IDE_r^s = \frac{1}{R} \sum_{v,r} xm_r^{s,v}. \quad (10)$$

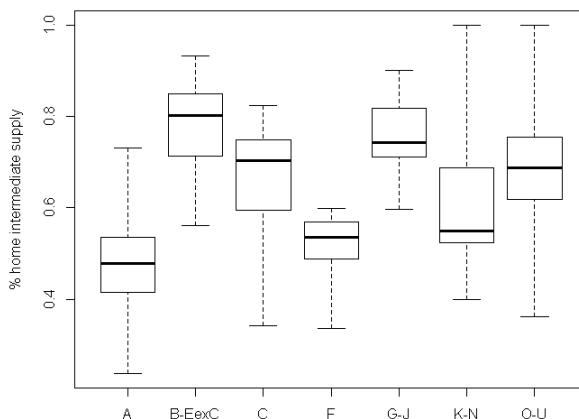
Ultimately, I treat the fraction $\phi^v = \frac{\sum_s \bar{x}^{v,s}}{\sum_s \bar{x}^{v,s} + IDE^v}$ (cf. Subtable 1c) as the share of within-region intermediate supply in total intermediate supply of commodity v . These means are presented in Table 2 and in Figure 4 as the boxplot's central lines. On top of that, I compute the standard deviation (across r) of the terms $\frac{\sum_s x_r^{v,s}}{\sum_s x_r^{v,s} + \sum_s xm_r^{s,v}}$, ϕ_{SD}^v , as a rough proxy of the cross-regional differences, for the purpose of further prior elicitation (see the last column in Table 2 and the interquartile and maximum-to-minimum ranges in Figure 4).

Table 2: Share of home region as a destination of intermediate supply

NACE sections	intermediate supply to home region	
	mean [%, across regions]	S.D.[p.p., across regions]
A	47.1	11.3
BDE	77.5	10.8
C	66.0	12.7
F	52.0	6.8
G-J	75.1	8.2
K-N	61.9	15.2
O-U	68.8	12.8

Source: author.

Figure 4: Share of home region as a destination of intermediate supply



Source: author.

To translate the information from Table 2 into the prior distribution over the coloured set in Figure 3, one needs to make a statement about the distance. For this purpose, all the distances from the regional capitals to the region's border, measured on the straight lines to all the capitals of the neighbouring

Table 1: Use of data from Statistics Finland (2006) regional tables: illustrative scheme

(a) Original structure of regional I-O tables in Finland

		home region r			final demand	\sum supply
		sector A	sector B	...		
home region r	sector A	$x_r^{A,A}$	$x_r^{A,B}$			X_r^A
	sector B	$x_r^{B,A}$	$x_r^{B,B}$			X_r^B
	...					
imports from other regions	all sectors	xm_r^A	xm_r^B			
imports from abroad	all sectors					

(b) Step 1: sectoral disaggregation of domestic imports

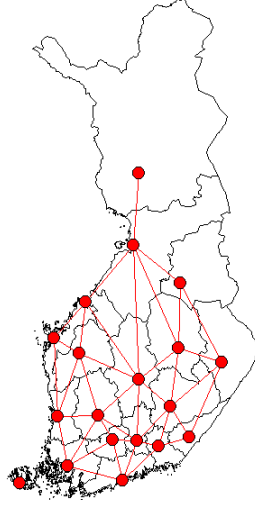
		home region r			final demand	\sum supply
		sector A	sector B	...		
home region r	sector A	$x_r^{A,A}$	$x_r^{A,B}$			X_r^A
	sector B	$x_r^{B,A}$	$x_r^{B,B}$			X_r^B
	...					
imports from other regions	sector A	$xm_r^{A,A}$	$xm_r^{A,B}$			
	sector B	$xm_r^{B,A}$	$xm_r^{B,B}$			
	...					

(c) Step 2: approximation of domestic exports with domestic imports for regional means

	average across regions		final demand	\sum (supply)
	own region's intermediate demand	other regions' intermediate demand		
sector A	$\sum_s \bar{x}^{A,s}$	IDE^A		X_r^A
sector B	$\sum_s \bar{x}^{B,s}$	IDE^B		X_r^B
...				
imports from other regions		

Source: author.

Figure 5: Interpretation of within-region supplies in Finland in terms of distance



Source: author.

regions (see Figure 5), were averaged. On that basis, within-region supplies are interpreted as up to 51.2 km.

Returning to Figure 3, for a given shape θ_1^s , the increase in scale makes the weighted average distance in the spatial decay profile grow. For the lowest feasible scale, the share of weights attributed to the distance below 51.2 km is near 100%. When the scale increases, this share is falling until it matches first $\phi^s + \phi_{SD}^s$, and then ϕ^s . Denote the latter value of scale as $\mu^s(\theta_1^s)$, and the difference between the former and the latter as $\sigma^s(\theta_1^s)$. Now I can elicit the prior for scale, independently for each sector, as truncated normal with the mean $\mu^s(\theta_1^s)$, the standard deviation $\sigma^s(\theta_1^s)$, the lower bound $T_{min}(\theta_1^s)$ and the upper bound $T_{max}(\theta_1^s)$:

$$\theta_2^s | \theta_1^s \sim TN[\mu^s(\theta_1^s), \sigma^s(\theta_1^s), T_{min}(\theta_1^s), T_{max}(\theta_1^s)]. \quad (11)$$

I complete the prior elicitation with a fairly standard choice of the inverted Wishart prior for the variance-covariance matrix of the residuals across 7 sectors, $\mathbf{\Omega}$:

$$\mathbf{\Omega} \sim invWishart(v, \mathbf{H}), \quad (12)$$

where $v = 7$ (the lowest possible number of degrees of freedom), diagonal elements $[H_{s,s}]$ equal to 40% of variance of va_r^s across r (to be tested for sensitivity), and the non-diagonal elements of \mathbf{H} set to imply prior correlations of 0.3, given the diagonal elements and v .

4 Posterior distributions and spatial decay profiles

4.1 Data

In the empirical analysis for Poland, the value added (**va**) data source is the Local Data Bank of Statistics Poland. **va** data are broken down into NUTS-3 regions (73 for Poland) and sectors (7 NACE 2.0 section groups). The data on final demand are available from I–O tables only in sectoral breakdown, but not in spatial breakdown. As of 2015 in Poland, the consumption (both government and private) accounted for more than half of the final demand, so – as a proxy – I interpolate the sectoral final demand proportionately to the local populations. The calibration of β and β_0 is based on the input-output table for Poland as of 2015, compiled by Statistics Poland. It is straightforward to extend the econometric problem to a multi-year, pooled panel specification. Since statistical offices publish I–O tables in 5-year intervals, I use annual data for the 5-year period centred around 2015, with constant parameters.

