Intraregional Trade Shares for Goods-Producing Industries: RPC Estimates using EU Data

Citation to Publisher Version: Lahr, Michael L., Ferreira, João Pedro & Többen, Johannes. (2020). Intraregional Trade Shares for Goods-Producing Industries: RPC Estimates using EU Data. *Papers in Regional Science ()*.

Citation to *this* Version: Lahr, Michael L., Ferreira, João Pedro & Többen, Johannes. (2020). Intraregional Trade Shares for Goods-Producing Industries: RPC Estimates using EU Data. *Papers in Regional Science ()*. Retrieved from http://dx.doi.org/doi:10.7282/t3-27sb-ra31.

INTRAREGIONAL TRADE SHARES FOR GOODS-PRODUCING INDUSTRIES: RPC ESTIMATES USING EU DATA*

Michael L. Lahr,[a] João Pedro Ferreira,[b] and Johannes Többen[c]

ABSTRACT. The lack of subnational trade data has dampened the development of reliable regional and multiregional models for regional policy development. So, most researchers and vendors of regional and interregional economic models continue to rely on location quotients, supply-demand pool techniques, or minor modifications of them, despite knowing that they under-estimate interregional trade. In this piece, we analyze the relative viability of estimates of intraregional trade—so called "regional purchase coefficients" (RPCs). We do so for manufacturing sectors in 28 EU countries using the World Input-Output Database. We introduce an RPC-estimating technique using a quasi-binomial regression approach for goods-producing industries; we apply standard supply/demand ratios as RPCs for service-based industries. We then apply the estimates to an aggregate EU input-output (I-O) table and measure how closely the results approximate the I-O tables (direct requirements matrices) for each of the 28 EU nations. We compare these findings to those obtained by other conventional approaches. We also evaluate their ability to replicate the country Leontief inverses and output multipliers. We find quasi-binomial regression approaches superior across the board.

1. Introduction

Constructing accurate (subnational) regional and multiregional accounts remains a focus of regional economists. This is because the extent of interindustry interactions and trade are of greater importance to more-open economies since, over time, they enable endogenous growth via Hirschman (1958) linkages. Nonsurvey approaches, often enhanced with superior data, the so-called "hybrid approach" (c.f. Lahr 1993, 2001a), remain the prime means of developing regional accounts. This is a result of the high costs in terms of both time and money of survey-based approaches. Nonsurvey regional input-output (I-O) accounts are constructed by departing from a given, pertinent national technology (national direct requirements matrix) and, then, adjusting it for prevailing trade and productivity differences (Lahr, 2001b, Sargento, 2009). That is, analysts assume that technology is spatially constant (i.e., knowledge is freely transmitted within a nation). A focus of this paper is modeling the magnitude of interregional trade—both regional gross outflows and inflows, not net trade; we leave the measurement of interregional productivity differentials (and concordant adjustments of regional accounts to them) for others to pursue, at least for the time being.

---

* We thank Ronald E. Miller for his casting his thoughts and eagle eye across our final draft. We thank both him and Geoffrey J.D. Hewings for their encouragement in this venture.

[a] Rutgers Economic Advisory Service (R/ECON™), Edward J. Bloustein School of Planning & Public Policy, Rutgers, The State University of New Jersey, **contact author**: lahr@rutgers.edu.

[b] Food and Resource Economics Department, University of Florida.

[c] Energy and Climate Group, Gesellschaft Für Wirtschaftliche Strukturforschung, Osnabrück, Germany.

As early as Stevens and Trainer (1976) researchers knew that simple regionalization approaches—typically some form of location quotient (LQ)—in raw form were problematic.[1] In particular, LQs do not permit cross-hauling. Treyz and Stevens (1981) note that cross-hauling is the rule rather than the exception in most subnational regional economies. So, unless transportation costs suddenly skyrocket (i.e. fuel prices surge), it is unlikely that the phenomenon of cross-hauling will go away anytime soon. Moreover, Polenske and Hewings (2004) suggest that increasing fragmentation of production and rising demand for product variety spark trade. This suggests a rising need for understanding the character of cross-hauling and, hence, trade. Indeed, a rather long list of researchers note that ignoring cross-hauling leads to significantly biased estimates of subnational trade (Round, 1983; McCann and Dewhurst, 1998; Lehtonen and Tykkyläinen, 2014; Boero et al., 2018).

We highlight cross-hauling since many heavily used regionalization approaches—in fact those most commonly used—ignore it. Among the various approaches that are most guilty on this count are truncated location quotients (LQ).[2] Employment LQs, income LQs, and, supply LQs (SLQs) are essentially rough proxies for the supply/demand ratio (SDR), so we also include the SDR as part of the "LQ family." Miller and Blair (2009), a widely recognized resource on I-O analysis, devotes more than ten pages in Section 8.2 of their book to such approaches; so, we do not do so here.

Flegg, Webber, and Elliott (1995), Flegg and Tomho (2013, 2016, 2018a, 2018b), Flegg, Mastronardi and Romero, 2016), Jahn (2017), and Kowalewksi (2015) have examined slightly more complex parametric transforms of LQs to estimate regional supply percentages. Like LQ approaches, when applied row-wise to a region's technology matrix, these so-called "FLQs" estimate a region's direct requirements. FLQ approaches also have been criticized (Lamonica and Chelli, 2018; Fujimoto, 2018) with respect to (1) their value given that interregional trade is known in the instances published and (2) the lack of meaningful economic content of some of the adjustments that they apply to the LQs.

Another approach that explicitly attempts to account for cross-hauling was the appropriately named "Cross-Hauling Adjusted Regionalization Method" (CHARM) and originally developed by Kronenburg (2009). Related to work by Grubel and Lloyd (1971) on intra-industry trade, the latest versions of CHARM employ rather complex formulae, which we will associate with the industry's RPC in a later section. CHARM variants relate a region's output by industry to its demand by industry via notions of "excess demand" and "excess output". But it is not much

[1] Stevens, Treyz, and Lahr (1989) note that cutting LQ estimates of RPCs to a third of their value would be better than the LQ itself as a proxy for RPCs in state-based U.S. manufacturing sectors.

[2] By "truncated," we merely refer to any large LQ, those with a value greater than 1, being set to 1.

more data-demanding than the supply/demand ratio (SDR), although its formula is much more complex as we show.

Perhaps the most theoretically grounded approach, which accounts for cross-hauling and is also mentioned briefly by Miller and Blair (2009, Section 8.2.7), is that first articulated by Stevens and Trainer (1976). This approach recognizes that the macroeconomic "cross-hauling problem"[3] (see, e.g., Begg and Isserman, 1986; Robison and Miller, 1988) is motivated by arbitrariness of subnational jurisdictional boundaries as well as by characteristics of industrial organization (monopolistic competition and oligopoly), actual differences in commodities traded by the industry (Kronenberg, 2009), and business location propensities. The work on this approach parallels that in international trade theory, albeit neglecting such matter as exchange rates and trade agreements, as detailed in a later section. It is similarly grounded in gravity models and statistics. Stevens and Trainer (1976) coined the term "regional purchase coefficient" when referring to the extent to which regional producers fulfill the same region's demand for a commodity and termed it a "regional purchase coefficient" (RPC). We stick with this terminology, despite a greater degree of logic to using Miller and Blair's (2009) "regional supply percentages" for the concept, since the term RPC has deeply permeated the extant literature.

Please note our focus in this is paper is rows-only approaches to trade-adjusting I-O direct requirements matrices. In applying such a focus, we are by no means disregarding Garhart and Giarratani's (1987) request for a full matrix of RPCs, but rather bypassing the request, at least for now, as an impracticality. Answering it requires at least some RAS-like data raking (Lahr and De Mesnard, 2004; Miller and Blair, 2009, Section 8.3), which requires not just n but rather at least n^2 pieces of information on industry product inflows, where n is the number of industries being measured in the economy. In most countries, even the n pieces of information must be estimated at the subnational level. In this vein, we do not compare regionalization approaches predicated upon RPC-estimates to those that employ RAS.[4] We recognize that the more real information is applied to the regionalization problem, the better the resulting I-O tables should be. Generally

[3] A commodity is cross-hauled when it is both imported into and exported by an economy. For example, the United States both imports and exports crude oil (in 2018, respectively, \$163.1 billion and \$47.2 billion). In this instance, industrial organization (differences in the country of origin of the firms that own oilfields) plays a major role. Meanwhile, the State of Idaho both receives inflows and ships outflows of potatoes. This likely has less to do with the oligopolies "counterproductively" trading their commodities across borders or even with different types of potatoes (those for baking versus boiling) being shipped but rather with the simple fact that Idaho's panhandle, which produces almost no potatoes, consumes potatoes produced in nearby eastern Washington State while southern Idaho, which is spatially distant from the panhandle, largely satisfies its own potato needs.

[4] We also do not analyze cross-industry location quotients (CILQs) either, but simply based on the principle that they are not strictly rows-only techniques.

speaking, as Lahr (2001a) notes, information on regional inflows by using industry are most difficult to find, precluding RAS a regionalization option in most cases.[5]

In essence there are four families of rows-only estimates: LQs, FLQs, CHARM, and econometric approaches. Of these, the LQ has the lowest data requirement—just needing data on employment, labor compensation, or production by industry. Both FLQs and econometric approaches require some knowledge of interregional trade relationships for a nation *a priori*. From those generalized relationships, researchers are able to infer the magnitude of intraregional trade for a specific industry in their region. While CHARM does not require any information on interregional trade, it does require complete information on regional supply and demand, which can make it more demanding than inference via FLQ relationships. From a data perspective, econometric approaches can be the most demanding, but they need not be any more demanding than FLQ approaches as we show later.

We understand that simple parametric transforms of LQs are used and will continue to be used. A main point of this piece is to suggest that, when intra-regional trade by industry is known or can be calculated from interregional trade flows, statistical techniques and theory approaching that used to measure international trade should be applied. That is, models of trade characteristics should be grounded in location theory. In this vein, we do our best here to re-ignite explorations into the determinants of subnational interregional trade in the development of RPCs. The issue of what approach should be used is almost strictly a matter of data availability. It may well be, of course, that in some regions of the world a location quotient of some sort is the best that can be done, at least if sharing out intranational trade across a multiregional framework is not so easy (see, e.g. Haddad and Hewings, 2005; Haddad, 2014; Elshahawany, Haddad, and Lahr, 2017).

