Dipartimento delle Finanze

working paper

Bilateral Regional Trade Flows in Italy: an Origin-Destination-Commodity GWR-SAR approach

DF WP n.18
May 2023

Alessio BALDASSARRE, Danilo CARULLO, Paolo DI CARO, Elisa FUSCO, Pasquale GIACOBBE, Carlo ORECCHIA

DF Working Papers, 2017

MEF Ministero dell'Economia e delle Finanze
Dipartimento delle Finanze
Via dei Normanni, 5
00184 Roma

Email df.wpapers@finanze.it
http://www.finanze.gov.it/

ISSN electronic edition 2610-8887

Papers can be downloaded free of charge from the MEF website:
http://www.finanze.it/opencms/it/il-dipartimento/documentazione/Collana-di-lavori-e-di-ricerca/

Bilateral Regional Trade Flows in Italy: an Origin-Destination-Commodity GWR-SAR approach*

Alessio Baldassarre[a], Danilo Carullo[a], Paolo Di Caro[b], Elisa Fusco[c,*], Pasquale Giacobbe[d], Carlo Orecchia[a]

[a] *Department of Finance; Italian Ministry of Economy and Finance, Italy.*
[b] *Department of Law, University of Catania, Italy.*
[c] *Department of Economic modelling and statistical analysis for policy making; Sogei SpA, Italy.*
[d] *Department of Business Administration and Law, University of Calabria, Italy.*

Abstract

The main purpose of this paper is to present an innovative approach to estimate the Italian inter-regional trade flows in terms of final and intermediate consumption. It contributes to the literature in several ways. The first innovative feature concerns the data used in the analysis. We reconstruct the flow of households' final consumption by using administrative data from the Italian VAT returns. The result is then used for estimating a traditional gravity model for final consumption trade; the estimated coefficients are furtherly exploited to compute the flows of intermediate consumption. The second contribution relates to the modeling approach: we combine the literature on gravity models with a spatial autoregressive specification, to take into account spatial dependence in the bilateral flows, and a geographically weighted regression estimator, to control for behavioral instability of data over space. In addition to that, our model controls for commodity dependence by including them as a fixed effect in a pseudo-panel view, where the time dimension captures the commodities dynamics. Therefore, the strategy here introduced is useful to consider both local level economic relations and spillovers, existing between regions, and the link among different types of products.
JEL classification *C21, D57, R15.*

Keywords: Inter-regional flows; Gravity models; O-D Spatial autoregressive models; Geographically Weighted Regression

---

*Corresponding author. E-mail: efusco@sogei.it

*This paper has been developed at the Department of Finance-Italian Ministry of Economy and Finance under the project 'Assistance for the assessment of environmental tax reforms', funded by the European Commission-DG Reform, in collaboration with The World Bank. We are particularly grateful to Stefano Curto, Marco Manzo, Maria Teresa Monteduro, Gavino Mura, Hasan Dudu, Louise Wandhal-Jensen, and Jan Ignacy Witajewski-Baltvilks. We wish to thank Sogei SpA for the support. The views expressed in the paper do not reflect those of the institutions of affiliation. The usual disclaimer applies. Disclaimer: The findings, interpretations, and conclusions expressed herein are those of the authors and do not necessarily reflect the view of the European Commission, Italian Ministry of Economy and Finance, Sogei SpA, World Bank Group, its Board of Directors or the governments they represent. The author, Pasquale Giacobbe, was working at Sogei SpA at the time of writing.

1. Introduction

Understanding intra-national interactions in economic aggregates such as trade, investments, and human capital is crucial for analyzing some of the mechanisms underlying economic growth and regional development gaps. Thissen *et al.* (2019), among others, point out how: "*Economic development is inter-regional in nature, with economic growth being determined by physical and technological proximity identified by interregional and national cross-border interactions*".

The trade literature conventionally focuses on foreign trade while the within-country dimension of the trade flows has been underinvestigated. This gap has initially been justified by arguing that concepts like comparative advantage were less influential in a regional view since regions share language, culture, and legislation and are characterized by the absence of tariffs, easier factor mobility, and fewer difficulties like shorter distances (Bentivogli *et al.*, 2019), which turn out in lower frictions affecting within-country trade flows. On the other hand, the most severe obstacle in carrying out regional trade flow modeling consists of data availability. Inter-regional trade does not get statistically recorded (Ivanova *et al.*, 2009) and this translates into a lack of official data for the trade flow of goods and services between regions.

From a technical point of view, trade flows estimations have been widely dealt with through gravity models. Gravity models are based on the assumption that trade relationships are influenced by two factors: the *(i)* economic sizes of two places and their *(ii)* distance (Tinbergen, 1962). Such models have been defined as *"Some of the clearest and most robust findings in empirical economics."* (Leamer & Levinsohn, 1994). Disdier & Head (2008), for example, in their meta-analysis on the effect of distance on trade, cover 1,052 separate estimates in 78 papers. Gravity models are also widespread in other areas of application concerning different factor movements (Anderson, 2011), for instance, domestic tourism flows (Song *et al.*, 2012; Marrocu & Paci, 2013; Morley *et al.*, 2014; Santeramo & Morelli, 2016), freight or passenger transport (Grosche *et al.*, 2007; Doucet *et al.*, 2014; Zhang *et al.*, 2018), human mobility flows (Barbosa *et al.*, 2018), migration flows (Lewer & Van den Berg, 2008; Beine *et al.*, 2016).

Besides these aspects, the estimation of the inter-regional trade plays a paramount role in the multi-regional input-output analysis. Trade among regions is a complex and multidimensional issue, indeed, it depends on many factors, such as geographic and cultural proximity of the trading partners, consumer preferences, trade costs, and the structure and size of trading economies (Ivanova *et al.*, 2009). Thus, proper modeling of the inter-regional trade flow permits to disentangle of potential feedback effects and spillover characterizing regional interconnections (Sargento *et al.*, 2012) and get robust inter-regional trade estimations which, in turn, serve as a cornerstone in the construction of multi-regional input-output systems. In addition, the estimation of inter-regional trade flows represents a crucial issue when dealing with Computable General Equilibrium (CGE) models (Johansen, 1960). Several works investigate social-economic differences across the administrative units (*e.g.*, regions) within countries. Regional CGE models analyze how different regions would respond to a given shock, and for this purpose, the inter-regional trade flows need to be estimated. Among CGE models, RHOMOLO is a dynamic spatial general equilibrium model that aims at investigating ex-ante the impact of policy instruments for 267 NUTS2 regions of the EU (Lecca *et al.*, 2018). As we remarked before, the estimation of inter-regional trade flows constitutes a crucial step also for this model: the authors, indeed, combine the national information from the World Input-Output Database (WIOD) as macro-constraints with the prior data on trade flows developed by Thissen *et al.* (2013).

The arguments exposed so far shed light on the importance of properly estimating the patterns, the intensity, and the determinants of bilateral trade flows.

By deeply analyzing the literature on bilateral trade flows, we noted an important drawback in the absence of solid empirical evidence on Italy. Although Italy is characterized by a high heterogeneity among its regions, as far as we know, just a few papers investigated this topic with a focus on the Italian context. Among the notable ones, Paniccià & Rosignoli (2018) estimated the inter-regional flows through a "*deterrence function*" which assumes that the trade interaction between two regions decreases if distance, cost, and travel time increase and, on the contrary, increases if the amount of activity (production) increases. The deterrence factor can be defined as the ratio of flows with transaction costs to flows without costs, assuming a proportional relationship between costs and trade flows, namely if costs are too high the interaction between the two regions decreases and vice versa. The estimation procedure involves distinct OLS (*Ordinary Least Squares*) regressions for manufacturing and the services sectors and specific analysis for special products. Thissen *et al.* (2019) proposed a constrained nonlinear programming method to determine the trade flow between two regions as a "minimum path". Flows are determined in terms of the distance between regions, weighted by the probability of passing through intermediate logistics centers. Specifically, they use a two-stage method: *(i)* in the first step they determine some matrices (one for each product considered) constituted by the probability that a product is carried from one region to another passing through 0, 1, 2, 3 or 4 logistics centers; *(ii)* in a second step the distance of the trade flows is minimized using a non-linear programming problem. The ENI foundation method (Standardi *et al.*, 2014) defines, from transport data, a system of national accounting equations from which they derive macro-area level output.