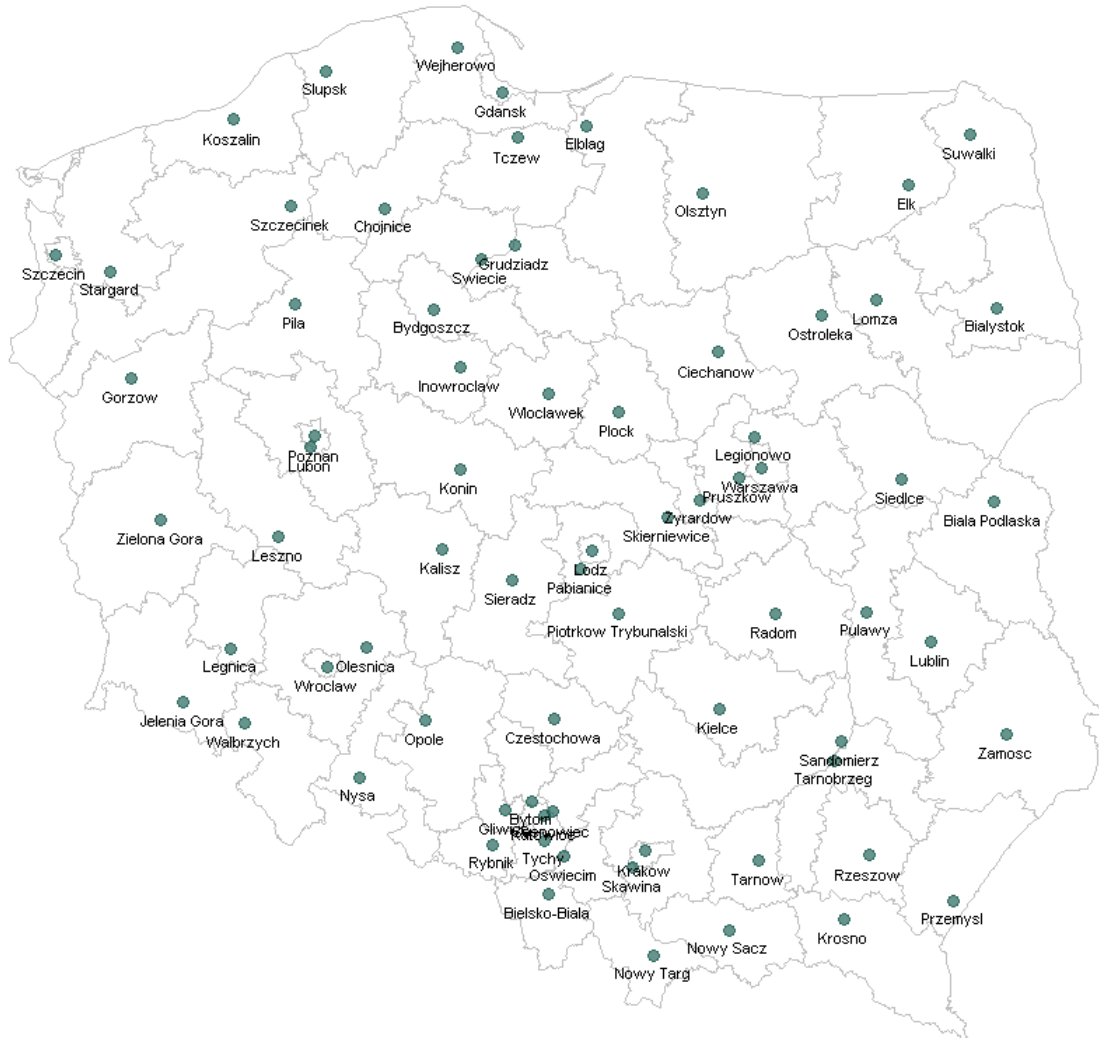
As opposed to the previous related literature, I abandon the simplifying about \mathbf{W}^* that the distance is measured on the straight line between centroids. Instead, I consider the real-world driving distance between the most populated cities in each pair of regions. The obvious consequence thereof is that \mathbf{W}^* exhibits some minor deviations from symmetry. To understand further implications, one needs to take additional insights into the map of Polish NUTS-3 regions. First, Figure 6 demonstrates that the most populated cities – treated as economic centers of gravity – are often located on the periphery when one considers the region’s geometric shape (see Szczecinek or Suwałki). Second, a number of “ring” regions around big metropolies do not even encompass their centroids. This may produce a false impression of negligibly low distance between the city and its ring in an agglomeration. Figure 12a provides an example, and additionally demonstrates how illusive the straight-line distance measurement can be.

4.2 Parameter distributions and convergence diagnostics

The posterior distribution has been simulated with the random walk Metropolis-Hastings algorithm in 2 chains, each of them of 3 million draws. The first 1.5 million draws were rejected as burn-in from the initial values set at prior means. Two features of the simulating process should be noted. First, given the specific shape of the domain (see Figures 2 and 3), I prefer to work with $\ln(\theta_1^s)$ and $\ln(\theta_2^s)$ rather than θ_1^s and θ_2^s in a procedure where the step is drawn with a constant variance and the responsiveness of distance profiles to the level of shape and scale is (graphically) very different in various parts of the domain. Second, as a candidate generating density, I use multivariate normal density with zero covariances and parameter-specific variances: $5 \cdot 10^{-5}$ for θ_1^s and $25 \cdot 10^{-5}$ for θ_2^s , $25 \cdot 10^{-5}$ for error variances σ_s^2 and $2 \cdot 10^{-5}$ for cross-sectoral error correlations $\rho_{s,v}$ (or, more precisely, for their transformations α_s such that $\rho_{s,v} \equiv 2 \cdot \Phi(\alpha_s) - 1$). The mean acceptance probabilities in both chains amounted to 39% and 41%.

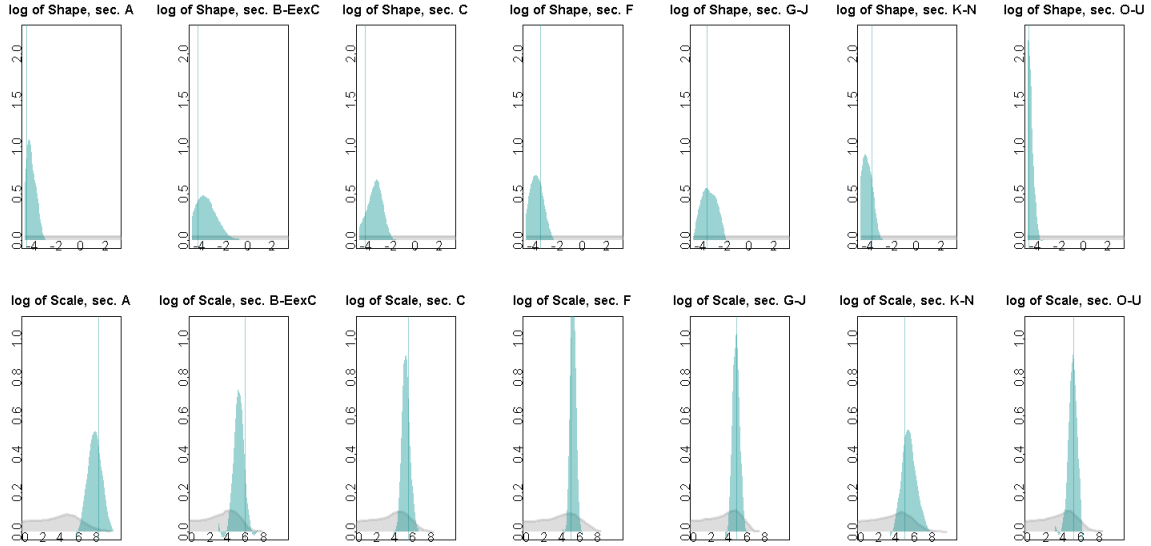
Figure 7 presents the posterior densities (in green) against their prior counterparts (in grey, presented incompletely). The variance appears to have shrunk in a satisfactory way, and the posterior densities

Figure 6: Most populated cities of NUTS-3 regions



Source: author.