We explore the comparative accuracy of these various approaches. We do so by exploiting one of the many international multiregional I-O (MRIO) tables available—the World Input-Output Database (WIOD). We first characterize EU manufacturing RPC by country and industry within the EU—the focus of our work. We then compare the various approaches to estimating these RPCs, acknowledging the extent of data requirements for each: LQs, simple parametric LQ transforms, CHARM, and a quasi-binomial statistical estimate in the spirit of Treyz and Stevens (1985). We compare the approaches via estimated "regionalized" direct requirements matrices and corresponding Leontief inverses and total output multipliers by country to the corresponding actual country tables in WIOD. Our hypothesis in these comparisons is very roughly similar to that employed by Szyrmer (1989) who sequentially added more information when reproducing a

[5] Regional value-added (not including imports or inflows) by industry is sometimes available, however, so RAS can be helpful in estimating regional technology matrices if output estimates are possible, enabling the derivation of intermediate industry demand as suggested by Lahr (2001b).

national table. That is, we expect approaches that use more information in estimating RPCs should yield better results. In this vein, we expect our quasi-binomial-based estimate of RPCs to perform best and we expect conventional employment LQ approaches to be the worst. Other approaches should fall somewhere in between.

2. Regional Purchase Coefficients in WIOD

Of the international I-O data bases available we opted to use WIOD. We made this selection because it is our understanding that this research team (Timmer et al., 2015) focused its effort on balancing trade between origin and destination countries. This is undoubtedly the same reason that Lamonica and Chelli (2018) used it. The 2016 version of WIOD includes more countries and more industries for 2000-2014. Rather than use all of the countries in the database as Lamonica and Chelli did, we focus on just the 28 countries of the European Union (EU) to estimate the RPCs for each goods-producing industry by country.

To make a realistic sort of experiment, we apply the RPCs by industry that result from each approach examined to an I-O table for the EU—an aggregate of all 28 national tables into a single "national" account. This EU aggregate proxies as the national I-O matrix for our tests. Our thought is that the meta-region of analysis should be as much like a unified country, at least tradewise, as it can possibly be for an international trade database.[6] The EU is exactly that. For this set of 28 countries we first estimate RPCs (ρ) for goods-producing industries by country for the year 2014:

$$\rho_i = \sigma_i / d_i = \left(q_i \ell_i\right) / d_i \tag{1}$$

where,

$$0 \leq \rho_i \leq 1$$

For any given region (EU country here), σ_i is the local supply of the good or service *i* by the region to itself, d_i is the demand for *i*, q_i is the corresponding total supply or output of *i* in the region and ℓ_i is the share of total output *i* shipped to destinations within the region. As ρ_i approaches 1, regional production satisfies almost all of the local demand. As ρ_i approaches 0 local demand is mostly fulfilled by supplies from elsewhere. Alternatively, Miller and Blair (2009) represent it as:

$$\rho_i = \frac{q_i - x_i}{q_i - x_i + m_i} \tag{2}$$

[6] Lamonica and Chelli's (2018) probed further into the δ value in the flurry of work by A.T. Flegg and colleagues. They did so by selecting the broadest possible geographical scope of the WIOD database. In so doing they analyze the comparative value of LQs for estimating international trade, rather than for subnational trade as FLQs are intended. This is our sole criticism of the Lamonica and Chelli piece.

where x_i and m_i are regional outflows and inflows of commodity *i*, respectively, albeit with slightly modified notation.

Figure 1 shows the distribution of RPCs, via box and whisker plots, for goods-producing industries by EU country.[7] Figure 2 specifically focuses on services (including electricity and construction services). As it is possible to see in Figure 1, the countries are ordered from left to right by lowest to median RPC in the 23 goods-producing industries. The first three countries have an (unweighted) average RPC value of less than 0.10 and also, not surprisingly, are among the smallest EU countries (Cyprus, Luxembourg and Ireland). Contrarily, Spain, Romania, and Italy reportedly have median RPCs higher than 0.60. Moreover, Italy's case is extreme in that all WIOD manufacturing industries have an RPC higher than 0.24, which is higher than the median value of seven EU countries. Alternatively, Luxembourg's least internally demand-driven goods-producing industry, "Manufacturing other non-metallic mineral products," has an RPC of 0.63, which is lower than the median RPC in Italy or Spain.

From a joint assessment of Figures 1 and 2, it should be immediately clear that, compared to the demands for goods, service demands are more likely to be satisfied by local suppliers. Without exception, the median RPC for good-production is lower than that for services in a given country. Indeed, only three countries have median RPCs for services below 0.80—again, Luxembourg, Ireland, and Malta. There is more apparent variation by sector, however: "Air Transport" (0.02), "Publishing activities" (0.09) or "Motion picture, video and television program production, sound recording and music publishing activities; programming and broadcasting activities" (0.09). Indeed, both Luxembourg (0.34) and Ireland (0.44) have two of the lowest RPC's in the "Financial service activities, except insurance and pension funding" while Ireland is the extreme case for "Telecommunications (0.49).

[7] Box-and-whisker plots depict the minimum value (the lowest dot, if it is not also part of the lower whisker), first quartile (the lower whisker), mean (the line in the box), third quartile (the upper whisker) and maximum value (the uppermost dot, if it is not also part of the upper whisker). In later figures, outliers in addition to the maximum and minimum values are the other dots that are beyond the whiskers, and the median is denoted by the × in the box.

Figure 1: EU RPC's in agriculture and manufacturing, 2014

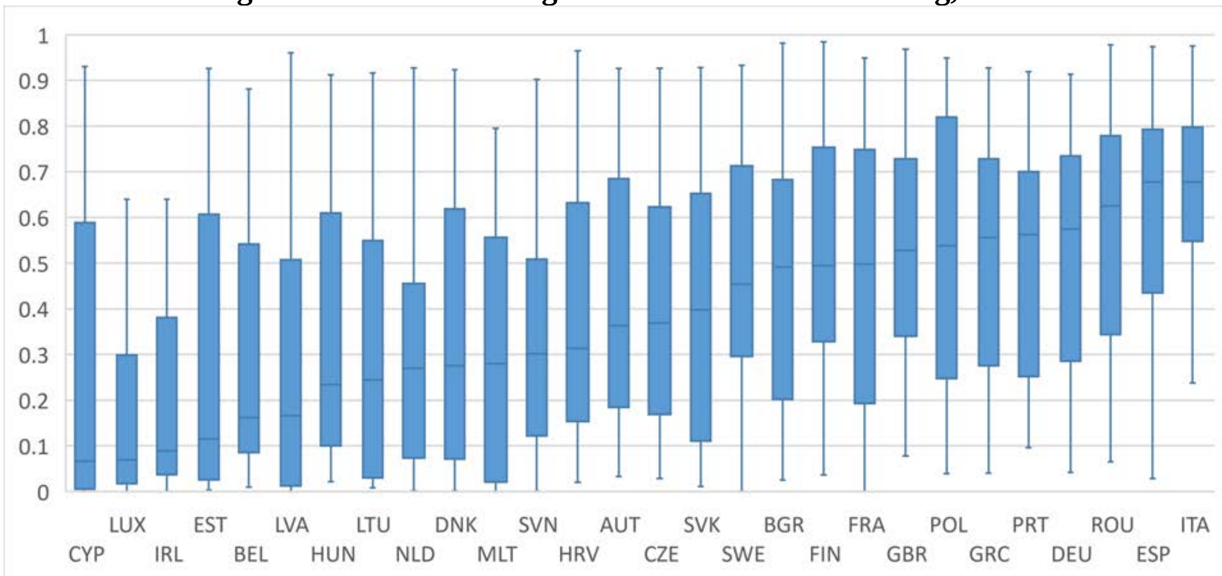


Figure 2: EU RPCs in services, 2014

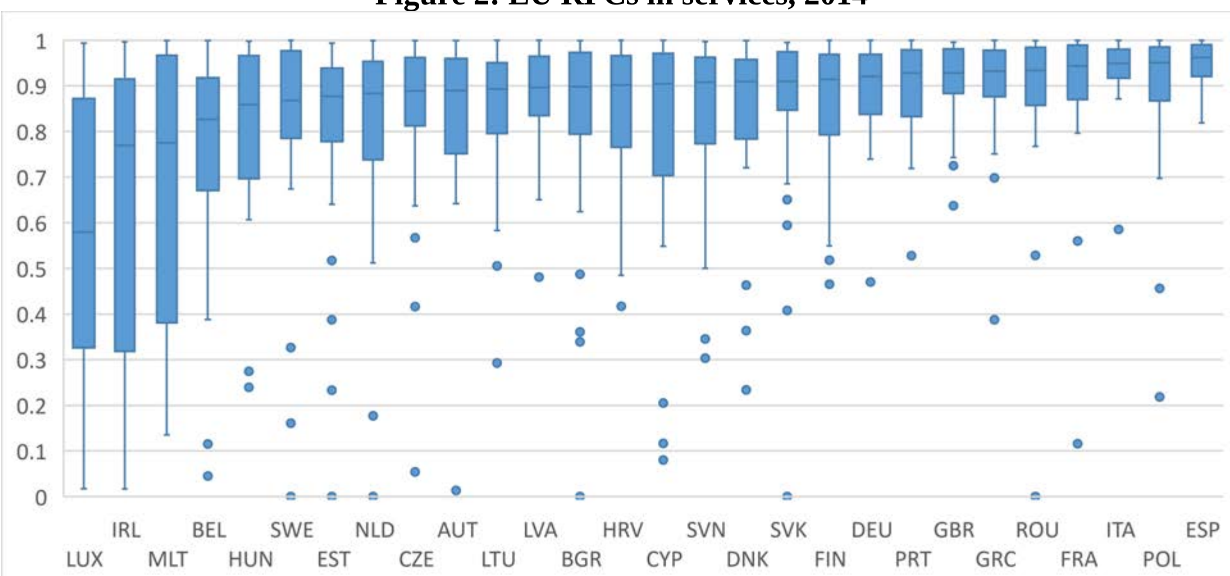


Table 1: RPC's in goods-producing industries across counties in the EU in 2014

	Code	Output ($ thousand)	RPCs	
			Average	Std. Dev.
Crop and animal production, hunting and related service activities	A01	557,404	0.72	0.15
Forestry and logging	A02	52,705	0.78	0.26
Fishing and aquaculture	A03	18,246	0.45	0.31
Mining and quarrying	B	203,440	0.27	0.23
Manufacture of food products, beverages and tobacco products	C10-C12	1,422,996	0.61	0.18
Manufacture of textiles, wearing apparel and leather products	C13-C15	272,266	0.19	0.21
Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	C16	160,205	0.61	0.19
Manufacture of paper and paper products	C17	225,067	0.43	0.25
Printing and reproduction of recorded media	C18	119,001	0.85	0.19
Manufacture of coke and refined petroleum products	C19	595,117	0.38	0.28

Manufacture of chemicals and chemical products	C20	681,020	0.22	0.18
Manufacture of basic pharmaceutical products and pharmaceutical preparations	C21	297,666	0.21	0.15
Manufacture of rubber and plastic products	C22	364,604	0.38	0.20
Manufacture of other non-metallic mineral products	C23	267,191	0.65	0.13
Manufacture of basic metals	C24	484,703	0.27	0.23
Manufacture of fabricated metal products, except machinery and equipment	C25	614,847	0.57	0.19
Manufacture of computer, electronic and optical products	C26	365,499	0.12	0.13
Manufacture of electrical equipment	C27	354,298	0.18	0.15
Manufacture of machinery and equipment n.e.c.	C28	803,512	0.25	0.20
Manufacture of motor vehicles, trailers and semi-trailers	C29	986,997	0.21	0.20
Manufacture of other transport equipment	C30	271,691	0.28	0.23
Manufacture of furniture; other manufacturing	C31-C32	276,604	0.37	0.20

RPCs also help give some perspective on the EU's sectoral behavior in trade. Table 1 shows the total gross output for the 28 EU countries, by industry and the weighted average RPC and its standard deviation by industry. According to Table 1, a good degree of heterogeneity is also consistent among goods-producing sectors. Among the commodities that have a greater propensity to be traded we find the "Manufacture of computer, electronic and optical products", the "Manufacture of electrical equipment" and the production of "Textiles, wearing apparel and leather products". In the EU, at least, national production tends to satisfy less than 20% of the local demand for these commodities. Contrarily, it satisfied 70% or more of national demands in "Printing and reproduction of recorded media", "Forestry and logging" and Agriculture products. It is expected that "Fishing" would have a higher-than-average standard deviation since a number of EU countries lack a coast, and, hence, must import this product.