A noteworthy methodological aspect comes from the fact that classical gravity models assume that individual flows are independent of each other, and thus, no spatial autocorrelation must be present in the residuals. However, this assumption is not realistic, as it is almost sure that the proximity of the regions influences trade flows and, if spatial dependence is detected, the OLS estimated parameters may be biased, bringing to overestimates or underestimates of the unknown true value. This concern is strengthened by the fact that, as highlighted in Sargento (2007), a gravity model can be seen as a spatial interaction model, where the spatial interaction flows can be known or unknown a priori. In the former case, the flows are explained through econometric modeling, in the latter the model is applied to assess the unknown flows.

Therefore, to deal with this issue and to incorporate spatial autocorrelation, some scholars have proposed a Spatial Autoregressive (SAR) specification of the gravity model (LeSage & Pace, 2008; LeSage & Thomas-Agnan, 2015). Compared to the classical statistical models, here the flows are characterized by three types of dependency: *(i)* from an origin region to neighbors of the destination region, *(ii)* from neighbors of an origin region to destination region, *(iii)* the interdependence from neighbors to the origin region to neighbors of the destination region. Moreover, Liu *et al.* (2015), following the SAR approach, proposed to estimate the model through a GWR (*Geographically Weighted Regression*) to account also for spatial heterogeneity.

In this paper, to consider the location and geographic concentration of economic activities and, therefore, the economic relationships existing at the local level, and the behavioral instability of data over space, the approach proposed to estimate Italian inter-regional trade flows grounds on the idea of LeSage & Pace, 2008, as regard the model specification, and of Liu *et al.* (2015) as regard the choice of a GWR estimator. In addition to that, to take into account the dependence among commodities, a *pseudo-panel* is constructed where the time-dimension represents the commodities side and commodities' dummy variables are included as fixed effects in the model.

The paper is organized as follows. In Section 2 the statistical methodology is described; Section 3 discusses the empirical strategy for the Italian case; in Section 4 dataset construction details and variables explanation are provided; Section 5 presents the estimation results and a comparison with some institutional data sources; Section 6 concludes.

2. Methodological proposal

The application of gravity equations to the empirical analysis of international trade was pioneered by Tinbergen (1962), Pöyhönen (1963), Pulliainen (1963) and Sawyer (1967). Gravity models are based on the assumption that trade relationships y_i^{od} are influenced by the economic sizes of two regions and by their distance.

$$y_i^{od} = e^{\beta_0} \cdot \frac{(X_i^{o\cdot})^{\beta_1} \cdot (X_i^{\cdot d})^{\beta_2}}{(d^{od})^{\beta_3}}, \forall i = 1, ..., I \tag{1}$$

where y_i^{od} are the trade flows of commodity i from the origin region $o = 1, ..., O$ to the destination region $d = 1, ..., D$; $X_i^{o\cdot}$ are the total outflows of commodity i from region o (supply region; $X_i^{\cdot d}$ are total inflows of commodity i to region d (demand region); d^{od} is the distance between regions o and d; and β_0, β_1, β_2, β_3 are the relative parameters to be estimated.

A log-linearisation of Equation 1 is applied to perform estimation using linear regression models[2]:

$$ln(y_i^{od}) = \beta_0 + \beta_1 ln(X_i^{o\cdot}) + \beta_2 ln(X_i^{\cdot d}) + \beta_3 ln(d^{od}) + \epsilon, \forall i = 1, ..., I \tag{2}$$

and a matrix formulation is provided to generalize for the case of n regions:

$$Y = \beta_0 L_n + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{3}$$

where L_n is a $N \times 1$ matrix with all elements equal to 1; X_1 is the total supply from supply regions for commodity i; X_2 is the total demand from demand regions for commodity i; X_3 is the distance between two regions.

Thus, spatial terms are included as follows:

$$Y = \rho_o W_o Y + \rho_d W_d Y + \rho_w W_w Y + \beta_0 L_n + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{4}$$

where $W_o Y$ measures the spatial dependence on the origin of the trade flow; $W_d Y$ measures the spatial dependence on the destination of the trade flow; $W_w Y$ measures the interdependence between the origin and the destination of trade flow, and ρ_o, ρ_d, ρ_w are additional parameters, to be estimated and that provide information on the spatial dependence in the dependent variable Y (see among others LeSage & Pace, 2008).

The main complexity of such a model is that origin-destination flows require, as a first step, the vectorization for each commodity of the $n \times n$ square matrix of inter-regional flows from each of the n origin regions o to each of the n destination regions d. The result of this operation consists of a $n^2 \times 1$ vector of trade flows obtained by stacking the columns of the flow matrix into a vector representing our dependent variable Y.

More in detail, the idea behind LeSage & Pace (2008) is that: "*(i) large commodity flows from region o (origin) to region d (destination) might be accompanied by similarly large flows from neighbors of the region o to region d; (ii) large commodity flows from region o to region d might be accompanied by similarly large flows from region o to neighbors of the region d; (iii) large commodity flows from region o to region d might be accompanied by large flows from neighbors of the region o to neighbors of region d*" (Lesage & Polasek, 2008). The trade flows referred to in point *(i)* represent what LeSage & Pace (2008) describe as *origin-based dependence*, point *(ii)* is labeled as *destination-based dependence*, and point *(iii)* constitutes the *origin-destination dependence*. This decomposition of the spatial dependence starts from the usual row-standardized spatial weight matrix W, reflecting relations be-

[2]Note that the sign of the estimated β_3 is expected to be negative.

tween the origin/destination regions, and brings to the construction of three spatial weight matrices accounting for these three different types of spatial connectivity between origin and destination regions: *(i)* W_o is the spatial weight matrix capturing "origin-based" spatial dependence relations using an average of flows from neighbors of each origin region to each destination region[3] $(W \otimes I_n)$; *(ii)* W_d is the spatial weight matrix capturing the "destination-based" spatial dependence *i.e.* the connectivity relations between the flows from an origin region to neighbors of the destination region $(I_n \otimes W)$; *(iii)* the third type of dependence to consider is $W_w = W \otimes W$ that is a spatial weight matrix reflecting an average of flows from neighbors of the origin region to neighbors of the destination region. This methodology allows framing the connectivity between n^2 origin-destination pairs of regions (*i.e.* bilateral trade flows) by providing a framework for modeling the connectivity of origin-destination regions in a perspective consistent with the usual spatial autoregressive models.

However, it is important to note that the estimation of Equation 4 is set to be carried out for each commodity $(i = 1, ..., I)$ assuming implicit independence among commodities; to overcome this drawback, direct incorporation of the commodities is proposed by constructing a *pseudo-panel*, where the time-dimension represents the commodities side. This implies, that connectivity between $n^2 \times I$ origin-destination pairs of regions by commodities have to be modeled, which translates into a n origins $\times$ n destinations $\times$ I commodities sample as shown in Table 1.

Table 1: Dataset example

Origin	Destination	Commodity	Variable
Region 1	Region 1	Commodity 1	...
Region 1	Region 2	Commodity 1	...
. . .	....	...	...
Region 1	Region d	Commodity i	...
....	...	...	...
Region O	Region D	Commodity I	...

In this case, to consider the type of commodity in the estimation, a fixed effect is included in Equation 4, becoming an LSDV (*Least squares with dummy variables*)

[3]Note, $\otimes$ is the *Kronecker product* that allows obtaining vectors without having to deal directly with $n^2 \times n^2$ matrices improving computational efficiency. Such vectors aid equally in the interpretation of the origin-destination dependence (LeSage & Thomas-Agnan, 2015).

model with a SAR specification:

$$\begin{aligned} Y =& \rho_o W_o Y + \rho_d W_d Y + \rho_w W_w Y + \beta_0 L_n + \\ & \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \sum_{i=1}^{I-1} \gamma_i D_i + \epsilon \end{aligned} \tag{5}$$

where D_i are the commodities' dummy variables.