Figure 7: Shape and scale: prior (gray) and posterior (green) distributions



Source: author.

are regular and unimodal. The convergence is confirmed by the potential scale reduction factors below 1.2 for all the parameters.

4.3 Spatial decay profiles

The values reported in Table 3 are difficult to interpret by a mere look at the posterior mean. Nor can be figured out the amount of uncertainty implied by the respective 95% highest posterior density intervals. Instead, the economic reading of these estimates consists in plotting the spatial decay profiles implied by every shape and scale pair, and evaluating the amount of uncertainty around that profile.

Figure 8 presents the supply geography profile for each commodity. The black lines demonstrate baseline profiles, corresponding to the posterior means of each production sector's shape and scale. Elements $\hat{w}_{r,p}^s$ in the initial version of $\hat{\mathbf{x}}$ (before balancing), proportional to the preferred locations from which a firm located in region p orders commodity s , depend on the distance from r to p demonstrated on individual panes of this figure, for each s separately.

The computation of the confidence sets around the profiles builds on the 1.5m valid draws from the posterior distribution. Bearing in mind Figure 3, one can expect a given average supply distance to fit various combinations of θ_1^s and θ_2^s . As a result, for the same s , these two dimensions of each chain are strongly correlated, and the evaluation of θ_1^s marginal dispersion is economically meaningless without θ_2^s (and *vice versa*). Instead, I start with a separate, individual profile that every draw implies. This means that for each distance: 0, 1, ..., 800 km, 1.5 million values are generated. From each of these

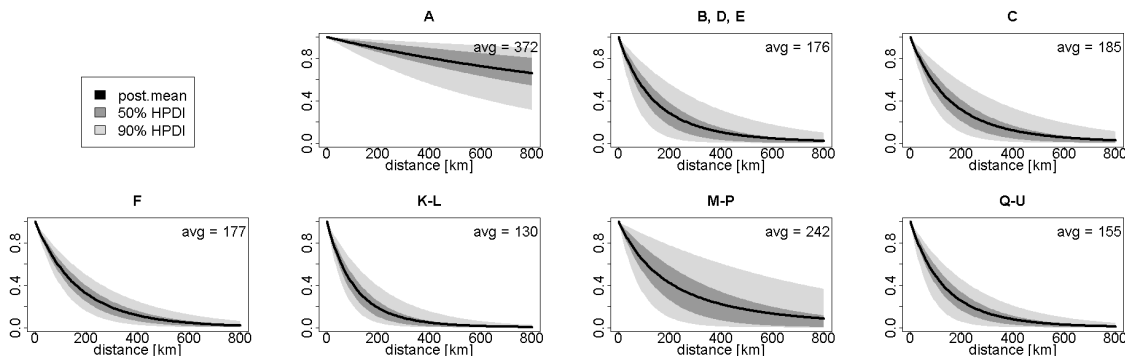
Table 3: Posterior distribution and convergence description

sector	shape				scale			
	post.mean	95% HPDI		R	post.mean	95% HPDI		R
A	0.018	0.010	0.043	1.02	2197.329	583.081	9227.817	1.01
B-E ex. C	0.040	0.011	0.226	1.17	151.985	55.870	406.063	1.12
C	0.039	0.012	0.125	1.02	170.120	76.896	428.969	1.02
F	0.023	0.010	0.061	1.01	175.197	98.720	313.323	1.02
G-J	0.033	0.011	0.100	1.01	114.063	58.227	243.523	1.00
K-N	0.016	0.010	0.035	1.09	236.287	68.046	976.590	1.02
O-U	0.013	0.010	0.021	1.02	139.397	58.915	283.047	1.01

R – potential scale reduction factor. R values for variances and covariances: from 1.00 to 1.02.

Source: author.

Figure 8: Estimated spatial decay profiles for Poland across sectors



The decay in weights on the vertical axes corresponds to the numerator in eq. (6), i.e. $1 - \Gamma(\hat{\theta}_1^s, \hat{\theta}_2^s, \bar{w}_{r,p})$, given the point estimates $\hat{\theta}_1^s, \hat{\theta}_2^s$ for sector s and the different distances between r and p , $w_{r,p}^*$, on the horizontal axis.

Source: author.

vectors, I read the quantiles of order 0.025 and 0.975 (0.05 and 0.95) to obtain the 95% (90%) confidence set, represented by the light- (dark-) grey areas in Figure 8. Such distance-quantile pairs form the vertices of these coloured areas.

Figure 8 demonstrates that sections A (agriculture, fishery, forestry) and M-P (advanced, professional services) sell value added to the most distant locations. The mean distance for a distance profile, computed as an average of the vector $0, 1, \dots, 800$ km weighted with the profile for posterior means, amounts to 363 km for A and 256 km for M-P. Sections K-L (simple services) exhibit an opposite behaviour, that is – they are mostly delivered locally with an average of 118 km.

5 Illustrative simulation results

I consider the direct, indirect and induced effects in the supply chain of a generic enterprise in the 'programming' sector located in the city Wrocław (NUTS-3 region PL514, one of $R = 73$ regions in Poland, henceforth referred to as "home region") (see Figure 9). The enterprise has a cost level of 100 m EUR and sales revenue of 200 m EUR.

The final form of the cost structure matrix used for simulating is additionally expanded by the household block (see block column and block row 2 in eq. 13) and the analysed enterprise block (block column and block row 3):

$$\hat{\mathbf{A}}^{[\mathbf{R}(\mathbf{S}+1)+1] \times [\mathbf{R}(\mathbf{S}+1)+1]} = \left[\begin{array}{c|c|c} \hat{\mathbf{A}}^{\mathbf{RS} \times \mathbf{RS}} & (1-g) \cdot \begin{bmatrix} c^1 \cdot \hat{\mathbf{W}}^1 \\ \vdots \\ c^S \cdot \hat{\mathbf{W}}^S \end{bmatrix} & \mathbf{a}_p^v \\ \hline [\mathbf{L} \cdot (1-\tau) \otimes \mathbf{X}] \otimes \mathbf{C} & \mathbf{0}^{\mathbf{S} \times \mathbf{S}} & \mathbf{1}_p^v \\ \hline \mathbf{0}^{1 \times \mathbf{RS}} & \mathbf{0}^{1 \times \mathbf{S}} & 0 \end{array} \right], \quad (13)$$

where $\mathbf{L} = [L^1 \dots L^S]$ is the horizontal vector of labour costs across $S = 77$ sectors (source: Statistics Poland), τ – country specific labour tax wedge (fraction, source: OECD), g – country-specific households' savings rate (fraction, source: Eurostat), c^s is the country-wide share of commodity s in total household consumption and $\hat{\mathbf{W}}^s = \hat{\mathbf{W}}^s$ for all s except retail trade, in which case $\hat{\mathbf{W}}^s = \mathbf{I}_R$. This structure of $\hat{\mathbf{W}}^s$ assumes that the regional distribution of demand for a given sector's products is equivalent to the estimated structure in the business-to-business relationships (with the only exception of the retail trade, where the only output are trade margins and those are assumed to be entirely local).