3. RPC estimates

We combine information from the WIOD database with data from other official statistical sources[8] to identify determinants of RPCs for every goods-producing industry in each of the 28 EU nations. We start simply and move to more complex approaches. We begin with estimates generated by truncated employment location quotients (LQs) and supply/demand ratios (SDRs)—basic fallback approaches. We also test a representative FLQ. We next regress the LQs and examine them in tandem with FLQ estimates. We follow by examining the relative merits of CHARM-based estimates of trade shares (Többen and Kronenberg, 2015; Fujimoto, 2015). We then apply a quasi-binomial form informed in an economic-theoretic sense by Treyz and Stevens (1985). We then compare outcomes from all three of these approaches.

[8] Our database includes variables from WIOD Socio-Economic Accounts, or the EUROSTAT or OECD databases.

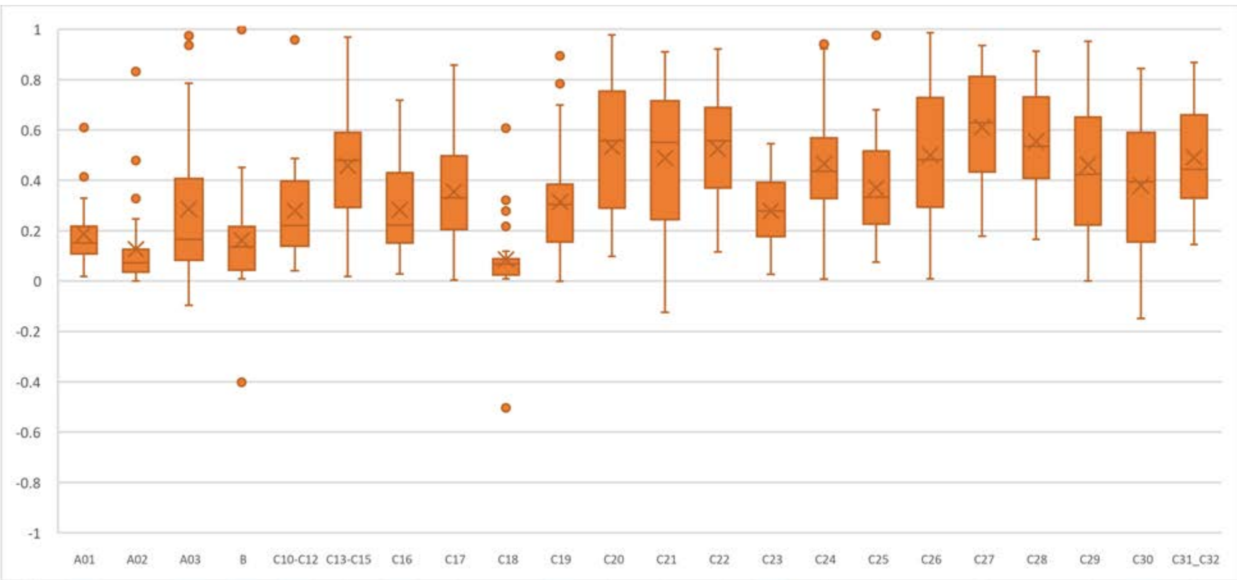
3.1. Relative bias in short-cut estimates of RPCs

We next compare RPC estimations. We examine those obtained with the quasi-binomial linear approach and compare them to those obtained via LQ and SDR approaches as well as from FLQs, admittedly raw versions of those used in research following Flegg, Webber, and Elliott (1995).[9]

3.1.1. SDRs and LQs

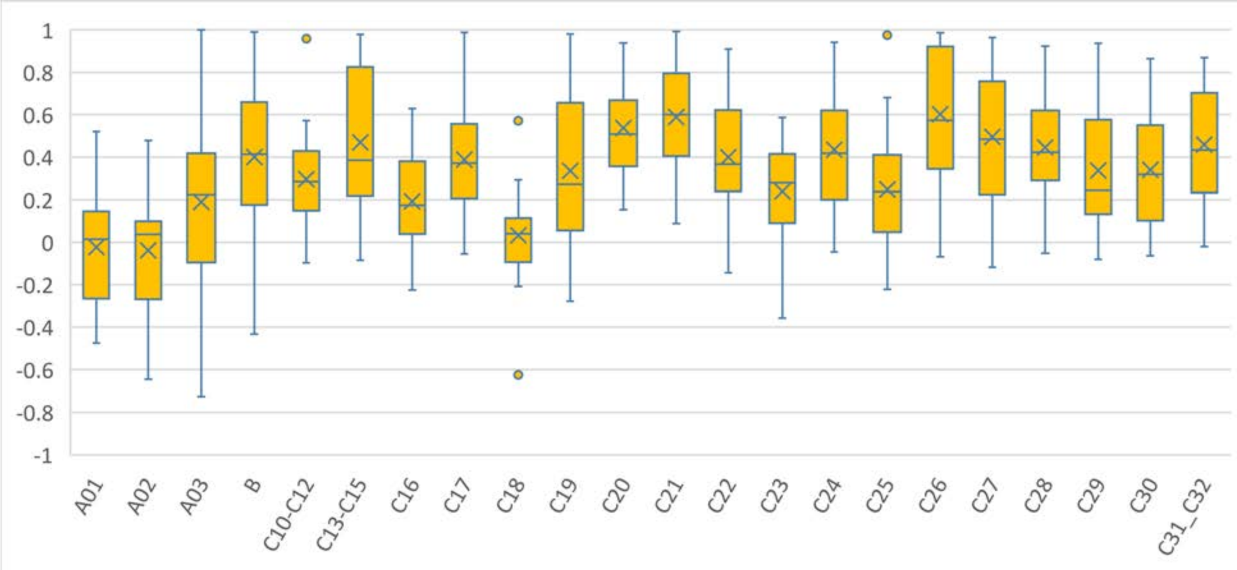
Figure 3a and 3b present box and whisker plots for SDRs and LQs, respectively, compared to the true RPC values. The SDRs and employment LQs estimates of RPCs are clearly both biased (i.e., the error term has a non-zero mean) and inefficient (i.e., the error terms have relatively high standard deviations). In the case of SDR, the error term has an average value of 0.372 and a standard deviation of 0.269. For the LQ those same values are, respectively 0.334 and 0.330. The offset of efficacy for bias makes a generalized preference between the two rather difficult. In the end, the explanatory power of the SDR gives it the edge, at least for estimating intranational trade in the EU.

Figure 3a: Errors by manufacturing sector using the SDRs, R^2 = .324



[9] Some of these methods had already been criticized by underestimating imports from other regions and the lack of flexibility in adapting to the size of the region (Flegg and Webber, 1997). Oddly enough, Lamonica and Chelli (2017) show that the FLQ and AFLQ approaches are inferior to the truncated LQ. This can be particularly problematic when the adjusting parameter δ in these methods is above .3.

Figure 3b: Errors by sector using LQs, R^2= .140



3.1.2. FLQ estimates of RPCs[10]

Besides testing the performance of SDR and the LQ as RPC estimates, we also do so for an approach that follow FLQ techniques. For this, we consider (Flegg 2013, 2018, 2019):

$$\dot{\rho}_i^c \equiv SLQ_i^c \times \lambda^* \qquad (3)$$

where $\dot{\rho}_i^c$ denotes an estimate of the actual RPC $\left(\rho_i^c\right)$ for industry *i* in country *c* within the EU.

$$SLQ_i = \begin{cases} \left(q_i^c / q_\bullet^c\right) / \left(q_i^\bullet / q_\bullet^\bullet\right), & \text{if } \left[\left(q_i^c / q_\bullet^c\right) / \left(q_i^\bullet / q_\bullet^\bullet\right)\right] \leq 1.0 \\ 1 & \text{, if } \left[\left(q_i^c / q_\bullet^c\right) / \left(q_i^\bullet / q_\bullet^\bullet\right)\right] > 1.0 \end{cases} \qquad (4)$$

and $\lambda^* \equiv \delta \log_2(1 + e_\bullet^c / e_\bullet^\bullet)$

(5)

In the above, $e_\bullet^c$ and $q_\bullet^c$ are total employment and output for country *c*, $e_\bullet^\bullet$ and $q_\bullet^\bullet$ are total EU-wide employment and output. Thus, for example, $e_\bullet^c / e_\bullet^\bullet$ is country *c*'s share of EU employment and $q_i^c / q_\bullet^c$ is industry *i*'s share of country *c*'s output. Thus, its estimates differ directly with δ and the truncated SLQ. Jahn et al. (2020) recommend a δ value of 0.3. Using WIOD data, we performed a rough search over values of δ; our results are presented in the Table 2.

[10] We acknowledge an anonymous referee who urged us to include this section on the FLQ due to its broad proliferation in the literature on I-O table regionalization.