Finally, if a GWR-SAR is implemented, with the aim of considering also the spatial heterogeneity (see Jaya *et al.*, 2018; Geniaux & Martinetti, 2018; Tomal, 2020), Equation 5 can be rewritten as follows

$$\begin{aligned} Y =& \rho_o(u_r, v_r) W_o Y + \rho_d(u_r, v_r) W_d Y + \rho_w(u_r, v_r) W_w Y + \beta_0(u_r, v_r) L_n + \\ & \beta_1(u_r, v_r) X_1 + \beta_2(u_r, v_r) X_2 + \beta_3(u_r, v_r) X_3 + \sum_{i=1}^{I-1} \gamma_i(u_r, v_r) D_i + \epsilon \end{aligned} \tag{6}$$

where (u_r, v_r) denotes the longitude and latitude of a region r. GWR is a weighted regression where weights are given by the geographical distance between regions and the weight is higher if the geographical distance to the point is smaller and gradually reduces if the distance increases. Therefore, the regression coefficients are not fixed but depend on the geographical coordinates of observations allowing local rather than global parameters to be estimated. Note that permitting ρ_o, ρ_d, and ρ_w parameters to vary over regions means estimating also a local spatial dependence other than a global one.

3. Empirical strategy

The empirical strategy aims at estimating inter-regional flows for both final and intermediate consumption. First, we reconstruct the households' final consumption flows by using administrative data from the Italian Value Added Tax (VAT) returns database. Second, we estimate the determinants of households' final consumption flows through a traditional gravity trade model. Third, the estimated coefficients are furtherly exploited to compute the flows of intermediate consumption. Finally, we correct the results with specific adjustment factors.

In particular, in the first step, inter-regional flows of the households' final consumption are reconstructed, starting from fiscal administrative data of VAT (calculation procedure is explained in detail in Subsection 4.1). Then, in the second step, we model the determinants of regional bilateral flows in terms of households' final consumption. Hence, following the log-linear specification expressed in Equation 5 we obtain

$$Y_{hh} = \rho_o W_o Y_{hh} + \rho_d W_d Y_{hh} + \rho_w W_w Y_{hh} + \beta_0 L_n + \beta_1 X_1 + \beta_2 X_{2hh} + \beta_3 X_3 + \sum_{i=1}^{I-1} \gamma_i D_i + \epsilon \tag{7}$$

where Y_{hh} is the inter-regional flow for the final consumption of households derived from administrative data.

Then, the estimated $\widehat{\beta}$ parameters are used to predict inter-regional flows ($\widehat{Y}_{IC}$) for intermediate consumption IC by assuming that the size of final consumption determinants is equal to the size of intermediate consumption determinants:

$$\widehat{Y}_{IC} = \widehat{\rho}_o W_o Y_{hh} + \widehat{\rho}_d W_d Y_{hh} + \widehat{\rho}_w W_w Y_{hh} + \widehat{\beta}_0 L_n + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_{2IC} + \widehat{\beta}_3 X_3 \sum_{i=1}^{I-1} \widehat{\gamma}_i D_i \tag{8}$$

where X_{2IC} is the total demand of intermediate commodities from demand regions.

Finally, by following the approach proposed by Liu *et al.* (2015), inter-regional flows for intermediate consumptions are corrected through two adjustment factors which capture the regional interaction in production (C) and the vertical integration of sectors (θ):

$$Y'_{IC} = \frac{\widehat{Y}_{IC}}{C^{\theta}} \tag{9}$$

The regional interaction in production C is calculated for each sector s and each pair of region od by using location quotients (Flegg & Tohmo, 2013). The C factor can be expressed as follows

$$C_s^{od} = \begin{cases} \frac{\mu_s^o + \mu_s^d}{|\mu_s^o - \mu_s^d| + min_{r=1,2,\ldots,n}\mu_s^r} & \text{if } o \neq d \\ 1 & \text{if } o = d, \end{cases} \tag{10}$$

where the degree of interaction C_s^{od} of sector s denotes the regional degree of interaction of the sector s producing commodity i; μ_s^o and μ_s^d are location quotients of sector s producing commodity i for region o and d.

The factor capturing the vertical integration of sectors θ_s is calculated as

$$\theta_s = \bar{\delta} - \delta_s, \tag{11}$$

where δ_s is the ratio between the amount of commodity i produced and implied in sector s for the production process and the total amount of intermediate commodities used by the same sector s; $\bar{\delta}$ is the average of δ_s over all sectors.

4. Data and variables

In this Section, the variables involved in the analysis and the necessary steps to reconstruct the final consumption of households are described. The methodology explained in Section 3 is applied on 20[4] Italian NUTS2 regions (origin region o, destination region d), by considering 20 Nace Rev.2 sectors (s) and 20 Classification of Products by Activities (CPA) commodities (i)[5]. Data refer to the year 2016.

4.1. Final consumption of households reconstruction - Dependent variable

As discussed in Section 1, one of the critical factors in inter-regional trade estimation is the lack of data. In this paper, the just-mentioned drawback is addressed by employing an innovative procedure based on administrative data. Thus, the VAT returns database is used as initial data for the estimation of households' final consumption flows, which constitutes our dependent variable. VAT database is used by the Department of Finance at the Italian Ministry of Economy and Finance for several analyses, such as the development of a distributional microsimulation model for households (Cirillo *et al.*, 2021). It is an important source of information since it contains all the data on VAT, including the tax domicile and the activity sector of the taxpayer. In particular, the VT panel reports the amount of sales of goods and services, from each VAT payer to final consumers and VAT holders, by differentiating for the buyers' location.

Table 2: Initial data from VT panel

Region	Sector				
	s_1	...	s	...	s_{20}
d_1	$vt^{domicile}_{od_1,s_1}$				
...		...		...	
d			$vt^{domicile}_{od,s}$		
...		...		...	
d_{20}					$vt^{domicile}_{od_{20},s_{20}}$

Notes: For each origin region $o_1 \dots o_{20}$ where the VAT payers are domiciled, the amount of sales of goods and services $vt^{domicile}_{od,s}$ for each sector $s_1, \dots, s_{20}$ and destination region $d_1, \dots, d_{20}$ is reported.

However, to properly reconstruct inter-regional flows of households' final consumption, these data need some adjustments. More specifically, the VT panel has to be expressed in terms of plant location and commodity, instead of domicile and activity sector, respectively; furthermore, we have to correct for transport and trade

[4]The two autonomous provinces of Trento and Bolzano are considered as a single region labeled Trentino Alto-Adige.

[5]For the list of Italian NUTS2 regions, Nace Rev.2 sectors and statistical classification of products by activity - please see Tables A1,A2 andA3, respectively.

margin bias and, in the end, we must transpose the VT panel data from supply-side to demand-side.

In the first step, a bridge matrix derived from IRAP (Italian Regional tax on Activity Production) database[6] is applied to VT data $vt^{domicile}_{od,s}$ for the transformation from location by tax domicile to the location by the production plant. In detail, we use the information about the (net) value of production provided by fiscal domicile and plant location. This operation is conducted by calculating a re-proportioning coefficient $\xi_{o,s}$ for each origin region o and sector s, which is the ratio between the value of net production per plant $VNP^{plant}_{o,s}$ and the value of net production per domicile $VNP^{domicile}_{o,s}$.

$$\xi_{o,s} = \frac{VNP^{plant}_{o,s}}{VNP^{domicile}_{o,s}} \tag{12}$$

Then, this coefficient is multiplied by the taxable transactions $vt^{domicile}_{od,s}$ obtaining a re-proportioned value of the total amount of taxable transactions $\sum_d vt^{plant}_{od,s}$ in the region o and each sector s.

$$\sum_d vt^{plant}_{od,s} = \sum_d vt^{domicile}_{od,s} \xi_{o,s} \tag{13}$$

Finally, these values are distributed for each pair of regions od according to the weight of the taxable operations in the initial data $vt^{domicile}_{od,s}$.

$$VT^{plant}_{od,s} = \frac{vt^{domicile}_{od,s}}{\sum_d vt^{domicile}_{od,s}} \sum_d vt^{plant}_{od,s} \tag{14}$$

In a second step, an additional bridge matrix Φ, derived from national accounting data[7], is applied to VT data by plant and sector $VT^{plant}_{od,s}$ to distribute them by commodities i. Φ is a bridge matrix with sectors s and commodities i as rows and columns, respectively. Each value Φ_{is} of this matrix represents the national output of commodity i produced by sector s, *i.e.*, $Output_{nat,is}$ with respect to the total output of commodity i:

$$\Phi_{is} = \frac{Output_{nat,is}}{\sum_i Output_{nat,is}} \tag{15}$$

Hence, we obtain the estimated output of commodities for each region.

$$VT^{plant}_{od,i} = \sum_s VT^{plant}_{od,s} \Phi_{is}, \tag{16}$$

[6] IRAP database from Department of Finance-MEF.