Matrix \mathbf{C} is related to the cross-regional commuting. I assume that households can spend a non-negligible portion of their income (calibrated discretionally at 50%) in a different region than the region of employment of the working household members. For this purpose, the scheme of commuting $\tilde{\mathbf{C}}$ needs to be established and $\mathbf{C} = 0.5 \cdot \tilde{\mathbf{C}} + 0.5 \cdot \mathbf{I}$. In $\tilde{\mathbf{C}}$ (and \mathbf{C} , both sized $R \times R$), columns correspond to the region of employment and sum to unity, thus indicating the structure of employees' residence arranged in rows.

I compute $\tilde{\mathbf{C}}$ using the unique dataset provided by Datarino (a technological company based in Poland, collecting geolocation data from mobile devices, mostly smartphones¹). The data on a device's anonymous advertising ID and geographic coordinates was captured over a 14-day period in 2019. A device was sending the data through a set of collaborating applications, whenever it fell within a range of a wi-fi network. It did not rely on the device's GPS positioning, but Google's service on wi-fi network positioning. Altogether, the dataset involved over 4.9 million records, related to over 355 thousand unique advertising IDs, derived from the complete database: (i) observations collected during 5-minute intervals starting at 3:00, 11:00, 17:00 and 23:00 (23%) and (ii) all observations for

¹<https://ceo.com.pl/powstala-platforma-do-analizy-i-gromadzenia-informacji-nt-zachowan-klientow-w-sklepie-dzieki-niej-sprzedawcy-moga-p>

randomly selected advertising IDs collected in the course of the 14-day sampling period (77%). Out of that set, the following records have been filtered out:

1. records of advertising IDs observed only once, i.e. with no tracking value (over 4.8m records for 254 142 IDs left);
2. records located outside Poland (over 4.8m records for 253 340 IDs left);
3. records observed on weekends, and on working days from 14:00 to 23:00 and from 6:00 to 10:00, difficult to interpret as home-time or work-time observations (871 508 records for 158 235 IDs left);
4. advertising IDs with diverse worktime locations, i.e. the dominant location representing less than 60% of cases or observed only once (84 517 IDs left);
5. advertising IDs with diverse hometime locations (conditions as above), and – further – for which either hometime or worktime location remained unidentified (46 249 IDs left).

The contingency table for the homeplace and workplace frequencies of the remaining 46 249 IDs, with normalized column sums, is reported in Appendix A as matrix $\tilde{\mathbf{C}}$. This table should be treated with due caution, since material discrepancies between observation counts in individual columns and rows occur, and hence some cells are underpopulated (see Tables 5-7). Further, while the sampling scheme from the full database was random, there is no reason to assume that the full database coverage is random with respect to the entire population. However, 89% of the observations are located on the diagonal of this contingency matrix, which suggests that the use of \mathbf{C} can be treated as a minor refinement as compared to the alternative of using an identity matrix instead, which would assume away any cross-regional commuting.

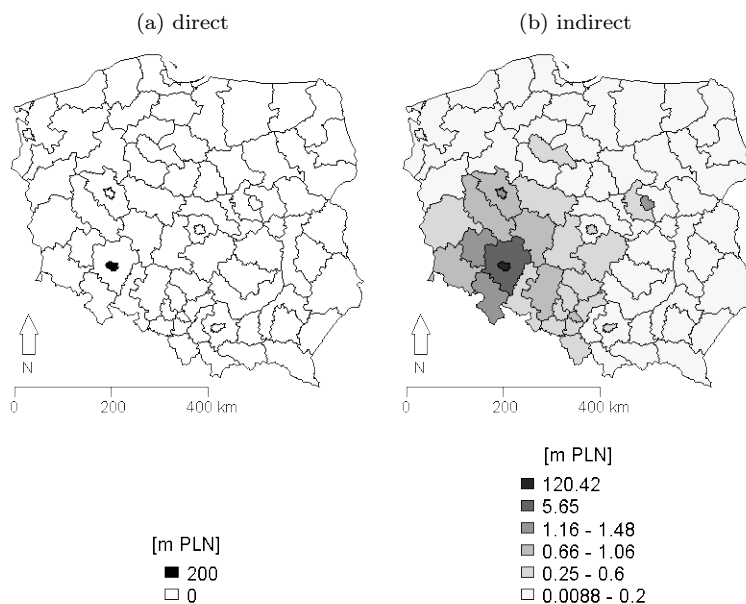
The last column describes the company’s cost structure, and in practical situations should rely on the company’s purchasing data (across both sectors and domestic regions) divided by total sales revenue; for the generic company, I set it equal to \mathbf{a}_p^v , that is the column of $\hat{\mathbf{A}}^{\mathbf{RS} \times \mathbf{RS}}$ representing sector v and region p to which this company belongs. In line with the above, the vertical vector \mathbf{l}_p^v is equal to the column of $[\mathbf{L} \cdot (1 - \tau) \otimes \mathbf{X}] \otimes \mathbf{C}$ – block (2,1) on the right-hand side of 13 – corresponding to sector v and region p .

The standard simulation formula reads as follows:

$$\Delta \mathbf{X}^{\mathbf{R}(\mathbf{S}+1)+1} = \left\{ \mathbf{I}_{\mathbf{R}(\mathbf{S}+1)+1} - \hat{\mathbf{A}}^{\mathbf{R}(\mathbf{S}+1)+1 \times \mathbf{R}(\mathbf{S}+1)+1} \right\}^{-1} \cdot \Delta \mathbf{Y}^{\mathbf{R}(\mathbf{S}+1)+1}, \quad (14)$$