Table 2: Summary statistics of the RPC-estimating capacity of the FLQ method for different values of δ

δ value	Mean error	Standard deviation	Share of RPCs overestimated	R^2
0.1	0.11	0.280	66.3%	0.184
0.2	-0.05	0.267	48.0%	0.194
0.3	-0.15	0.266	36.8%	0.183
0.4	-0.22	0.269	27.9%	0.164
0.5	-0.27	0.271	19.9%	0.145
0.6	-0.30	0.274	13.2%	0.129
0.7	-0.33	0.277	9.1%	0.115
0.8	-0.35	0.279	5.9%	0.103
0.9	-0.36	0.281	4.4%	0.093

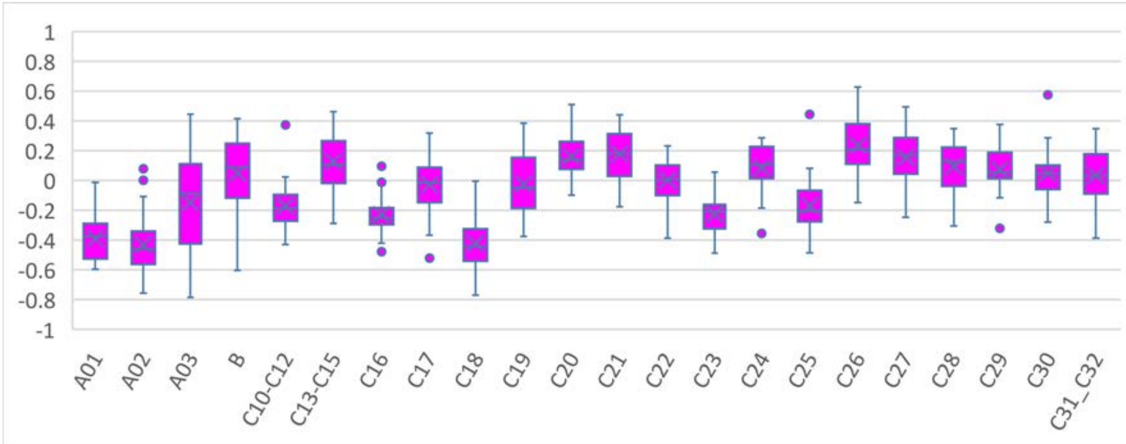
Table 2 shows that FLQ estimate of the RPC maximizes its descriptive power, as measured by R^2, when δ is close to 0.2. This is when $\sum_{c=1}^{28}\sum_{i=1}^{23}\left(\dot{\rho}_i^c - \rho_i^c\right)\Big/644$ is minimized and there is a "best" balance between over- and under-estimates of actual RPCs in WIOD. Interestingly, its standard deviation is higher and explanatory power, in terms of R^2, lower than that observed for SDRs.

Recall, as is always the case with the FLQ, that to identify the optimal δ requires knowing the actual RPCs. That is, until some cosmological δ is identified, we cannot divine δ a priori, which renders FLQs impractical in most cases. Either you have real RPCs and should use them, or you cannot know the optimal δ for your region of interest. In this vein, however, it is equivalent to econometric approaches.[11] In any case, we compare the FLQs with δ=.2 to other rows-only regionalization approaches in the remainder of the paper.

Figure 4 presents box-and-whisker plots for FLQ compared to actual RPCs. A quick glance suggests that the FLQ is superior to the LQ and SDR, not surprising since it relies on actual RPCs for parameterization. Its standard deviation, however, is about the same as that for SDRs. What is more interesting, however, is that it generates a better balance between over- and under-estimates of RPC values than do more sophisticated approaches that we discuss later. But, compared to them, it does so with a modicum of bias and with substantially lower ability to estimate true RPCs as measured by R^2 when compared to SDRs.

[11] Some vendors of regional economic accounts, who employ econometrically derived RPCs, extrapolate from the broader geography via which their model parameters are derived (say, from states to counties). There is no reason that FLQs could not be similarly applied.

Figure 4: FLQ error by sector, δ = .2 R^2 = .194



3.2. CHARM estimates of trade

We also calculated RPCs using CHARM, which was originally devised by Kronenberg (2009). Its formula has since been refined for interregional content by Többen and Kronenberg (2015) and Fujimoto (2015). We probe the accuracy of both.

Fujimoto (2018) found that his variant of CHARM works especially well compared to LQ-type techniques when estimating trade between non-contiguous regions. Többen and Kronenberg (2015) note that despite being an improvement over LQ-related approaches, CHARM-related approaches still over-estimate intraregional deliveries.

3.2.1. Többen and Kronenberg's CHARM formula

As we mentioned earlier, this method estimates gross imports and exports of a subnational region *c* from/to the rest of its country *R*, given estimates or data on regional production by industry i (q_i^c), total consumption $\sum_j z_{ij}^{\bullet c} + y_i^{\bullet c}$, as well as imports m_i^c and exports x_i^c from/to the rest of the world. Consequently, imports m_i^c and exports x_i^c refer are trade-flows of the country with non-European countries. Where $z_{ij}^{\bullet c}$ represents the total intermediate consumption and $y_i^{\bullet c}$ total final demand both for industry *i* in country *c*. The $\bullet$ denotes summation over the respective index, which is used here to clarify that $z_{ij}^{\bullet c}$ and $y_i^{\bullet c}$ consumption of products from domestic production, as well as imports from other EU countries and from the rest of the world. In this vein, it should be superior to LQ-based estimates.

The first step of CHARM, as developed by Többen and Kronenberg (2015), is to compute the cross-hauling potential θ_i^c of EU country c, i.e. the maximum amount of cross-hauling as per accounting balances:

$$\theta_i^c = 2\min\left[q_i^c - x_i^c; \sum_j z_{ij}^{\bullet c} + y_i^{\bullet c} - m_i^c; q_i^{\sim c} - x_i^R; \sum_j z_{ij}^{\bullet R} + y_i^{\bullet R} - m_i^R\right] \quad (6)$$

The right-hand side is multiplied by two (2) since cross-hauling is present in both import *and* exports.

But what fraction of the cross-hauling potential is actually realized? Többen and Kronenberg (2015) denote this fraction by h_i and estimate it for each product from the EU I-O table as

$$h_i^{EU} = \frac{\theta_i^{EU}}{2\min\left(q_i^{EU} - x_i^{EU}, \sum_j z_{ij}^{\bullet EU} + y_i^{\bullet EU} - m_i^{EU}\right)}, \tag{7}$$

where θ_i^{EU} denotes the amount of cross-hauling by product i in the EU I-O. It can be computed as $\theta_i^{EU} = x_i^{EU} + m_i^{EU} - \left|x_i^{EU} - m_i^{EU}\right|$, where $v_i^{EU} = x_i^{EU} + m_i^{EU}$ denotes the trade volume and $b_i^{EU} = e_i^{EU} - m_i^{EU}$ denotes the trade balance. The share h_i^{EU} is naturally applied to Equation (4). In applying this share to a region of a nation (or here to a country member of the EU), the underlying assumption is that the share of cross-hauling potential actually traded is the same for the region as it is for the nation. If the cross-hauling potential is fully realized, then $h_i = 1$; if not, then $h_i = 0$ (i.e., there is no cross-hauling). So, in the case of the CHARM method, the estimate of heterogeneity parameter and its relationship with the RPC depends upon whether the country is a net-exporter, i.e. $\min\left(q_i^c, \sum_j z_{ij}^{\bullet c} + y_i^{\bullet c}\right) = \sum_j z_{ij}^{\bullet c} + y_i^{\bullet c}$, or a net-importer, i.e. $\min\left(q_i^c, \sum_j z_{ij}^{\bullet c} + y_i^{\bullet c}\right) = q_i^{\bullet c}$ of a product. For the sake of simplicity, we do not distinguish between trade with the rest of the EU and the rest of the world in the following.

Case 1: The region is a net-exporter (i.e. x>m)

When a region is a net-exporter, Equation 5 becomes

$$h_i = \frac{x_i + m_i - x_i + m_i}{2\sum_j z_{ij} + y_i} = \frac{m_i}{\sum_j z_{ij} + y_i}. \tag{8}$$

Hence, the parameter h_i is the share of imports in total consumption, and relationship with the RPC is simply:

$$\dot{\rho}_i = 1 - h_i \tag{9}$$

Case 2: The region is a net-importer (i.e., x<m)

In case a region is a net-importer, however, Equation 5 becomes

$$h_i = \frac{x_i + m_i - m_i + x_i}{2q_i} = \frac{x_i}{q_i}. \tag{10}$$

Thus, the parameter h_i is the share of exports in gross output. In this case, the RPC formula is somewhat more complex, as the nominator and the denominator need to be transformed from exports and output, respectively, to an expression of domestic consumption and imports. Substituting the nominator with the accounting balance $x_i = q_i + m_i - \sum_j z_{ij} - y_i$, as well as the denominator by the accounting balance $q_i = \sum_j z_{ij} + y_i + x_i - m_i$ yields

$$h_i = 1 - \frac{\sum_j z_{ij} + y_i - m_i}{\sum_j z_{ij} + y_i + x_i - m_i} = 1 - \frac{\sum_j z_{ij} + y_i - m_i}{\sum_j z_{ij} + y_i + b_i} \tag{11}$$

The denominator can be re-written as $\sum_j z_{ij} + y_i + b_i = \left(\sum_j z_{ij} + y_i\right) * \left(1 + \frac{b_i}{\sum_j z_{ij} + y_i}\right)$. Multiplying Equation 9 by it and dividing by total consumption, $\left(\sum_j z_{ij} + y_i\right)$, then yields

$$h_i \left(1 + \frac{b_i}{\sum_j z_{ij} + y_i}\right) = \left(1 + \frac{b_i}{\sum_j z_{ij} + y_i}\right) - \frac{\sum_j z_{ij} + y_i - x_i}{\sum_j z_{ij} + y_i} = \left(1 + \frac{b_i}{\sum_j z_{ij} + y_i}\right) - \dot{\rho}_i \qquad (12)$$

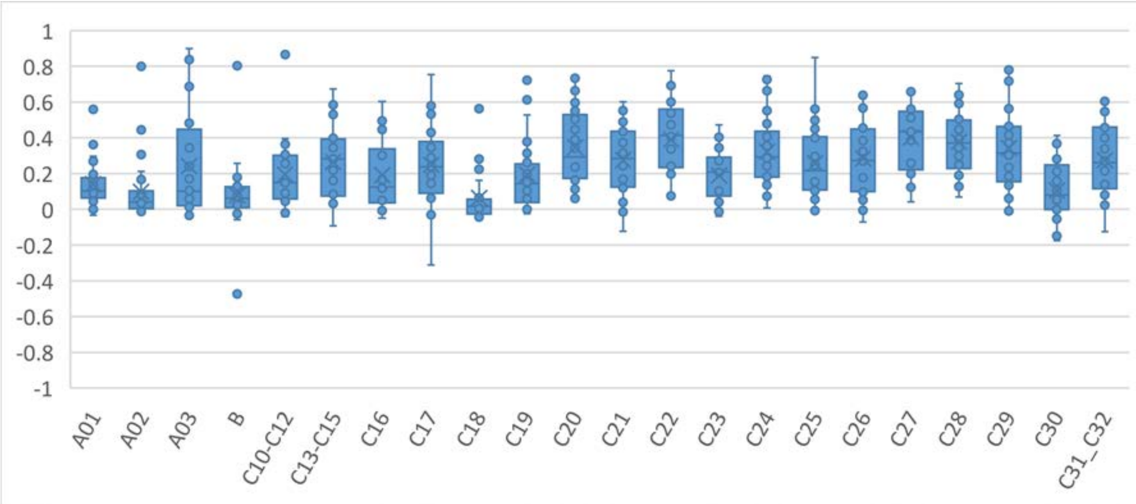
The second term on the right-hand side denotes domestic consumption over total consumption and is, thus, equal to the RPC, albeit expressed in term of intermediate and final demand. Finally, solving Equation 10 for the RPC delivers

$$\dot{\rho}_i = (1 - h_i)\left(1 + \frac{b_i}{\sum_j q_{ij} + y_i}\right) \qquad (13)$$

Thus, in the case of a net exporting sector a further scaling is required, to assure the value is between zero and one, as the trade balance is negative and always smaller than total consumption.