[7] ISTAT - Input-Output table, 2016, supply table at purchasers' prices.

In a third step, once the VT data framework by plant and commodity for each region are obtained, we need to correct for margins' bias on electricity, trade, and transport. These three sectors represent a subset $g \in I$, where I is the commodities' set. The VT data, by construction, attribute to margins' commodities g some taxable transactions which belong to other commodities $i \neq g$, *e.g.*, the agriculture commodity produced by region o and transported and sold to region d is attributed to the transport commodity instead of agriculture commodity.

To avoid this bias, we have to identify for each region o and commodity g, how much of the taxable transactions value (linked to g) should be attributed to margin and how much to the amount of produced goods and services. This operation is conducted by using data on national trade, transport margins[8] and national output[9]. Then, we re-distribute the identified margin values to the commodities object of the effective trade flow.

The margins (labeled as Mar in the formulas) identification is obtained by applying shares σ_g to the VT data on plant and commodity. σ_g defines for each commodity g the amount of output to be considered as margin (*e.g.*, 14%, 84%, and 15% of the electricity, trade, and transport commodities are margins and have to be allocated among all the other commodities as flows). These shares are calculated as follows

$$\sigma_g = \frac{Mar_{nat,ii}}{\sum_s Output_{nat,is}}, \forall i = g \tag{17}$$

This step leads to the estimation of the margins $M_{od,g}$ for the three commodities g, which indicate the amount of sales of electricity, trade, and transport that has to be reallocated between all the other commodities.

$$M_{od,g} = VT^{plant}_{od,i}\sigma_g \tag{18}$$

At this stage, we need to reallocate these quantities $M_{od,g}$ across all the other commodities. Therefore, we define η_{ig} which represents the distribution of margin g among all the other commodities i. These values are calculated as follows

$$\eta_{ig} = -\frac{Mar_{nat,ig}}{\sum_i Mar_{nat,ig}} \tag{19}$$

Now we apply these percentages to the margins $M_{od,g}$, in order to reallocate the margins of commodity g to all the others as follows

$$\Delta M_{od,ig} = \eta_{ig} M_{od,g} \tag{20}$$

Hence, we correct the VT data by plant and commodities by reallocating the margins

[8]ISTAT - National Accounting Matrix, 2014.

[9]ISTAT - Input-Output table, 2016, supply table at purchasers' prices.

$$\hat{VT}_{od,i}^{plant} = VT_{od,i}^{plant} + \sum_g \Delta M_{od,ig}, \tag{21}$$

where $\hat{VT}_{od,i}^{plant}$ represents the value of the trade flow of commodity i from region o (by plant) to region d (i.e., source-side).

At this stage, we need to change the point of view by switching from the source side to the destination side (i.e., $\hat{VT}_{od,i}^{plant} \rightarrow \hat{VT}_{do,i}^{plant}$).

Table 3: Final VT after transformations

Region	Commodity				
	i_1	...	i	...	i_{20}
o_1	$\hat{VT}_{do_1,i_1}^{plant}$				
...		...		...	
o			$\hat{VT}_{do,i}^{plant}$		
...		...		...	
o_{20}					$\hat{VT}_{do_{20},i_{20}}^{plant}$

Notes: for each demanding region $d_1 \ldots d_{20}$ and each commodity $i_1 \ldots i_{20}$, it is indicated the flow of the household's final consumption by destination region and commodity $\hat{VT}_{do,i}^{plant}$ for each origin region $o_1, \ldots, o_{20}$.

Finally, to estimate administrative data coherently with the national accounting data, we transform $\hat{VT}_{do,i}^{plant}$ in shares $\tau_{d,i}$. These shares are then applied to aggregated regionalized households expenditure by CPA[10] net to imports linked to households[11].

$$\tau_{d,i} = \frac{\hat{VT}_{do,i}^{plant}}{\sum_o \hat{VT}_{do,i}^{plant}} \tag{22}$$

$$\hat{VT}_{do,i}^{plant} = (HH_{d,i} - Import_{d,i}^{HH})\tau_{d,i}, \tag{23}$$

where $HH_{d,i}$ is the households' consumption in the destination region d[12].

[10]ISTAT - Regional account, 2016.

[11]ISTAT-ICE yearbook "Commercio estero e attivita' internazionali delle imprese", ed 2017, data 2016. The imports are then split in $Import_{d,i}^{HH}$ and $Import_{d,i}^{IC}$ based on the weight of import for household final consumption and import for intermediate consumption, respectively, on the aggregate demand.

[12]National accounts regional main aggregates: Final consumption expenditure of households by expenditure item (Coicop 2 digit) and durability, 2016.

We obtain the flows of the household's final consumption between each pair of origin-destination region and commodity $\hat{VT}_{do,i}$ which represent the response variable, namely, Y_{hh}, used in the second step of the analysis.

4.2. Explanatory variables

The explanatory variables considered in the estimation are:

- X_1: total regional supply (net to exports) of commodity i of region o. We estimate X_1 as the difference between the output $Output_{o,i}$ and the regional exports[13] (Equation 25), where $Output_{o,i}$ is calculated through the application of the location quotients (Flegg & Tohmo, 2013) as in Equation 24.

$$Output_{o,i} = \sum_{s}(Output_{naz,is} * AFLQ^{EMP}_{o,is}), \tag{24}$$

  where $Output_{naz,is}$ is the national output for commodity i and sector s[14], and $AFLQ^{EMP}_{o,is}$ is the Augmented Flegg Location Quotient built on employment data for region o, commodity i, and sector s[15] taking into account the sector specialization and region dimension. At this point, the total regional supply net exports is calculated as follows

$$X_1 = Output_{o,i} - Export_{o,i} \tag{25}$$

- X_2: total regional demand (net to imports) of commodity i of region o. According to our purpose, we split X_2 in:

  - $X_{2,hh}$, (used for estimation): total regional households final consumption demand (net of final consumption imported commodities) of commodity i in region d. $X_{2,hh}$ is calculated as follows

$$X_{2,hh} = HH_{d,i} - Import^{HH}_{d,i} \tag{26}$$

  - $X_{2,IC}$ (used for prediction): total regional intermediate consumption (net of intermediate consumption imported commodities) of commodity i in region d. The total intermediate consumption per region d and commodity i is calculated through the application of the $AFLQ$ on the national intermediate consumption[16] as follows

---

[13]ISTAT-ICE yearbook "Commercio estero e attivita' internazionali delle imprese", ed 2017, data 2016.

[14]ISTAT Input-Output table, 2016, supply table at purchasers' prices.

[15]National accounts regional main aggregates: Declared and undeclared employment by industry and population, 2016.

[16]ISTAT Input-Output, 2016, use table at purchasers' prices.

$$IC_{d,i} = \sum_{s} (IC_{naz,is} * AFLQ_{d,is}^{EMP}) \tag{27}$$

Hence, $X_{2,IC}$ is

$$X_{2,IC} = IC_{d,i} - Import_{d,i}^{IC} \tag{28}$$

- X_3 is the road distance in kilometers calculated as the minimum path between regional centroids from the *OpenStreetMap*'s maps (a visual example is provided in Figure 1);
- W the 20×20 spatial weight matrix deriving from ISTAT shapefiles (neighborhood is identified by considering a radius of 250 kilometers (Figure 2)).