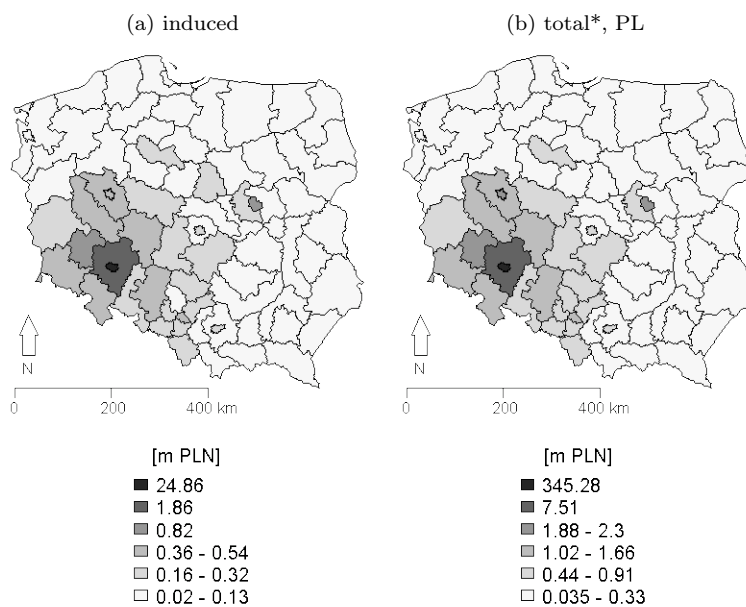
where $\Delta \mathbf{Y}^{\mathbf{R}(\mathbf{S}+1)+1}$ is the vertical vector of final demand (with the analysed enterprise’s sold output value as the last element and zero otherwise) and $\Delta \mathbf{X}^{\mathbf{R} \cdot \mathbf{S}+1}$ is the resulting global output vertical vector across all sectors (including households) and regions.

Figure 9: Simulation: direct and indirect effects on global output



Source: author.

Figure 10: Simulation: induced and total* effects on global output



* = direct + indirect + induced.

Source: author.

Table 4: Share of indirect effects in the domestic region in total indirect effects with confidence interval

confidence level	upper limit	lower limit
90%	72.9%	93.7%
95%	71.5%	92.4%

Share simulated for posterior means: 82.2%.

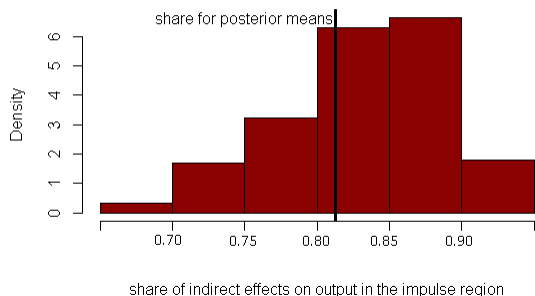
Source: author.

The highest indirect and induced effects are allocated to the regions that (i) add most value to manufacturing and (ii) are situated near Wrocław. The territorial units hosting the highest indirect effects can be identified as the top value-adding regions within 100 km from Wrocław, but also the remote capital city of Warsaw (Figure 9b). This is because in some sectors, such as in advanced services (see Fig. 8), distance is not material in the spatial distribution of suppliers. Thus, even relatively distant metropolies benefit from the impulse in terms of indirect effects.

The central question in the communication between entrepreneurs and local governments (or local governments between one another) is the share of indirect effects attributable to the host region. This question is highly case-specific and depends on the spatial structure of firm’s purchases. In this illustrative analysis, the answer largely depends on the choice of the branch subject to the impulse, and the spatial decay profiles linked to its most purchased commodities. Bottom line, 82.2% of the indirect effects arise in the city Wrocław. This fraction originates from the choice of posterior means reported in Table 3 for building $\hat{\mathbf{A}}$. While the deterministic, national I-O implies a certain value of nation-wide indirect effects, and the various θ values do not change the sum of effects over regions (see Toró, 2016 for a formal proof), the uncertainty around θ affects only their spatial distribution.

As a consequence, the question arises about the uncertainty around this fraction due to the uncertainty around θ as reflected in the posterior distribution. To provide the answer, I consider a random subsample of 1000 draws of θ from the posterior distribution, and – for each of them – I re-compute the $\hat{\mathbf{A}}^{[\mathbf{R}(\mathbf{S}+1)+1] \times [\mathbf{R}(\mathbf{S}+1)+1]}$ matrix. Then, I re-run the simulation (14) and re-evaluate the share of indirect effects in the home region in total domestic indirect effects. Table (4) and Figure (11) present the results as, respectively, the confidence intervals and histogram around the fraction of indirect effects in the host region. The 95% HPDI ranges from 71.5% to 92.4%. Note, however, that in some practical applications – when the last column of matrix $\hat{\mathbf{A}}$ is known to an enterprise as its spatio-sectoral cost structure – this interval would be much narrower, as the impulse would remain unchanged between individual simulations. The provided interval width can be adequate to a situation in which the enterprise knows its sectoral cost structure, but not the spatio-sectoral one, and decides to approximate the latter with the estimates from Section (4) conditionally on its own sector and location.

Figure 11: Share of indirect effects located in the domestic region: point and interval assessment



Source: author.

6 Sensitivity analysis

6.1 Sensitivity to distance measure

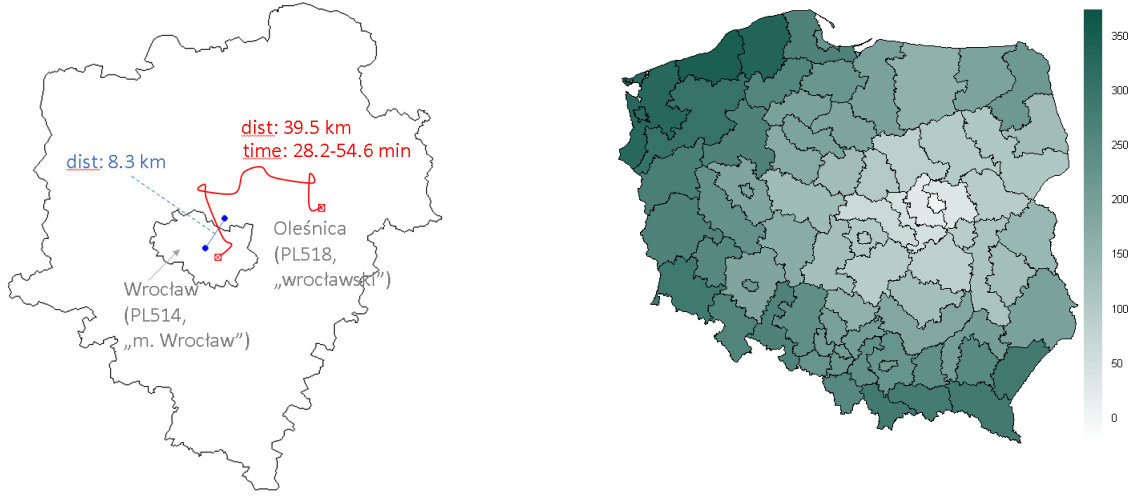
As a robustness check, I also consider the driving times as an alternative \mathbf{W}^* matrix builder. If a given sector (say, simple services) values driving time more than any other category of the transport cost, it can be sensitive to the specific choice of \mathbf{W}^* . This can be observed in both Figure 12a (driving time doubles in the rush hours) and 12b. The latter illustrates the driving times from Warsaw to the given region’s “capital” (defined as previously) and reveals that some destinations are available more quickly than others, in spite of the apparently equal distance from Warsaw. The list of possible reasons includes at least (i) differences in road quality (e.g. highways), (ii) real-world road length (cf. Fig. 12a), (iii) exact location of the region capital, and possibly more. To ensure that the time-based version of \mathbf{W}^* , henceforth referred to as $\tilde{\mathbf{W}}$, is not biased by a specific day- or week-time of the data acquisition, and the calculation does not rely on the data provider’s averaging algorithms, the following sampling scheme has been applied. The data on driving time between each pair of points was downloaded from Google Maps API 42 times, in 4 hour intervals, over the timespan of one week (11-17. September 2021) at 0:00, 4:00, 8:00, 12:00, 16:00 and 20:00. A single download last for approximately 30 minutes, and the download was effected so as to account for the current traffic (function *google_directions* from the R package *googleway*, with departure time set at the moment of download, and the traffic model “best guess”). The average of the 42 matrices was used as $\tilde{\mathbf{W}}$.