A comparison of Figure 5 to those in Figure 3, suggests that CHARM performs somewhat better than LQ and SDR approaches. The error of the CHARM estimates of RPCs has a mean of .240 and average standard deviation of .217. In terms of over estimation of the RPC, however SDR, LQ, and CHARM overestimate 99, 86, and 89 percent of the time, respectively. This highlights how the SDR, LQ and CHARM tend to overestimate the intraregional linkages and ultimately providing higher multipliers than the ones that are observed in the reality and an error that is not negligible. But, of the three nonparametric approaches, CHARM performs best. Still, much like the LQ and SDR estimates, it substantially underestimates cross-hauling, albeit not as badly. Subtracting the mean error from the CHARM RPC estimates values would remove its bias for goods-producing industries in the EU and likely improve the accuracy of accounts that would derive from the approach.

Figure 5: Errors by sector using CHARM, R^2 =.486



In any case, it clear that nonparametric methods used to estimate intraregional trade tend to substantially overestimate RPCs (underestimate interregional trade). Higher RPCs induce severe upward bias in the goods-producing aspects (at least) of the regional direct requirements matrix and, hence, the resulting Leontief inverse and output multipliers. This confirms findings of Többen and Kronenberg (2015).

3.2.2. Fujimoto's modified CHARM formula

Fujimoto (2018) proposes an alternative to the assumption that the share cross-hauling in the cross-hauling potential in interregional trade of a region is the same as that of the nation in international trade, as shown by Equation 5. His alternative, more strictly based on the Grubel-Lloyd (GL) index (Grubel and Lloyd, 1971), is measures the share of intra-industry trade (i.e. cross-hauling) in the trade volume of a product for a country. The GL index is defined as

$$GL_i^{EU} = \frac{v_i^{EU} - |b_i^{EU}|}{v_i^{EU}} \quad (14)$$

Fujimoto further assumes that the share of cross-hauling in the volume of interregional trade is the equivalent to the nation's volume of international trade, such that $h_i^c = GL_i^{EU}$ so that interregional trade is estimated by

$$v_i^c = \frac{|b_i^c|}{1-h_i^c}. \quad (15)$$

The trade balance of a region with the rest of the country, b_i^c, can then be computed as the difference between (i) regional gross output less exports abroad and (ii) total regional consumption less imports from abroad. Comparing interregional imports estimated by this procedure with survey-based imports, as well as output multipliers, Fujimoto (2018) shows this approach yields estimates of trade with significantly lower error and bias compared to Többen and Kronenberg's (2015) version.

But Fujimoto's approach does not guarantee that the resulting interregional trade estimates are consistent with regional accounting balances. This is the rationale that Többen and Kronenberg use when applying cross-hauling potential of a region in the denominator of their CHARM estimator. If the nation, for example, displays a non-zero GL index for a product that is not produced within the region, Equation 13 delivers a positive value even though output is zero; this forces imports to exceed total regional consumption. So, the resulting RPC is negative, which when applied to the direct requirements matrix would yield a problematic set of regional accounts. Moreover, this is no wild-eyed theoretical argument. Using Fujimoto's MCHARM formula to WIOD data, we obtained negative RPCs for 337 of 1568 national industries (21.5%). The "true" RPCs for the industries tended to be small, but no rule for adjustment immediately emerged upon observation. In fact, Fujimoto (2018) acknowledges this problem and defines consistency constraints on the estimates, but does not discuss how adherence to these constraints is achieved. For lack of a reasonable way to assign a set of nonnegative values to this minority of negative RPCs obtained by MCHARM formulae, we opted to not include MCHARM in the comparison of rows-only IO regionalization approaches that we perform in later section of this paper. Still, we note that future research might consider some sort of hybrid that combines Fujimoto's approach with Többen and Kronenberg's, since its development would be sufficient scope for another full paper.

3.3. Econometrically derived estimates

As noted in the introduction, Stevens and Trainer (1976) used a regression-based approach based on location theory to estimate intraregional trade. Since that work, a substantial number of articles has examined the multifarious nature of the determinants of subnational interregional trade, which this more recent spate of literature appears to avoid. But fairly early on, Stevens et al. (1980, 1983), Treyz and Stevens (1985), Despotakis (1985), and Stevens, Treyz, and Lahr (1989) examined the case of states of the U.S. economy. Since then, a number of other research teams have attempted to estimate intraregional and interregional trade, c.f. Celik (2004), Jackson et al. (2005, 2006), Lindall, Olson, and Alward (2006), Robinson and Liu, 2006), Llano et al. (2010), Jackson and Schwarm (2011), Chun, Kim, and Kim (2012), Hwang et al. (2016), LeSage and Llano (2016), Sun (2018), and Zhang, Chen and Jin (2019).

Clearly, trade across space has been heavily studied lately. Much of the work on an international scale evaluates determinants of the "propensity to export." Francois and Manchino

(2006), who summarize this literature, note that economies export more or less depending on their relative wealth (GDP per capita), quality of their institutions (e.g., liberalization of trade agreements), distances to countries that have demands unmet from domestic sources, access to low-cost freight transportation and the quality of their communications infrastructure. Economic base theory[12] further identifies significant differences in the relative tradability of commodities produced by agriculture, mining, manufacturing sectors when compared to services. We know that openness is related to productivity since more-productive firms are better equipped to compete in more distant markets (del Gatto et al. 2008); less-productive ones are likely to confine trade to closer environs. In turn, openness and tradability are related to country size and government's share of GDP (Alesina and Wacziarg, 1998). Further cultural distance and linguistic distance can play a role (Boisso and Ferrantino (1997), and Polachek's (1992) notion that the likelihood of political conflict is reduced by the magnitude of trade relations has similar implications as well. Colonial origins have been shown to play a decisive role too (Ortega and Peri, 2014). In addition, in a review of the interplay between exchange rates and international trade, Auboin and Ruta (2013, p. 600) find the relationship "a complex one… Empirical studies also broadly confirm that contingent trade measures are used in response to trading partners' currency depreciation."

At a subnational level, the theory is somewhat less "complex." Exchange rates and trade agreements do not exist; technological differences and cultural distances (e.g., language and colonial origins) are generally far less pronounced (i.e., have lower statistical variance).

Still, small regions are expected to be more open and, hence, more subject to swings in trade than are larger regions. Smaller regions are more likely to have more extensive problems with regard to undisclosed data (c.f. Sargento et al., 2012; Rodrigues and Lahr, 2018). Services are also expected to be less tradable than are goods-producing industries, with the exception, perhaps, of accommodation services and some producer services, e.g. freight transportation arrangement, venture capital services, patent attorney offices, architects' offices, engineering services, computer systems design, research & development services, and management consulting.

Treyz and Stevens (1985) show that RPCs for commodities across U.S. states are functions of each state's areal size, total sectoral demand, SDR, weight/value ratio, relative establishments size in terms of employees, and other sectoral and geographic fixed effects. The rationale behind the dependent variable, RPC, and the set of explanatory variables we use (and largely based on those used by Treyz and Stevens) is that geography (*land area*) and economy are linked since geographically smaller economies are less likely to be capable of satisfying the local demand of their firms and households. Thus, countries with larger land areas should have higher RPC values;

[12] Sombart (1916) appears to be the originator of economic base theory. Both Lösch (1954) and Andrews (1953) with whose work on this topic regional economists are more apt to be familiar; both cite what is ultimately Sombart's work.

this helps to explain the lower positions of Cyprus, Luxembourg and Ireland in Figure 1. While *land area* explains the geographical size of a market, sectoral demand, SLQ, and SDR (again, both truncated here) measure economic size. Regions with high demands are more likely to have diversified economies, and regions with more relative supply than demand should better equipped to satisfy their own demands. Finally, weight/value ratio is expected to have a significant correlation with the RPC since it can be considered a proxy of transportation costs.[13] Low-valued products are likely to be shipped by less-expensive means since freight carrying costs (insurance and storage-related costs) are of less concern. Further if the value is extraordinarily high per unit weight (like specialized medical devices), less of the item is shipped. Therefore, demands for relatively low-valued products (like stone and gravel) are more apt to be fulfilled by local suppliers and less so by more-distant competitors. We add to this mix a variable that reflects tourism intensity—the number of person-nights in hotels in each country. Our rational for including this variable is that tourism enhances local goods production as visitors seek unique local experiences and fare as well as locally produced goods to take home during their vacations.

Table 2: Quasi-binomial RPC estimates

	(1) LQ only	(2) FLQ equivalent	(3) Full
Constant	-0.785	-3.308	-6.881
SLQ truncated	0.370	0.401	0.621
Employment share	---	0.302	---
ln(land area)			0.280
ln(hotel room-nights/area)			0.192
ln(ton/$)			0.102
S/D truncated			2.024
Agriculture and Forestry (A01, A02)			1.481
Food, Beverage & Tobacco (C10, C11, C12)			0.708
Wood and wood products (C16)			0.666
Printing and recorded media (C18)			2.154
Pharmaceutical & chemical products (C20, C21)			-0.750
Electrical & electronic equipment (C26, C27)			-0.962
R^2	0.088	0.191	0.660

Note: All the variables are statistically significant at the 0.001 level with a two-tailed test "***"

We apply a quasi-binomial functional form. Like the somewhat convoluted functional form used by Treyz and Stevens (1985), it too constrains estimates of RPCs within the admissible interval of 0 and 1. Gelman and Hill (2007) and De Witte and Verschelde (2010) underline that the parameters of the binomial unbounded linear predictors may yield over- and under-dispersed values. As long as affiliated noise is relatively low, such that the observations are close to the frontier, a high share of estimated values are likely to be close to 1. The quasi-binomial approach

[13] These data were derived from the International Trade Center database, see https://www.trademap.org/

introduces a scale parameter that alters the scale of the variance. This more-flexible error structure of the quasi-binomial structure should yield higher levels of dispersion in the mean-variance relationship. The results of this regression are presented in Table 2 above. As elsewhere in this paper, the dependent variable is the RPC by manufacturing sector in each EU country. Note that we also perform such a regression on the SLQ alone and on an FLQ equivalent. The binomial yields unbiased estimates for both. Interestingly, applying the approach to the truncated SLQ reduced its explanatory power as indicated by $R^2 = .088$, and which in raw form was .140. The FLQ equivalent in quasi-binomial form suffers almost no loss in explanatory power while losing bias and gaining economic interpretability in its parameters.