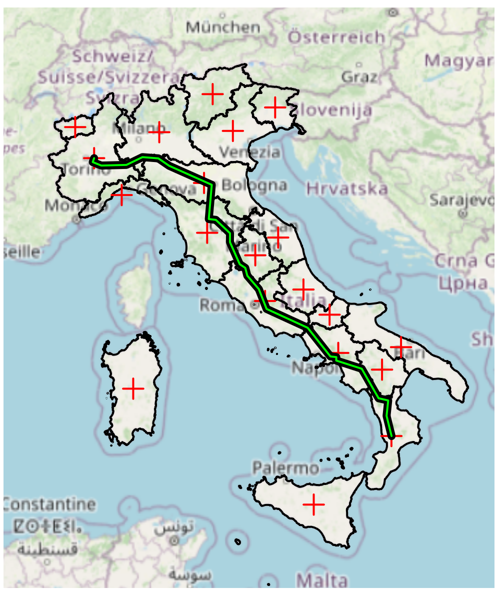


Figure 1: Example of the minimum path between regional centroids

Figure 2: Spatial connectivity structure (250 kilometers)

Descriptive statistics are shown in Table 4. The values for the minimum and maximum show considerable heterogeneity for average values of all economic variables, pointing out consistent differences among the regions in the covariates of our empirical model. Given the logarithmic specification, any null values are forced to 0.1.

Table 4: Descriptive statistics

Statistic	Mean	St. Dev.	Min	Max
Final consumption of households	112	1,021	0	39,361
Total supply	6,622	11,156	0	126,434
Total demand of households	2,573	6,294	4	75,018
Intermediate consumption	3,572	10,310	0	147,548
Road distance	644	427	0	2,210

Note: Observations=8,000. Trade flows are expressed in thousands of euros.

Finally, the dataset comprises 8000 records, that is, the number of supply regions 20 × the number of destination regions 20 × the number of the considered commodities 20. The final look is the same as expressed in Table 1.

5. Estimation results

The selection of the most suitable model specification, for estimating the Italian inter-regional bilateral trade flows, has involved several steps. In detail, to figure out if the chosen variables work properly and to eventually detect any source of bias in this baseline estimate, firstly a classical gravity model has been estimated by using an OLS regression. Then, a SAR specification of the baseline OLS regression has been adopted to get more precision in the estimation of the impact of the spatial dimension of the inter-regional trade. Finally, commodities' dummy variables have been added to take into account the relationship among commodities[17].

In Table 5 results related to the three different specifications of the model are presented. In particular: model *(1)* is the basic gravity model; model *(2)* is the spatial lag gravity model; model *(3)* is the spatial lag gravity model with control for commodities. Results reported in column 2 of Table 5 (model *(1)*) are statistically significant at 1% and their signs appear to be coherent with what is traditionally expected in the trade literature. More in detail, there is a big negative impact of road distance on the final consumption of households' bilateral trade flows (−0.76), indicating a decay of flows when distance increases, while the total supply and the demand of households present a similar positive effect, even though their magnitude is smaller with respect to the road distance. In model *(2)* (column 3) the negative sign of ρ_o and ρ_d, states that neighbors at origins or destinations influence the origin-destination flows under consideration slightly negatively[18]. The parameter ρ_w, which measures the influence of the interaction term reflecting connectivity between neighbors to the origin and neighbors to the destination, is positive,

[17] The detailed commodity labeling is provided in the appendix in Table A3.

[18] A sensitivity analysis, which led to the choice of a radius of 250 km, has been carried out with different radius (from 100 km to 300 km by 50) and # nearest neighbors (from 2 to 5) and results are available upon request.

highlighting positive spatial dependence between flows from an origin-destination pair and flows from neighbors to the origin and neighbors to the destination regions. Finally, in model *(3)* (column 4) the introduction of the commodity dummies (base="households services") generates a change in the magnitude of the total supply (0.572) and total demand of households (0.350), so the former gives now more incentive to create bilateral flows than the latter, while, the effect of the road distance is stable. Dummy variables coefficients indicate that final consumption of households trade flows are mainly facilitated, with respect to the base group ("households services"), for "mining and quarrying" products, "agriculture, forestry, and fishing" products, and "manufactured" goods, and, on the contrary, negatively impacted for services like "constructions and construction works", "wholesale and retail trade services; repair services of motor vehicles and motorcycles", "professional, scientific and technical services" or "public administration and defense services; compulsory social security services". The GOF (goodness-of-fit) of all models, expressed by R^2, is about 0.685 for *(1)* and *(2)* models and 0.731 for *(3)*, and the results recorded by AIC, Log-likelihood, and the LR test jointly suggest to prefer the spatial model *(3)*. However, the spatial LM (Lagrange multiplier) test confirms that, while we are able to reduce/eliminate the spatial dependence in Y (see column 4 of LM test - spatial lag), at the same time, this correction is not enough to deal with the spatial dependence in the error term (see column 4 of LM test - spatial error) probably due to an unobserved spatial heterogeneity on regressors, which strengthens the choice of a GWR-SAR estimator.

Results in Table 6 show that the impact of all covariates cannot be considered as homogeneous from a territorial point of view pushing to sharply reject a prediction model assuming the same patterns in the flows of inter-regional trade. See, for example, the total supply coefficient that varies across regions from 0.501 to 0.696 or the total demand of households (from 0.278 to 0.435). In Figure 3 it is also possible to appreciate the spatial distribution over the Italian regions of the estimated GWR-SAR coefficients of these variables. In particular, it can be observed a greater impact of total supply on the final consumption of household flows in the North-Eastern and Central regions than in the North-Western and Southern ones and a decreasing effect from north to south of the impact of the total demand of households. Strong differences can also be seen in the commodities coefficients as well. Moreover, enhancements in estimation are confirmed by the R^2 rising from 0.731 to 0.758 and the AIC going down from $22,260$ to $21,400$.

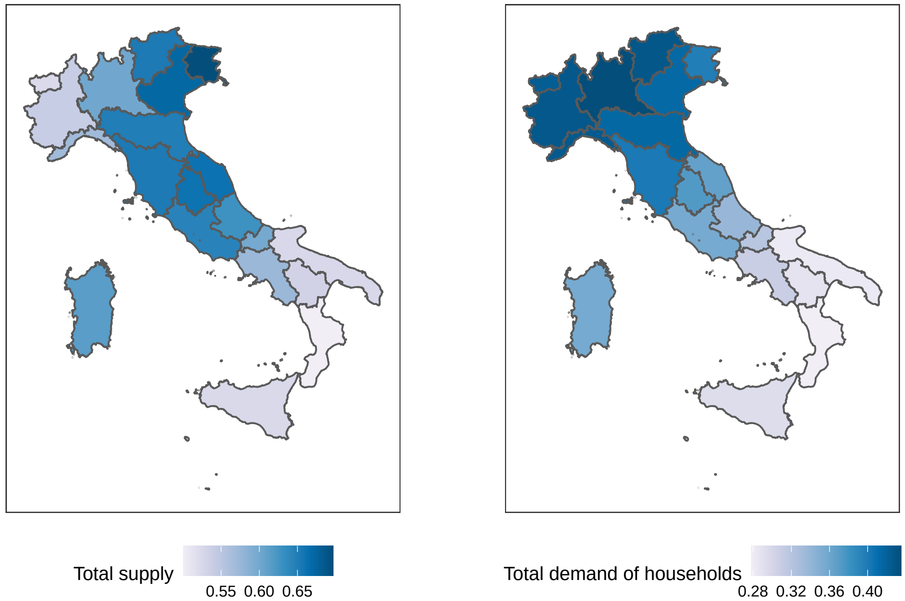


Figure 3: GWR spatial distribution of beta parameters

Finally, as discussed in Section 3 the estimated $\widehat{\beta}$ parameters are used to predict inter-regional flows for intermediate consumptions corrected by the degree of vertical integration of sectors θ and the degree of regional interaction in production C shown in Tables 7 and 8. For example, from Table 7 can be seen that a high degree of interaction exists between Lombardy and Tuscany related to "mining and quarrying" sectors or among Molise and Sicily and Sardinia with regard to "agriculture, forestry and fishing" sectors. On the contrary, a low degree of interaction can be found among Aosta Valley and some other regions in "mining and quarrying" sectors. Moreover, from Table 8, $\theta > 0$ indicates a low degree of vertical integration and, conversely, $\theta < 0$ indicates a high one.