For the sake of notational convenience, consider a scalar-normalized version of $\tilde{\mathbf{W}}$ such that the sum of its elements is equal to its counterpart from \mathbf{W}^* . In the generalized version of the model, eq. (6) reads as follows:

Figure 12: Illustrative comparison of distance metrics

(a) from Wrocław city to wrocławski subregion

(b) driving time from Warsaw to a given region's "capital"



Source: author.

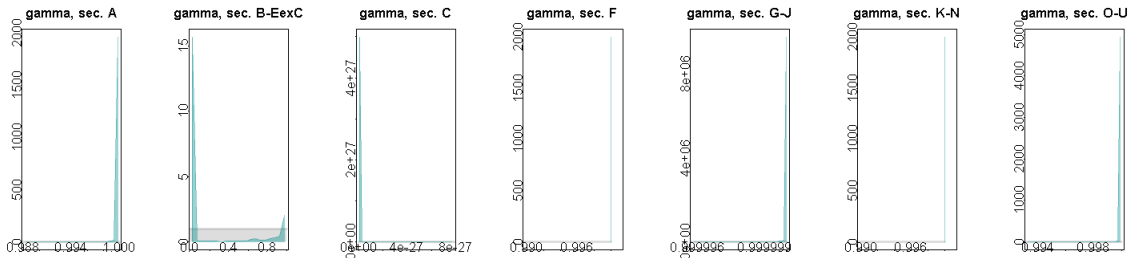
$$\hat{w}_{r,p}^s \equiv \frac{1 - P_{\Gamma} \left[\theta_1^s, \theta_2^s, (w_{r,p}^*)^{1-\gamma^s} \cdot (\tilde{w}_{r,p})^{\gamma^s} \right]}{\sum_{r=1}^R \left\{ 1 - P_{\Gamma} \left[\theta_1^s, \theta_2^s, (w_{r,p}^*)^{1-\gamma^s} \cdot (\tilde{w}_{r,p})^{\gamma^s} \right] \right\}}. \quad (15)$$

The priors of γ^s are set as uniform, independent of one another and of other parameters:

$$\gamma^s \sim U(0, 1), \quad s = 1, \dots, S \text{ i.i.d.} \quad (16)$$

Note that the posterior distributions of γ^s tend to concentrate on the right of the feasible $[0; 1]$ interval.

Figure 13: Parameter γ : prior and posterior distributions



Source: author.

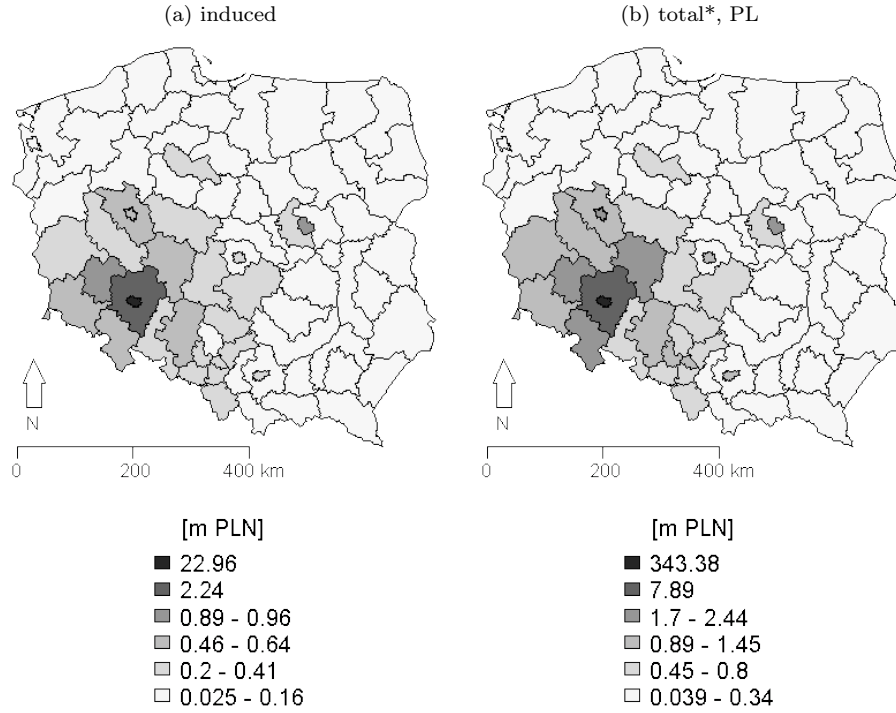
The driving time is a dominant metric for all kinds of services (G-J, K-N, O-U), as well as the agriculture and construction. For the rest of industry (B-E), including manufacturing (C), the opposite is the case, i.e. the driving distance appears as a more relevant measure. However, the posterior distributions derived for the hybrid measure of distance remain roughly unchanged (see Appendix B for the counterpart of Fig. 7). The spatial decay profiles computed with the posterior means of θ for the generalized model (see Appendix B for and Fig. 8) are not straightforward to interpret due to the hybrid distance matrix, but – as a general rule – the alternative parameter set implies slightly shorter average supply distance for all commodities except agricultural ones. Consequently, the simulated indirect effects resemble those presented in Figure 9, with a minor upward correction in favour of the areas located closer to the home region.

Since the two versions of the distance matrix are, all in all, relatively similar and the downward shift in distance is clearly included in the confidence intervals around the decay profiles, one cannot overstate these differences and has to verify whether one version of the model dominates the other in terms of the empirical fit, given the degree of parsimony. Let M_1 be the general model discussed here, and M_2 – the model described in Section 4. I evaluate the marginal likelihood of both models with the bridge sampling technique (see Meng and Wong, 1996; Meng and Schilling, 2002) implemented in the R package *bridgesampling* (Gronau et al., 2020), and then compute the Bayes factor $BF\left(\frac{M_1}{M_2}\right) = 4.9 \cdot 10^{10}$. This means very strong evidence (Kass and Raftery, 1995) in favour of M_1 . Hence, despite the relatively low discrepancy between the spatial decay profiles

6.2 Sensitivity to commuting effects

By comparing Figures 10a and 14, one can notice that the use of the commuting matrix \mathbf{C} has slightly reduced the fraction of induced effects allocated to the region of impulse origins compared to the model version where commuting effects are assumed away (24.9 vs 23 m EUR). This difference obviously depends on the location, and its biggest impact can be observed when the analysis focuses on firms or projects located in the centres of agglomerations. On the one hand, the difference – 2.92 m EUR – appears to be non-negligible as it accounts for 4.9% of the total induced global output. On the other hand, it is rather tiny once as compared to the total effects. The raw data indicate the “ring” NUTS-3 region wrocławski as the dominant region of origin for commuters working in Wrocław city, and – unsurprisingly – this area becomes the main beneficiary in terms of the indirect effects once the shift is accounted for.