In sum, the RPC is expected to be higher when we are in the presence of larger regions, where industries are fairly highly concentrated (high LQ and SDR), a greater intensity of tourism, and industries in which products produced are more expensive per unit of weight. That is, signs of all coefficients have expected outcomes. Also, despite the inclusion of weight/value ratio *(ton/$1,000*) variable, the demands for agriculture, food products, wood products, and media were still more apt to be fulfilled via local supplies than the average industry product. Alternatively, lower-weight manufacturing products as pharmaceutical and chemical products and electrical and electronic equipment had lower than otherwise expected RPCs, suggesting that local production was more likely to satisfy more-distant markets than otherwise expected from the weight/value ratio.

Most of the variables were estimated using data from the WIOD socio-economic accounts but also other data sources such as EUROSTAT and OECD. We used this opportunity to test the statistical significance of other variables such as R&D's share of the GDP as well as other indicators related to productivity, like a region's share of national employment and compensation. These other variables were not statistically significant and thus not included in the final analysis.[14] Regardless, by modeling RPCs it is clear that it is not explained well by a truncated LQ alone; even employment share does not add much extra explanatory power. But other locational variables had substantial influence. Indeed, more than 66 percent of the variance in the RPC—the share of local demand for a commodity that is fulfilled by local suppliers—was explained by the model.[15]

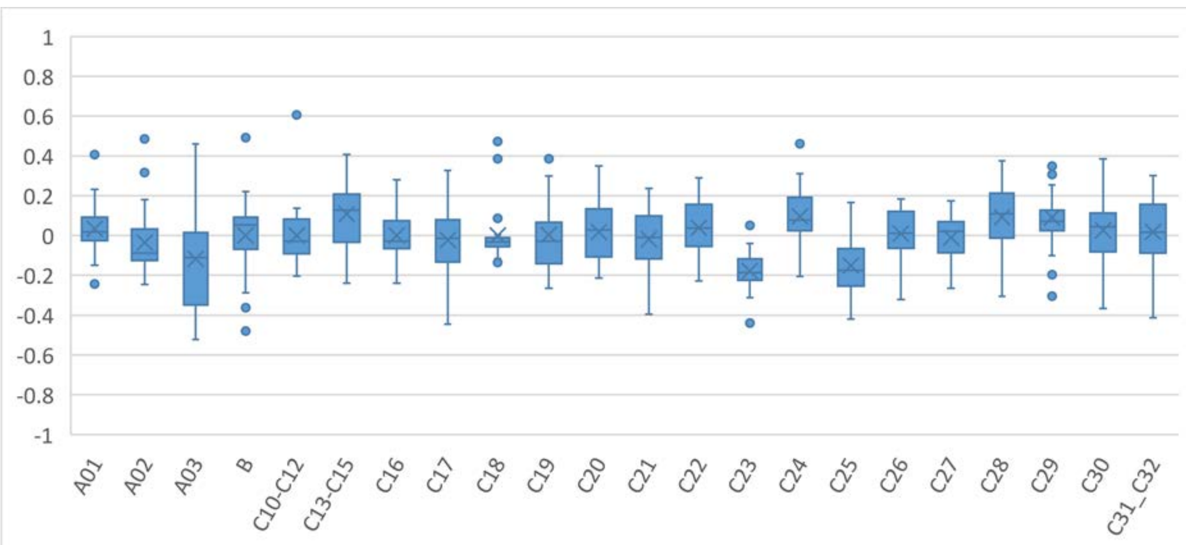
[14] Note, despite having the information by sector across counties, we did not implement the relative establishment size variable as per Stevens and Treyz (1985). Its collinearity with other key variables, namely the weight/value ratio, caused issues. We also take this opportunity to note that using binary variables for more industries would effectively have made the coefficient for the weight/value ratio variable such that it would not have been statistically significant from zero. Meanwhile, the value of R^2 would have risen by about .02. Given the level of aggregation and minimal degrees of freedom, we deemed it more desirable to retain weight/value ratio.

[15] Our quasi-binomial formulation was also applied to the Exiobase (Wood et al., 2015), a multiregional IO dataset that includes 27 of the 28 European countries (Croatia is excluded) and a greater detail in goods-producing sectors (93 industries). But the absence of data on employment, number of establishments and other relevant data reduced the number of variables that we could tested in this framework. Nevertheless, for the year 2014, the quasi-binomial representation of the RPC had an R^2 of .69 and most

Table 2 also reveals how the RPC changes with changes in independent variables, *ceteris paribus*. Then, the RPC value is expected to be bigger, (or in other words, the sectors in each country are expected to satisfy a bigger share of their local consumption) if the area of a country has a larger than average size, the number of nights spent is greater than the EU average, the SDR is larger, the total demand is higher, and the industry is relatively more concentrated in the region (higher than average LQ).

The effort to estimate RPCs using quasi-binomial regression approach proves worthwhile (see Figure 6). Its error term, not surprisingly, has a zero average and a standard deviation of just 0.17. Moreover, it is a much better predictor than are other alternatives that must be gauged via pre-existing knowledge of the rows-only RPC, namely the FLQ.

Figure 6: Errors by sector using quasi-binomial procedure, R^2= .660



4. Error and bias in regionalization

So, assuming we have made a convincing case for quasi-binomial estimates of RPCs, how much difference do improved RPCs make to the accuracy of regional accounts? That is, exactly how much upward bias does the underestimation of cross-hauling in goods-producing industries (only) induce in practice? To answer this question, we developed nonsurvey tables for the 28 EU countries. To do this, we first built a pseudo-national table—an aggregate EU table that is the sum of the 28 WIOD national accounts of the countries that form the EU. We next produced an average technology or direct requirements matrix for this meta-region for our simulations by column normalizing the result. We subsequently generated 28 nonsurvey 'regional' direct requirements

of the variables considered on Table 2 proved to be statistically significant (an exception was hotel room-nights/area and the binary variable for "Pharmaceutical & chemical products"). This suggests that variables such as area, SDR, weight/value ratio, SLQ, and the binary variables are likely to be valuable in modeling trade in multiregional and multicountry frameworks.

matrices by pre-multiplying the national technology matrix by the diagonalized matrix of RPCs, $\hat{\boldsymbol{\rho}}$ for each country of the EU. This means we need to also estimate RPCs for non-goods-producing industries as well. We opted to use a truncated SDR to estimate RPCs in such cases, and we use them across for all the tables.[16] That is, we use the CHARM, truncated LQ and SDR, as well as the quasi-binomial RPC-estimates for goods-producing sectors but always use SDR for all other sectors. Further, in producing the 28 nonsurvey tables for the EU nations, we further assumed, per unit of output, that shares of value added and international trade were fixed for a sector. This mimics a fairly un-informed regionalization approach in which no information on productivity differentials and international trade is available at a regional level.[17] We next compared the resulting nonsurvey tables (28 'regions' for each of 4 types of RPC estimates: LQ, SDR, CHARM, and quasi-binomial).

The degree of accuracy of finding for a particular application depends on how one uses I-O accounts. We opted to cover a broad swath of territory in this regard; so, we compare differences between actual and estimated direct requirement matrices, Leontief inverses, and output multipliers. To measure differences between actual-estimate array pairs, we use two different measures: the mean absolute deviation (MAD) and the weighted absolute difference (WAD).[18] Lower values indicate less bias in both measures. If we let a_{ij} denote an element from the actual array and $\tilde{a}_{ij}$ denote its estimate, then the below are the mathematical representations of the measures:

$$\mathrm{MAD} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}-\tilde{a}_{ij}|}{mn}$$

$$\mathrm{WAD} = \frac{\sum_i \sum_j a_{ij} \times |a_{ij} - \tilde{a}_{ij}|}{\sum_i \sum_j a_{ij}}$$

MAD is simple and readily understood; it is simply the average absolute bias in the array elements. WAD is the MAD but weighted by the size of the actual array element, and thus assigns greater benefit to being accurate on larger array elements. When the focus is technology (the **A** matrix) or the Leontief inverse (**L**), WAD provides solid guidance since variance in zero values and values close to zero in such matrices can be regarded as noise. Finally, In the case of the output multipliers ($\boldsymbol{\mu}$), as per Oosterhaven (1981, p. 49) we subtract 1.0 from each element of the vector prior to testing; the same goes for the diagonal elements of the Leontief inverse (**L**). The idea behind this is that 1.0 represents the direct effects, which is constant across and inherent to all the output

---

[16] We used the SDR, rather than the LQ since it correlated better with the RPCs of nongoods-producing sectors as estimated in WIOD. To be perfectly clear, here we are not referring to estimates for goods-producing sectors as presented in Figures 3a and 3b.

[17] Lahr (2001b) discusses how to incorporate such data at the regional level if it is available.

[18] See Lahr (2001a) for more information on the rationale for choosing these two measures among the many available. None of these measures is entirely satisfactory for measuring differences between interindustry matrices. Structural decomposition approaches are better suited, but the selection of a proper decomposition to apply is not a simple choice.

multipliers; thus, any error or bias in multipliers can attach to the rest of an output multiplier's value.

In Table 3, we present test results of comparing the identified nonsurvey array type to the "actual" equivalents as reflected in WIOD. The bolded numbers in Table 3 identify the lowest value by error by each country for each array—**A**, **L** or **μ**, respectively, the direct requirements matrix, total requirements matrix, and vector of output multipliers. Findings in Table 3 suggest that RPCs generated by the quasi-binomial method are, indeed, superior to equivalents produced via LQ, SDR or CHARM. In the case of the direct requirements (**A**) matrix, the quasi-binomial performs better in 13 of the 28 cases. Moreover, in those cases in which it does not, its WAD is second lowest and not far off the best. CHARM performs better in 10 of 28 cases and, collectively, the SDR and LQ perform better in the remaining three. Interestingly, the quasi-binomial estimates do not perform as well in cases in countries that have higher average RPC values.