Table 5: OLS Results

	Dependent variable:		
	Final consumption of households		
	(1)	(2)	(3)
Total supply (X_1)	0.434***	0.433***	0.572***
Total demand of households (X_2)	0.425***	0.437***	0.350***
Road distance (X_3)	−0.763***	−0.764***	−0.763***
Spatial dependence on the origin (W_oY)		−0.040***	−0.010
Spatial dependence on the destination (W_dY)		−0.017**	0.013
Spatial interdependence (W_wY)		0.036***	0.007
Agriculture			0.759***
Mining			1.612***
Manufacturing			0.448***
Electricity			0.243***
Water			0.296***
Construction			−0.666***
Trade			−0.302***
Transport			0.074
Accomodation			0.237***
IC			0.430***
Finance			0.329***
Estate			−0.031
Professional			−0.312***
Administration			0.296***
Public			−0.284***
Education			0.013
Health			−0.213***
Arts			0.324***
Others			0.347***
Constant	−0.318***	−0.235**	−1.143***
R^2	0.684	0.685	0.731
AIC	23,498	23,470	22,260
LogLik	-11,744	-11,727	-11,103
LR (Chisq)	-	34.05***	1,282***
LM test - spatial lag			
W_o	14.61***	8.97**	0.04
W_d	4.34*	1.57	3.12
W_w	3.30	0.03	2.13
LM test - spatial error			
W_o	253.57***	220.86***	218.84***
W_d	0.04	0.07	0.00
W_w	3.33	4.37*	6.93**
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 6: GWR vs OLS Results - *model (3)*

	Min	Q1	Median	Mean	Q3	Max	Global(OLS)
Total supply (X_1)	0.500	0.538	0.608	0.601	0.653	0.696	0.572
Total demand of households (X_2)	0.278	0.316	0.369	0.366	0.417	0.435	0.349
Road distance (X_3)	-0.786	-0.772	-0.758	-0.757	-0.746	-0.725	-0.763
Spatial dependence on the origin (W_oY)	-0.013	-0.012	-0.01	-0.01	-0.009	-0.007	-0.01
Spatial dependence on the destination (W_dY)	-0.005	0.000	0.012	0.011	0.019	0.033	0.013
Spatial interdependence (W_wY)	0.004	0.005	0.007	0.008	0.01	0.011	0.007
Agriculture	0.368	0.581	0.838	0.822	1.073	1.175	0.759
Mining	1.017	1.45	1.872	1.726	2.026	2.158	1.612
Manufacturing	0.293	0.356	0.386	0.423	0.434	0.654	0.449
Electricity	0.068	0.146	0.247	0.276	0.367	0.565	0.243
Water	0.048	0.150	0.299	0.323	0.488	0.593	0.296
Constructions	-0.885	-0.794	-0.724	-0.701	-0.663	-0.417	-0.665
Trade	-0.512	-0.442	-0.407	-0.308	-0.216	0.073	-0.302
Transport	-0.265	-0.183	-0.033	0.03	0.149	0.577	0.074
Accomodation	0.049	0.14	0.21	0.244	0.318	0.496	0.238
IC	0.102	0.166	0.346	0.43	0.594	1.057	0.431
Finance	-0.077	0.057	0.303	0.349	0.555	0.899	0.329
Estate	-0.178	-0.14	-0.121	-0.043	0.009	0.249	-0.030
Professional	-0.481	-0.451	-0.411	-0.337	-0.288	0.024	-0.312
Administration	0.049	0.103	0.235	0.302	0.468	0.64	0.297
Public	-0.483	-0.457	-0.348	-0.278	-0.154	0.101	-0.284
Education	-0.401	-0.309	-0.006	0.023	0.259	0.625	0.014
Health	-0.498	-0.413	-0.282	-0.214	-0.069	0.196	-0.213
Arts	0.130	0.201	0.315	0.342	0.483	0.551	0.324
Others	0.154	0.224	0.347	0.36	0.469	0.592	0.348
Constant	-2.524	-2.253	-1.796	-1.515	-0.874	0.128	-1.140
AIC = 21,400							
$R^2 = 0.758$							

Table 7: Regional interaction in production

Origin/Destination	Commodity	C value
Lombardy-Tuscany	Mining	12.294
Molise-Sicily	Agriculture	12.268
Molise-Sardinia	Agriculture	8.65
Trentino Alto Adige-Sardinia	Mining	8.601
Sicily-Sardinia	Agriculture	8.536
...	...	...
Aosta Valley-Trentino Alto Adige	Mining	1.113
Aosta Valley-Sardinia	Mining	1.099
Lombardy-Calabria	Agriculture	1.086
Aosta Valley-Abruzzo	Mining	1.072
Aosta Valley-Basilicata	Mining	1.048

Table 8: Vertical integration of sectors

Commodity	θ value
Agriculture	-0.048
Mining	0.146
Manufacturing	-0.388
Electricity	-0.224
Water	-0.048
Constructions	-0.097
Trade	0.176
Transport	-0.211
Accomodation	0.229
IC	-0.120
Finance	-0.326
Estate	0.161
Professional	-0.115
Administration	0.109
Public	0.234
Education	0.164
Health	0.020
Arts	-0.123
Others	0.218
Households	0.243

To test the robustness of the obtained results, comparisons with data from other institutional sources such as, respectively, ISTAT's annual Road Freight Transport survey[19] for the year 2016 and the study by Paniccià & Rosignoli (2018) have been carried out. Given the differences in the units of measurement[20] of our variables of interest and in the estimation approaches[21], the comparison has been made on rankings through a Spearman correlation analysis.
Results in Table 9 show that Spearman correlations[22] are very high, suggesting a reasonably good consistency of our estimated results.

[19]The Road Freight Transport survey is a sample survey under EU Regulation No 70/2012. The collected data relate to the loading and unloading of goods, their type and quantity, and the geography of the journeys made by the vehicle transporting them. For each trip, up to three types of goods are collected from the vehicle according to the NST/2007 classification for transport statistics. The recorded trips can be both national and international (in particular Italy-foreign and foreign-Italy flows).

[20]The road freight phenomenon is measured in tonnes and tonne-kilometers; the former measures the quantity of goods transported, the latter the performance level of the transport service.

[21]In Paniccià & Rosignoli (2018) not all commodities are estimated through an econometric model but ad-hoc analyses are made (mining, public administration, construction, real estate services, and energy)

[22]The Spearman rank correlation index is a non-parametric statistical measure of rank correlation (statistical dependence between the rankings of two variables). It varies between -1 and 1, where -1 indicates the maximum negative correlation and 1 indicates the maximum positive correlation.

Table 9: Comparison with RFT ISTAT and Paniccià & Rosignoli (2018) inter-regional trade flow estimation - Spearman correlation

	Commodity	Final cons.	Intermediate cons.	Total cons.
ISTAT	Agriculture	0.686	0.729	0.72
	Mining	0.738	0.688	0.737
	Manufacturing	0.714	0.854	0.848
IRPET	Agriculture	0.550	0.476	0.510
	Mining	0.662	0.664	0.672
	Manufacturing	0.751	0.795	0.802

Finally, the obtained inter-regional flows can be represented with a rope graph or through geographical maps. These graphs allow an immediate visualization of the major trade flows between regions in the sectors examined. As an example, the "manufacturing" sector will be analyzed in detail. Figure A.4 in Appendix shows the trade flows between regions of manufacturing goods relative to final household consumptions. In particular, the largest flows originate in Lombardy (grey ropes) which sells more to Piedmont, Lazio, and Emilia Romagna; Lazio (red ropes) which sells more to Campania, Tuscany, and Piedmont; Veneto (brown ropes) which sells to Lombardy, Emilia Romagna, and Piedmont. Considering instead the intermediate manufacturing goods, Figure A.5 in Appendix shows that Lombardy (grey ropes) remains the region with the most outflows, Veneto (brown ropes) has an increase in flows, on the contrary, Lazio (red ropes) has much less flows and Emilia Romagna (orange ropes) becomes the second region from which these flows originate.

6. Final remarks

This paper presents a novel framework to estimate households' final consumption and intermediate consumption bilateral trade flows among Italian regions. The approach consists of starting from administrative data on VAT returns for reconstructing the final consumption trade flows. Hence, the lack of data for inter-regional trade has been addressed by exploiting the availability of Italian VAT returns. Then, for the estimation of the determinants of final consumption, we control for spatial dependence (origin, destination, and origin-destination linkages) and commodities relationship by relying on a gravity model approach. This model is applied to a novel pseudo-panel, where the sectors are treated as time dimensions and included in the model as fixed effects. Once the determinants of final consumption are investigated, we estimate intermediate consumption bilateral trade flows, adjusting for regional interaction in production and vertical integration of sectors. Spearman correlations with ISTAT's annual Road Freight Transport survey and Paniccià & Rosignoli (2018) study were found to be very high, suggesting a reasonably good consistency of our estimation.