Figure 14: Simulation: induced and total* effects on global output



* = direct + indirect + induced.
 Source: author.

6.3 Sensitivity to prior hyperparameters: scale matrix in inverse-Wishart distribution

Lastly, I investigate the sensitivity of all the results presented above to the values of potentially fragile prior hyperparameters: the diagonal elements of \mathbf{H} in eq. 12. Whereas the baseline elicitation puts there 40% of the variance of respective dependent variable, I consider a range from 30% to 50%. All the results remain qualitatively unchanged.

7 Conclusions

This study reconsiders the previous approaches taken in the literature to build the interregional input-output tables by means of spatial econometric methods. I use the previously proposed spatial Durbin model's likelihood function, but analyse the spatial model with Bayesian methods. This yields two kinds of advantages. Firstly, it solves the convergence problems faced by the local (or even global) likelihood maximization algorithms since it assigns zero prior density to certain areas where the likelihood becomes invariant to parameter values. Such a phenomenon arises due to the use

of – otherwise parsimonious and flexible – gamma cumulative distribution functions in the role of the functional form building spatial decay profiles describing the supply geography for each sector. Secondly, it enables using other data sources to elicit priors. I decided to use to this aim the well-known dataset provided by Statistics Finland (2006).

The empirical analysis is conducted for Poland, with the data from the period 2013-2017 and NUTS-3 (subregional) spatial granularity. The posterior means for the shape and scale parameters in the sector-specific spatial decay profiles indicate that agricultural commodities and advanced services are supplied to the most distant locations, while the simple services – to the least distant ones. The highest uncertainty around the spatial decay profile characterizes the former group of sectors. Furthermore, while the illustrative simulation allocates 82.2% of the indirect effects to the home region, multiple simulations conducted with various IRIO matrices built conditionally on the parameters drawn from the posterior distribution suggest a confidence interval of 71.5% to 92.4% for this fraction.

In this study, I used two big data repositories helpful for investigating spatial relationships. Firstly, the distance matrix has been build by measuring the real-world driving distance between the most populated cities of regions, obtained via Google Maps API. I also consider an alternative distance measure – the driving time, averaged over 42 equidistant moments during a 7-day week. A generalized version of the model, in which the hybrid road- and time-based matrices serve as the distance matrices (with sector-specific weighting), does not make a material difference, but – statistically – is highly preferred, according to the Bayes factor. The posterior distribution of weights on both distance matrices concentrate around the extreme values that strongly prefer one of the two: the driving distance for industry except construction (including manufacturing), and the time distance – elsewhere.

Secondly, the simulation of the spatial induced effects accounted for cross-regional commuting of employees. To this aim, the geolocation data from mobile devices, provided by Datarino, was used. The results were close to, but not exactly the same as the results obtained with no commuting. In the considered scenario, where the impulse occurred at the heart of a big agglomeration (Wrocław city), commuting shifts 4.9% of the induced effects outside the home region.

Despite this study solves selected problems emerging in the previous spatial approaches to IRIO table generation, a new challenge emerges. The use of the Finnish dataset for prior elicitation excludes the possibility of their sensible use as a benchmark. Therefore, the data for validations should be sourced elsewhere. One potential way could be using the *Standard Audit File for Tax* database, owned by the Polish tax authorities. When merged with the REGON database containing enterprise-level dominant NACE codes, it could also be summarized as a spatio-sectorial $(R \cdot S) \times (R \cdot S)$ matrix. While not free from some other drawbacks (like missing entity-level data for monopolistic branches resulting from data protection laws, or the mismatches between invoicing and value-adding locations), it could provide a valuable material for comparison. Future research should also investigate the issue of potential error non-normality by building alternative, adequate likelihood functions.

Acknowledgements

The support of the National Science Centre in Poland is gratefully acknowledged (research project 2018/31/D/HS4/00316). The author is also grateful to the participants of 47th Macromodels International Conference 2021 in Wieliczka, in particular to Jacek Osiewalski and Justyna Wróblewska, for useful comments, as well as to Piotr Pękała for invaluable support in the data acquisition process. The usual disclaimer applies.

References

- Bonfiglio A. (2009):** *On the parametrization of techniques for representing regional economic structures*, Economic Systems Research, 21(2), 115–127.
- Flegg A., Tohmo T. (2013a):** *Estimating Regional Input Coefficients and Multipliers: The Use of FLQ is Not a Gamble*, Regional Studies, 50(2), 310–325, doi: 10.1080/00343404.2014.901499.
- Flegg A., Tohmo T. (2013b):** *Regional Input-Output Table and the FLQ Formula: A Case Study for Finland*, Regional Studies, 47(5), 703–721.
- Flegg A.T., Tohmo T. (2016):** *Estimating Regional Input Coefficients and Multipliers: The Use of FLQ is Not a Gamble*, Regional Studies, 50(2), 310–325.
- Flegg A.T., Webber C.D., Elliott M.V. (1995):** *On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables*, Regional Studies, 29, 547–561.
- Folmer H., Nijkamp P. (1985):** *Methodological aspects of impact analysis of regional economic policy*, Papers of the Regional Science Association, 57(1), 165–181.
- Gordon I.R. (1976):** *Gravity Demand Functions, Accessibility and Regional Trade*, Regional Studies, 10, 25–37.
- Gronaua Q.F., Singmann H., Wagenmakers E.J. (2020):** *bridgesampling: An R Package for Estimating Normalizing Constants*, Journal of Statistical Software, (92(10)), 1–29.
- Haining R., Li G. (2020):** *Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach*, CRC Press.
- Jackson R.W., Schwarm W.R., Okuyama Y., Islam S. (2006):** *Annals of Regional Science*, A Method for Constructing Commodity by Industry Flow Matrices, 40, 909–920.
- Jahn M. (2017):** *Extending the FLQ formula: a location quotient-based interregional input-output framework*, Regional Studies, 51(10), 1518–1529.
- Kass R.E., Raftery A.E. (1995):** *Bayes Factors*, Journal of the American Statistical Association, 90(430), 773–795.