Table 3: Measures of Differences between estimated and observed I-O tables

	Quasi-binomial RPC			LQ			SDR			CHARM		
	A	**L**	**μ**	**A**	**L**	**μ**	**A**	**L**	**μ**	**A**	**L**	**μ**
	WAD	*WAD*	*MAD*	*WAD*	*WAD*	*MAD*	*WAD*	*WAD*	*MAD*	*WAD*	*WAD*	*MAD*
AUT	**0.033**	**0.053**	**0.121**	0.036	0.066	0.245	0.036	0.070	0.276	0.035	0.063	0.206
BEL	**0.023**	**0.037**	**0.130**	0.023	0.047	0.240	0.024	0.057	0.306	0.024	0.049	0.233
BGR	**0.034**	**0.045**	**0.173**	0.035	0.058	0.191	0.036	0.060	0.193	0.034	0.051	0.180
CYP	0.060	**0.055**	**0.174**	**0.058**	0.058	0.239	0.058	0.057	0.214	0.059	0.056	0.201
CZE	**0.037**	**0.054**	**0.151**	0.039	0.068	0.251	0.039	0.066	0.218	0.038	0.059	0.185
DEU	**0.023**	**0.035**	**0.080**	0.026	0.043	0.195	0.025	0.044	0.214	0.024	0.036	0.144
DNK	0.035	**0.038**	**0.100**	0.036	0.047	0.167	0.034	0.051	0.231	**0.034**	0.044	0.180
ESP	0.023	0.038	0.118	**0.022**	**0.036**	0.099	0.023	0.043	0.146	0.022	0.037	**0.097**
EST	**0.027**	**0.035**	**0.108**	0.030	0.056	0.274	0.028	0.050	0.228	0.028	0.045	0.185
FIN	0.025	0.039	**0.128**	0.024	0.044	0.185	0.024	0.044	0.184	**0.023**	**0.038**	0.134
FRA	0.022	**0.035**	**0.094**	0.024	0.040	0.104	0.022	0.043	0.171	**0.021**	0.039	0.114
GBR	0.026	0.043	**0.100**	0.027	0.045	0.112	0.024	0.043	0.160	**0.024**	**0.041**	0.112
GRC	**0.066**	**0.056**	**0.149**	0.067	0.062	0.168	0.067	0.063	0.172	0.067	0.059	0.155
HRV	**0.056**	**0.064**	**0.155**	0.056	0.083	0.306	0.056	0.077	0.246	0.057	0.072	0.201
HUN	**0.027**	**0.045**	**0.199**	0.028	0.066	0.392	0.027	0.062	0.367	0.028	0.055	0.300
IRL	**0.028**	**0.048**	**0.187**	0.028	0.056	0.247	0.029	0.064	0.327	0.029	0.058	0.280
ITA	0.023	0.040	0.194	0.019	**0.033**	**0.122**	**0.019**	0.034	0.126	0.020	0.034	0.132
LTU	0.029	**0.044**	**0.220**	0.029	0.058	0.373	0.029	0.059	0.391	**0.029**	0.053	0.334
LUX	0.038	**0.054**	**0.226**	0.037	0.064	0.325	0.037	0.068	0.352	**0.037**	0.064	0.319
LVA	0.048	**0.060**	**0.149**	0.047	0.068	0.184	0.047	0.066	0.174	**0.046**	0.063	0.206
MLT	0.036	**0.046**	**0.138**	0.034	0.051	0.202	**0.034**	0.049	0.181	0.034	0.047	0.165
NLD	**0.025**	**0.042**	**0.165**	0.026	0.047	0.229	0.026	0.051	0.334	0.025	0.046	0.266
POL	0.027	**0.041**	**0.100**	0.026	0.048	0.238	0.026	0.044	0.192	**0.026**	0.042	0.131

PRT	0.028	0.047	**0.091**	0.026	0.045	0.153	0.027	0.049	0.188	**0.026**	**0.045**	0.138
ROU	**0.024**	**0.043**	**0.136**	0.025	0.058	0.198	0.025	0.056	0.170	0.025	0.050	0.140
SVK	0.041	**0.055**	**0.104**	0.041	0.067	0.306	0.039	0.063	0.272	**0.039**	0.057	0.209
SVN	**0.026**	**0.044**	**0.119**	0.028	0.062	0.270	0.027	0.056	0.230	0.027	0.051	0.182
SWE	**0.022**	**0.040**	**0.129**	0.025	0.055	0.231	0.023	0.054	0.255	0.023	0.047	0.191

In a comparison of Leontief inverse, however, that the quasi-binomial model performs even better, having the lowest WAD in 23 of the 28 instances. LQ and CHARM perform better in Spain and Italy, again, countries that have high average RPCs in goods-producing sectors. In the case of Italy, a main source of bias is in the specific cases of machinery, transport equipment and transport manufacturing. In the case of Spain, the main "offending" sectors are primary industries and food manufacturing. These appear to be cases in which just a bad apple is spoiling the whole bunch. Just a bit of bias in **A** circulates and accumulates heavily in **L**. But a high WAD, which in **L** runs from 0.022 (France) to 0.067 (Greece), is clearly not a sufficient indicator of bias transmission from **A** to **L**. Clearly more research on the transmission of bias across interindustry arrays is needed.

Table 4: Share of the cells over and underestimated in the derived A matrix

	RPC Estimate		**LQ**		**SDR**		**CHARM**	
	Over	Under	Over	Under	Over	Under	Over	Under
AUT	56.9%	43.1%	80.2%	19.8%	80.4%	19.6%	75.0%	25.0%
BEL	70.2%	29.8%	85.5%	14.5%	87.8%	12.2%	83.8%	16.2%
BGR	44.4%	55.6%	74.3%	25.7%	64.6%	35.4%	67.1%	32.9%
CYP	60.2%	39.8%	76.4%	23.6%	71.2%	28.7%	75.0%	25.0%
CZE	52.1%	47.9%	75.3%	24.7%	76.1%	23.9%	69.9%	30.1%
DEU	54.5%	45.5%	83.8%	16.2%	86.8%	13.2%	78.9%	21.1%
DNK	66.1%	33.9%	82.6%	17.4%	87.4%	12.6%	82.9%	17.1%
ESP	37.2%	62.8%	49.8%	50.2%	60.2%	39.8%	75.0%	25.0%
EST	62.8%	37.2%	82.3%	17.7%	81.5%	18.5%	78.5%	21.5%
FIN	37.7%	62.3%	65.0%	35.0%	63.4%	36.6%	56.9%	43.1%
FRA	71.7%	28.3%	76.9%	23.1%	86.2%	13.8%	81.6%	18.4%
GBR	40.9%	59.1%	49.3%	50.7%	65.9%	34.1%	56.4%	43.6%
GRC	53.2%	46.8%	71.6%	28.4%	68.4%	31.6%	63.5%	36.5%
HRV	55.7%	44.3%	71.6%	28.4%	70.4%	29.6%	66.0%	34.0%
HUN	46.9%	53.1%	80.6%	19.4%	77.5%	22.5%	71.2%	28.8%
IRL	65.4%	34.6%	79.0%	21.0%	81.2%	18.8%	78.6%	21.4%
ITA	33.9%	66.1%	56.5%	43.5%	57.3%	42.7%	75.0%	25.0%
LTU	54.8%	45.2%	77.3%	22.7%	75.1%	24.9%	71.6%	28.4%
LUX	67.0%	33.0%	79.8%	11.1%	82.2%	8.7%	79.3%	11.6%
LVA	64.9%	35.1%	82.5%	17.5%	79.3%	20.5%	77.4%	22.6%
MLT	46.6%	53.4%	77.7%	17.9%	75.2%	20.4%	70.5%	25.1%
NLD	56.3%	43.7%	65.7%	34.3%	80.2%	19.8%	73.6%	26.4%
POL	30.5%	69.5%	64.0%	36.0%	60.1%	39.9%	50.7%	49.3%

PRT	44.7%	55.3%	73.8%	26.2%	73.1%	26.9%	67.1%	32.9%
ROU	36.4%	63.6%	63.3%	36.7%	63.2%	36.8%	57.7%	42.3%
SVK	48.2%	51.8%	75.2%	24.8%	72.8%	27.2%	68.2%	31.8%
SVN	34.3%	65.7%	66.1%	33.9%	61.2%	38.7%	54.4%	45.6%
SWE	48.4%	51.6%	66.9%	33.1%	71.0%	29.0%	64.5%	35.5%

The use of WAD and MAD does only explain the *absolute deviation* obtained. Neither identifies the extent of under- or over-estimation bias. This is despite the fact that such bias has very distinct economic implications in practice. Results in Table 4 are used to uncover the potential of such consequences; it presents the share of over- or under-estimated cell values in the 28 **A** matrices. It reveals that quasi-binomial estimates of goods-producing RPCs yields fewer overestimates across all the 28 cases and that CHARM estimates are second best in 19 cases. That is, the quasi-binomial approach yields more balanced results with over and underestimation occurring in different circumstances. It overestimates in just 15 cases and is more extreme for Belgium and France. It underestimates in 13 cases and most severely in Poland and Italy.[19] On the other hand, LQ only underestimates the **A** values in two cases, while the SDR and the CHARM methods never underestimate. These results give us a better picture of the estimated errors presented in Table 3 and of their direction and economic relevance. In essence, overestimation bias is likely present in interindustry arrays based on simple estimates of RPCs. One cannot say that as matter of fact in the case of arrays produced via econometrically produced versions of the RPC.

5. Conclusions

Over the past couple of decades, a good number of papers have been written that use nonsurvey techniques to estimate regional I-O tables. All of them assume that within a nation, technology is spatially constant (knowledge flows freely). Most have little other economic foundation, merely identifying parameters for location quotients (LQs) or supply/demand rations (SDRs) that estimate them "best" within specific nations. We point out that such research is at best misguided since it ignores a flurry of work done in the mid- to late-1980s that employs location theoretic grounds for advancing a theory of the content of regional supply percentages—now generally called "RPCs" (i.e., regional purchase coefficients). We analyze a core set of RPC-estimate alternatives with a focus on those that permit cross-hauling. We do so by examining the use of domestically produced goods across countries in the European Union (EU). In the vein of Treyz and Stevens (1985), who used transportation survey statistics to derive "true" RPCs across U.S. states, we apply a handful of locational variables from the international trade literature to explain a nation's propensity to use

[19] There is a residual number of cells that are not over or underestimated in the original and the estimated matrices. These are the cells that present null values both in the true **A** and in the estimated **A,** and they represent less than 0.1% of the cells in the true matrix.

local production. The fact that the EU has fairly unfettered trade amongst its member countries is critical to our work since it enables direct implications to the estimation of subnational RPCs.

After estimating RPCs through various well-known techniques, we regionalize an aggregate of the national tables in WIOD to derive an EU “national” direct requirements table. We then compare the resulting “regional” direct requirements, regional Leontief inverses, and vector of regional output multipliers to the actual equivalents for the respective country tables in WIOD. We find that our theoretically derived econometric approach by far provides the best estimates. All others have over-estimation bias inherent to them, this includes the latest rendition of CHARM, although it performs better than estimates predicated upon LQs and SDRs.

Our work suggests a continuing need for research on the propensity of a region to engage in interregional trade. Other related research avenues remain for RPC estimating that could supplement those taken here. The availability of subnational data within any given nation drives the sort of regionalization scheme that can be applied. So, as in the case of Egypt, it well may be that the very basic employment LQ is the best technique that can be achieved. Then again, as Haddad (2014) has shown, one can generate interregional trade-flow estimates using just travel times and a simple gravity model. The ease with which multiregional tables can now be developed and the vast differences in the sets of data available across regions should propel this sort of research for years to come.

REFERENCES

Alesina, Alberto and Romain Wacziarg (1998) Openness, country size and government. *Journal of Public Economics*, 69, 305–321.