The bilateral trade flows estimation can be useful for several kinds of analyses. For instance, when dealing with computable general equilibrium (CGE) models based on a regionalized Social Accounting Matrix, data about inter-regional flows between regions has to be taken into account for a proper description of the economic system.

References

Anderson, James E. 2011. The gravity model. *Annual review of economics*, **3**(1), 133–160.

Barbosa, Hugo, Barthelemy, Marc, Ghoshal, Gourab, James, Charlotte R., Lenormand, Maxime, Louail, Thomas, Menezes, Ronaldo, Ramasco, José J., Simini, Filippo, & Tomasini, Marcello. 2018. Human mobility: Models and applications. *Physics reports*, **734**, 1–74. Human mobility: Models and applications.

Beine, Michel, Bertoli, Simone, & Fernández-Huertas Moraga, Jesús. 2016. A practitioners' guide to gravity models of international migration. *The world economy*, **39**(4), 496–512.

Bentivogli, Chiara, Ferraresi, Tommaso, Monti, Paola, Paniccià, Renato, & Rosignoli, Stefano. 2019. Italian regions in global value chains: An input-output approach. *Politica economica*, 55–94.

Cirillo, Cristina, Imperioli, Lucia, Manzo, Marco, *et al.* 2021. *The value added tax simulation model: Vatsim-df (ii)*. Tech. rept.

Disdier, Anne-Célia, & Head, Keith. 2008. The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The review of economics and statistics*, **90**(1), 37–48.

Doucet, Romain, Margaretic, Paula, Thomas-Agnan, Christine, & Villotta, Quentin. 2014 (Mar.). *Spatial dependence in (origin-destination) air passenger flows*. TSE Working Papers 14-494. Toulouse School of Economics (TSE).

Flegg, Anthony T, & Tohmo, Timo. 2013. Regional input–output tables and the flq formula: a case study of finland. *Regional studies*, **47**(5), 703–721.

Geniaux, Ghislain, & Martinetti, Davide. 2018. A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional science and urban economics*, **72**(C), 74–85.

Grosche, Tobias, Rothlauf, Franz, & Heinzl, Armin. 2007. Gravity models for airline passenger volume estimation. *Journal of air transport management*, **13**(4), 175–183.

Ivanova, Olga, d Artis Kancs, & Stelder, Dirk. 2009. *Modelling inter-regional trade flows: Data and methodological issues in rhomolo*. EERI Research Paper Series 31/2009. Brussels.

Jaya, I. Gede Nyoman Mindra, Ruchjana, Budi Nurani, Tantular, Bertho, Zulhanif, & Andriyana, Yudhie. 2018. Simulation and application of the spatial autoregressive geographically weighted regression model (sar-gwr) (sar-gwr).

Johansen, Leif. 1960. *A multi-sector study of economic growth*. Vol. 21. North-Holland Publishing Company.

Leamer, Edward E., & Levinsohn, James. 1994 (Nov.). *International Trade Theory: The Evidence*. NBER Working Papers 4940. National Bureau of Economic Research, Inc.

Lecca, Patrizio, Barbero, Javier, Aaron, Christensen, Conte, Andrea, Francesco, Di, Jorge, Diaz-Lanchas, Olga, Diukanova, Mandras, Giovanni, Persyn, Damiaan, & Sakkas, Stelios. 2018 (05). *Rhomolo v3: A spatial modelling framework*. Tech. rept.

LeSage, James P., & Pace, R. Kelley. 2008. Spatial econometric modeling of origin-destination flows*. *Journal of regional science*, **48**(5), 941–967.

LeSage, James P., & Thomas-Agnan, Christine. 2015. Interpreting spatial econometric origin-destination flow models. *Journal of regional science*, **55**(2), 188–208.

Lesage, James Paul, & Polasek, Wolfgang. 2008. Incorporating transportation network structure in spatial econometric models of commodity flows. *Spatial economic analysis*, **3**(2), 225–245.

Lewer, Joshua J., & Van den Berg, Hendrik. 2008. A gravity model of immigration. *Economics letters*, **99**(1), 164–167.

Liu, Weidong, Li, Xin, Liu, Hongguang, Tang, Zhipeng, & Guan, Dabo. 2015. Estimating inter-regional trade flows in china: A sector-specific statistical model. *J. geogr. sci*, **25**, 1247–1263.

Marrocu, Emanuela, & Paci, Raffaele. 2013. Different tourists to different destinations. evidence from spatial interaction models. *Tourism management*, **39**, 71–83.

Morley, Clive, Rosselló, Jaume, & Santana-Gallego, Maria. 2014. Gravity models for tourism demand: theory and use. *Annals of tourism research*, **48**, 1–10.

Paniccià, Renato, & Rosignoli, Stefano. 2018. *A methodology for building multiregional supply and use tables for italy*. Tech. rept. IRPET - STUDI E APPROFONDIMENTI.

Pulliainen, K. 1963. A world trade study: An econometric model of the pattern of the commodity flows in international trade in 1948-1960. *Ekonomiska samfundets tidskrift*, **16**, 69–77.

Pöyhönen, Pentti. 1963. A tentative model for the volume of trade between countries. *Weltwirtschaftliches archiv*, **90**, 93–100.

Santeramo, Fabio Gaetano, & Morelli, Mariangela. 2016. Modelling tourism flows through gravity models: a quantile regression approach. *Current issues in tourism*, **19**(11), 1077–1083.

Sargento, Ana L.M. 2007. Empirical examination of the gravity model in two different contexts: Estimation and explanation. *Jahrbuch für regionalwissenscha*, **27**(4), 103–127.

Sargento, Ana L.M., Ramos, Pedro Nogueira, & Hewings, Geoffrey J.D. 2012. Inter-regional trade flow estimation through non-survey models: an empirical assessment. *Economic systems research*, **24**(2), 173–193.

Sawyer, John A. 1967. An econometric study of international trade flows. by hans linnemann. amsterdam: North-holland publishing company. 1966. pp. xiv, 234. *Canadian journal of economics and political science*, **33**(4), 633–634.

Song, H, Dwyer, L, Li, Gang, & Cao, Zheng. 2012. Tourism economics research: A review and assessment. *Annals of tourism research*, **39**(3), 1653 – 1682.

Standardi, Gabriele, Bosello, Francesco, & Eboli, Fabio. 2014 (Jan.). *A Sub-National CGE Model for Italy*. Feem working paper no. 4.2014, cmcc research paper no. rp0189. Fondazione ENI (Enrico Mattei).

Thissen, Mark, Van Oort, Frank, Diodato, Dario, & Ruijs, Arjan. 2013. *Regional competitiveness and smart specialization in europe: Place-based development in international economic networks*. Edward Elgar Publishing.

Thissen, Mark, Ivanova, Olga, Mandras, Giovanni, & Husby, Trond. 2019 (Jan.). *European NUTS 2 regions: construction of interregional trade-linked Supply and Use tables with consistent transport flows*. JRC Working Papers on Territorial Modelling and Analysis 2019-01. Joint Research Centre (Seville site).

Tinbergen, Jan. 1962. *Shaping the world economy; suggestions for an international economic policy*. T.

Tomal, Mateusz. 2020. Modelling housing rents using spatial autoregressive geographically weighted regression: A case study in cracow, poland. *Isprs international journal of geo-information*, **9**(6).

Zhang, Yahua, Lin, Faqin, & Zhang, Anming. 2018. *Gravity models in air transport research: a survey and an application*. Cheltenham, UK: Edward Elgar Publishing.