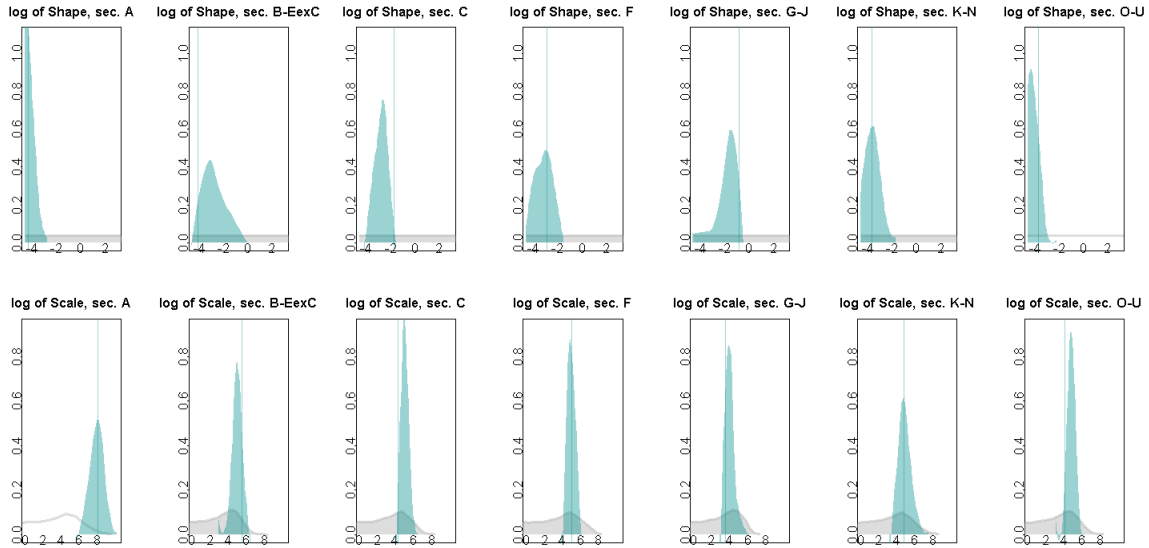
- Kronenberg T. (2009):** *Construction of Regional Input-Output Tables Using Nonsurvey Methods: The Role of Cross-Hauling*, International Regional Science Review, (32), 40–64.
- Lamonica G.R., Chelli F.M. (2017):** *The performance of non-survey techniques for constructing sub-territorial input-output tables*, Papers in Regional Science, doi:10.1111/pirs.12297.
- Lamonica G.R., Recchioni M.C., Chelli F.M., Salvati L. (2020):** *The efficiency of the cross-entropy method when estimating the technical coefficients of input-output tables*, Spatial Economic Analysis, 15(1), 62–91.
- Leontief W., Strout A. (1963):** *Structural Interdependence and Economic Development*, chapter Multiregional Input-Output Analysis, 119–149, London: Macmillan (St. Martin's Press).
- LeSage J.P., Pace R.K. (2009):** *Introduction to Spatial Econometrics*, Chapman & Hall.
- Lindall S., Olson D., Alward G. (2006):** *Deriving Multi-Regional Models Using the IMPLAN National Trade Flows Model*, Journal of Regional Analysis and Policy, 36, 76–83.
- Loveridge S. (2004):** *A Typology and Assessment of Multi-sector Regional Economic Impact Models*, Regional Studies, 38(3), 305–317.
- Meng X., Schilling S. (2002):** *Warp Bridge Sampling*, Journal of Computational and Graphical Statistics, 11(3), 552–586.
- Meng X.L., Wong W.H. (1996):** *Simulating Ratios of Normalizing Constants Via a Simple Identity: A Theoretical Exploration*, Statistica Sinica, (6), 831–860.
- Miller R.E., Blair P.D. (2009):** *Input-Output Analysis. Foundations and Extensions*, Cambridge University Press, Cambridge.
- Morrison W.I., Smith P. (1974):** *Nonsurvey input-output techniques at the small area level: an evaluation*, Journal of Regional Science, 14(1), 1–14.
- Polenske K.R. (1970):** *An Empirical Test of Interregional Input-Output Models: Estimation of 1963 Japanese Production*, American Economic Review, 60, 76–82.
- Rey S.J. (2000):** *Integrated regional econometric + input-output modeling: Issues and opportunities*, Papers in Regional Science, 79(3), 271–292.
- Round J.I. (1978):** *An interregional input-output approach to the evaluation of nonsurvey methods*, Journal of Regional Science, 18, 179–194.
- Statistics Finland (2006):** *Construction of Regional Input-Output Tables in Finland 2002*, 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece.
- Stone R. (1961):** *Input-Output and National Accounts*, OECD, Paris.

- Temursho U., Oosterhaven J., Cardenete M.A. (2021):** *A multi-regional generalized RAS updating technique*, Spatial Economic Analysis.
- Tobler W. (1970):** *A computer movie simulating urban growth in the Detroit region*, Economic Geography, 46(Supplement), 234–240.
- Torój A. (2016):** *Regional Economic Impact Assessment with Missing Input-Output Data: A Spatial Econometrics Approach for Poland*, Central European Journal of Economic Modelling and Econometrics, 8, 61–91.
- Torój A. (2021):** *Construction of multiregion-multisector input-output tables: a spatial econometric approach for Poland*, Spatial Economic Analysis, 16(4), 550–569.
- Wiedmann T., Wilting H.C., Lenzen M., Lutter S., Palm V. (2011):** *Quo Vadis MRIO? Methodological, data and institutional requirements for multi-region input-output analysis*, Ecological Economics, 70, 1937–1945.
- Yamagata Y., Seya H. (eds.) (2020):** *Spatial Analysis Using Big Data. Methods and Urban Applications*, Elsevier.

Appendix A: Commuting matrix C

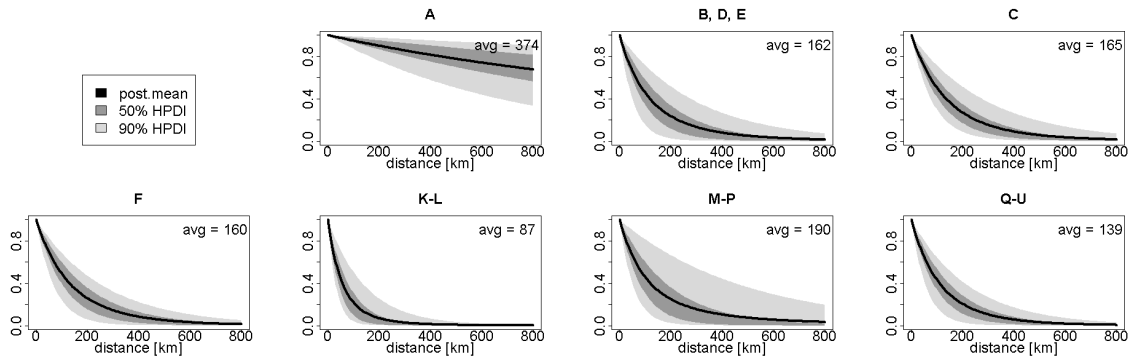
Appendix B: Priors, posteriors and spatial distance profiles for the general model in Subsection 6.1

Figure 15: Shape and scale: prior (gray) and posterior (green) distributions [W based on driving time]



Source: author.

Figure 16: Estimated spatial decay profiles for Poland across sectors [W based on driving time]



Source: author.