Appendix A. Appendix

Table A1: Italian NUTS2 regions

o,d subscript	Label
1	Abruzzo
2	Aosta Valley
3	Apulia
4	Basilicata
5	Calabria
6	Campania
7	Emilia-Romagna
8	Friuli-Venezia Giulia
9	Lazio
10	Liguria
11	Lombardy
12	Molise
13	Piedmont
14	Sardinia
15	Sicily
16	The Marches
17	Trentino-Alto Adige
18	Tuscany
19	Umbria
20	Veneto

Table A2: NACE rev.2 Level 1

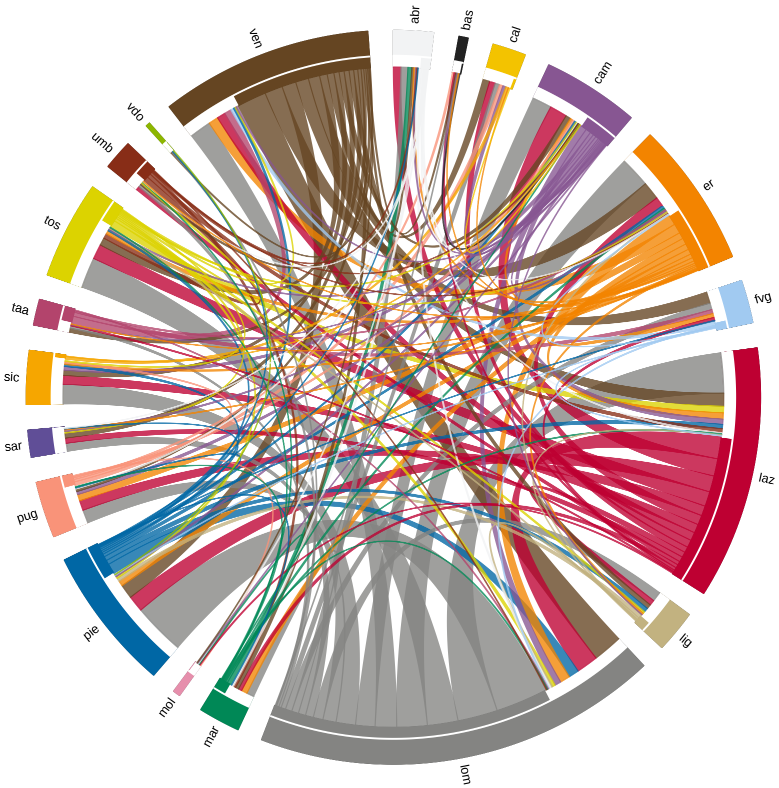
s subscript	Label
1	Agriculture, forestry and fishing
2	Mining and quarrying
3	Manufacturing
4	Electricity, gas, steam, and air conditioning supply
5	Water supply; sewerage, waste management, and remediation activities
6	Construction
7	Wholesale and retail trade; repair of motor vehicles and motorcycles
8	Transportation and storage
9	Accommodation and food service activities
10	Information and communication
11	Financial and insurance activities
12	Real estate activities
13	Professional, scientific, and technical activities
14	Administrative and support service activities
15	Public administration and defence; compulsory social security
16	Education
17	Human health and social work activities
18	Arts, entertainment, and recreation
19	Other service activities
20	Activities of households as employers; u0ndifferentiated goods- and services-producing activities of households for own use

Table A3: CPA 20 labels

i subscript	Label	Short label
1	Products of agriculture, forestry and fishing	Agriculture
2	Mining and quarrying	Mining
3	Manufactured products	Manufacturing
4	Electricity, gas, steam, and air conditioning	Electricity
5	Water supply; sewerage, waste management and remediation services	Water
6	Constructions and construction works	Constructions
7	Wholesale and retail trade services; repair services of motor vehicles and motorcycles	Trade
8	Transportation and storage services	Transport
9	Accommodation and food services	Accommodation
10	Information and communication services	IC
11	Financial and insurance services	Finance
12	Real estate services	Estate
13	Professional, scientific and technical services	Professional
14	Administrative and support services	Administration
15	Public administration and defence services; compulsory social security services	Public
16	Education services	Education
17	Human health and social work services	Health
18	Arts, entertainment and recreation services	Arts
19	Other services	Others
20	Services of households as employers; undifferentiated goods and services produced by households for own use	Households

Figure A.4: Manufactured products B2C trade flows

(a) Circle chart



(b) Map

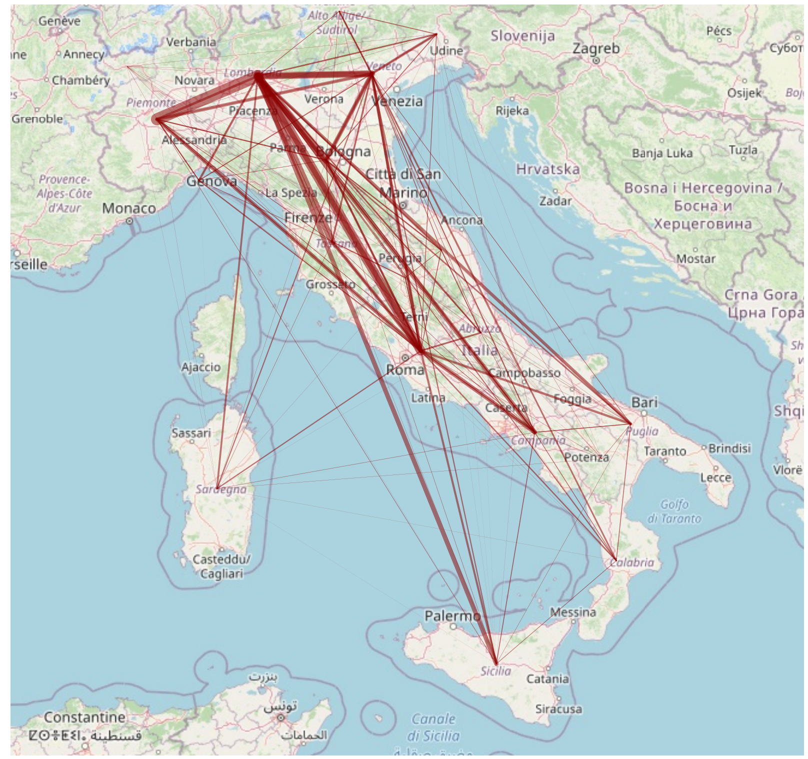
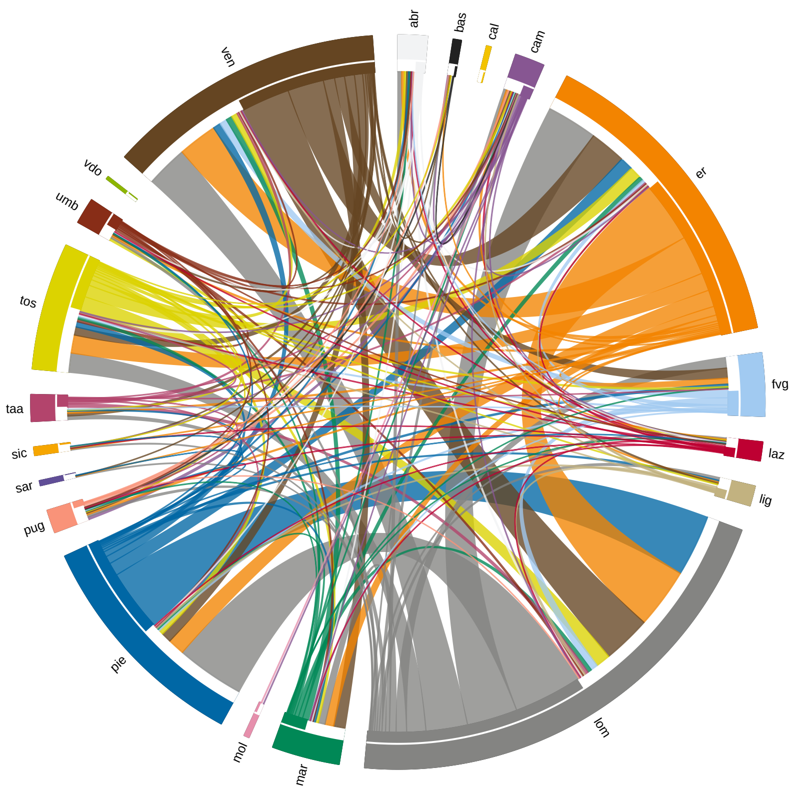


Figure A.5: Manufactured products B2B trade flows

(a) Circle chart



(b) Map