Andrews, Richard B. Mechanics of the urban economic base: Historical development of the base concept, *Land Economics*, 29(2), 162–166

Auboin, Marc and Michele Ruta. (2013) The relationship between exchange rates and international trade: A literature review. *World Trade Review*, 12, 577–605.

Begg, Robert and Andrew Isserman. (1986) Nonsurvey interregional I-O models. In 33rd North American Meetings of the Regional Science Association, November (Vol. 15).

Boero, Riccardo, Brian K. Edwards and Michael K. Rivera (2018) Regional input–output tables and trade flows: an integrated and interregional non-survey approach, *Regional Studies*. 52, 225–238.

Boisso, Dale and Michael Ferrantino. (1997) Economic distance, cultural distance, and openness in international trade: Empirical puzzles. *Journal of Economic Integration*, 456–484.

Celik, H. Murat. (2004) Forecasting interregional commodity flows using artificial neural networks: an evaluation. *Transportation Planning and Technology*, 27, 449–467.

Chun, Yongwan, Hyun Kim, and Changjoo Kim. (2012) Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: an application of the US interstate commodity flows. *Computers, Environment and Urban Systems*, 36, 583–591.

De Witte, Kristof and Marijn Verschelde. (2010) Estimating and explaining efficiency in a multilevel setting: A robust two-stage approach *(No. 10/657)*. Ghent University, Faculty of Economics and Business Administration.

Del Gatto, Massimo, Gianmarco I. P. Ottaviano, and Marcello Pagnini. (2008) Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms. *Journal of Regional Science*, *48*, 97–129.

Despokais, Konstantinus. (1985) *On the Calculation of Gross Regional Trade Flows*. Report EEA-85-01. Engineering-Economics Associates, Berkeley, CA.

Elshahawany, Dina N., Eduardo A. Haddad, and Michael L. Lahr. (2017) Accessibility, transportation cost, and regional growth: A case study for Egypt, *Middle East Development Journal*, 9:2, 256-277. DOI: 10.1080/17938120.2017.1366773

Flegg, Anthony T., Leonardo J. Mastronardi, and Carlos A. Romero. (2016) Evaluating the FLQ and AFLQ formulae for estimating regional input coefficients: empirical evidence for the province of Córdoba, Argentina. *Economic Systems Research* 28, 21–37.

Flegg, Anthony T. and T. Tohmo. (2013) Regional input–output tables and the FLQ formula: a case study of Finland. *Regional Studies*, 47, 703–721.

Flegg, Anthony T. and T. Tohmo. (2016) Estimating regional input coefficients and multipliers: the use of FLQ is not a gamble. *Regional Studies*, 50, 310–325.

Flegg, Anthony T. and T. Tohmo. (2018) The regionalization of national input–output tables: A review of the performance of two key non-survey methods. In *Applications of the Input-Output Framework*. Singapore: Springer, pp. 347–386.

Flegg, Anthony T. and T. Tohmo. (2019) The regionalization of national input-output tables: A study of South Korean regions. *Papers in Regional Science*, 98, 601–620.

Flegg, Anthony 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.

Francois, Joseph and Miriam Manchin. (2006) Institutional quality, infrastructure, and the propensity to export. Unpublished, January, World Bank, Washington, DC. Available online in October 2019 at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.557.5080&rep=rep1&type=pdf.

Fujimoto, Takashi. (2015) Quantitative analysis of the regional income determinant factors in a remote island economy: Generation and application of a regional input–output table. *Agricultural Economic Research*, 86, 257–272. doi.org/10.11472/nokei.86.257 [in Japanese]

Fujimoto, Takashi. (2018) Appropriate assumption on cross-hauling national input–output table regionalization, *Spatial Economic Analysis*, 14, 106-128.

Garhart, Robert E. and Frank Giarratani. (1987) Nonsurvey input-output estimation techniques: Evidence on the structure of errors. *Journal of Regional Science*, 27, 245–253.

Gelman, Andrew and Jennifer Hill. (2007) *Data Analysis Using Regression and Hierarchical/Multilevel Models*. New York: Cambridge University Press.

Grubel, Herbert G., and Peter J. Lloyd. (1971) The empirical measurement of intra-industry trade, *Economic Record*, 47, 494–517. doi:10.1111/j.1475-4932.1971.tb00772.x

Haddad, Eduardo A. (2014) Trade and interdependence in Lebanon: An interregional input-output perspective. *Journal of Development and Economic Policies*, 16, 5–45.

Haddad E. A. and G.J.D. Hewings. (2005) Market imperfections in a spatial economy: Some experimental results, *Quarterly Review of Economics and Finance*, 45, 476-496.

Hwang, Ho-Ling, Stephanie Hargrove, Shih-Miao Chin, Daniel W. Wilson, Hyeonsup Lim, Jiaoli Chen, Rob Taylor, Bruce Peterson, and Diane Davidson. (2016) *The Freight Analysis Framework Version 4 (FAF4)-Building the FAF4 Regional Database: Data Sources and Estimation Methodologies.* No. ORNL/TM-2016/489. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN.

Jackson, Randall W. and Walter R. Schwarm. (2011) "Accounting foundations for interregional commodity-by-industry input-output models," *Letters in Spatial and Resource Sciences,* 4, 187–196.

Jackson, Randall W., Walter R. Schwarm, and Yasuhide Okuyama. (2005) "An evaluation of interregional trade estimation methods," *Journal of Regional Analysis and Policy,* 36, 84–93.

Jackson, Randall W., Walter R. Schwarm, Yasuhide Okuyama, and Samia Islam. (2006) "A method for constructing commodity by industry flow matrices," *Annals of Regional Science,* 40, 909–920.

Jahn, Malte. (2017) Extending the FLQ formula: a location quotient-based interregional input–output framework." *Regional Studies*, 51,1518–1529.

Kowalewksi, Julia. (2015) Regionalization of national input–output tables: empirical evidence on the use of the FLQ formula. *Regional Studies*, 49, 240–250

Kronenberg, Tobias. (2009) Construction of regional input-output tables using nonsurvey methods: the role of cross-hauling. *International Regional Science Review*, 32(1), 40–64.

Lahr, Michael L. (1993) A review of the literature supporting the hybrid approach to constructing regional input–output models. *Economic Systems Research*, 5, 277–293.

Lahr, Michael L. (2001a) A strategy for producing hybrid regional input-output tables. In M. L. Lahr and H.W.A. Dietzenbacher, eds., *Input-output Analysis: Frontiers and Extensions*. Palgrave: New York, pp. 211–242.

Lahr, Michael L. (2001b) Reconciling domestication techniques, the notion of re-exports and some comments on regional accounting. *Economic Systems Research*, 13, 165–179.

Lahr, Michael and Louis De Mesnard. (2004) Biproportional techniques in input-output analysis: Table updating and structural analysis. *Economic Systems Research*, 16, 115-134.

Lahr, Michael L. and Benjamin H. Stevens. (2002) A study of the role of regionalization in the generation of aggregation error in regional input–output models. *Journal of Regional Science*, 42,477–507.

Lamonica, Giuseppe R. and Francesco M. Chelli. (2018) The performance of non-survey techniques for constructing sub-territorial input-output tables. *Papers in Regional Science*, 97, 1169–1202.

Lehtonen, Olli and Markku Tykkyläinen. (2014) Estimating regional input coefficients and multipliers: Is the choice of a non-survey technique a gamble? *Regional Studies*, 48, 382–399.

LeSage, James P. and Carlos Llano. (2016) A spatial interaction model with spatially structured origin and destination effects. In Roberto Patuellu and Giuseppe Arbia (eds), *Spatial Econometric Interaction Modelling*. Springer, pp. 171–197.

Lindall, Scott A., Douglas C. Olson, and Gregory S. Alward. (2006) "Deriving multi-regional models using the IMPLAN national trade flows model," *Journal of Regional Analysis and Policy,* 67–83.

Llano, Carlos, Almudena Esteban, Julian Pérez, and Antonio Pulido. (2010) Opening the interregional trade ''black box'': The C-intereg database for the Spanish economy (1995—2005), *International Regional Science Review*, 33, 302–337.

Lösch, August. (1954) *The Economics of Location.* (trans. H. H. Voglom) New Haven, Yale University.

McCann, Phil and John H. L. Dewhurst. (1998) Regional size, industrial location and input-output expenditure coefficients. *Regional Studies,* 32, 435–444.

Miller, Ronald E. and Peter D. Blair. (2009) *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press.

Oosterhaven, Jan. (1981) *Interregional Input-Output Analysis and Dutch Regional Policy Problems*. Aldershot, U.K.: Gower Publishing.

Ortega, F. and G. Peri. (2014) Openness and income: The roles of trade and migration. *Journal of International Economics*, *92*(2), 231–251.

Polachek, Solomon W. (1992) Conflict and trade: An economics approach to political international interactions, In Walter Isard and Charles H. Anderton (eds), *Economics of Arms Reduction and the Peace Process*. North-Holland, Amsterdam, pp. 89–120.

Polenske, Karen R. and Geoffrey J. D. Hewings. (2004) Trade and spatial economic interdependence. *Papers in Regional Science*, 83, 269–289.

Robinson, Dennis P. and Zuoming Liu. "The effects of interregional trade flow estimating procedures on multiregional social accounting matrix multipliers," *Journal of Regional Analysis & Policy,* 36, 94–114.

Robison, M. Hank and Jon R. Miller. 1988. Cross-hauling and nonsurvey input-output Models: Some lessons from small-area timber economies, *Environment and Planning A*, 20, 1523–1530.

Rodrigues, João and Michael Lahr. (2018) The reconciliation of multiple conflicting estimates: Entropy-based and axiomatic approaches. *Entropy*. 20(11): 815.

Round, Jeffrey I. (1983) Nonsurvey techniques: A critical review of the theory and the evidence. *International Regional Science Review*, 8, 189–212.

Sargento, Ana L.M. (2009) *Regional input-output tables and models: interregional trade estimation and input-output modelling based on total use rectangular tables* (Doctoral dissertation).

Sargento, Ana. L., Perdro N. Ramos, and Geoffrey J. D. Hewings. (2012) Inter-regional trade flow estimation through non-survey models: An empirical assessment. *Economic Systems Research*, 24, 173–193.

Sombart, Werner. (1916) *Der Moderne Kapitalismus: Historisch-systematische Darstellung des Gesamteuropäischen Wirtschaftslebens von Seinen Anfängen bis zur Gegenwart*. Verlag von Duncker & Humblot, München und Leipzig. https://play.google.com/books/reader?id=sMRaAAAAYAAJ&hl=en&pg=GBS.PR1

Stevens, Benjamin H. and Glynnis A. Trainer. (1976) The generation of error in regional input-output impact models. *Regional Science Research Institute Working Paper 76-A*. Regional Science Research Institute, Amherst MA.

Stevens, Benjamin H., George I. Treyz, David. J. Ehrlich, and James R. Bower. (1980) Estimation of regional purchase coefficients and their use in the construction of non-survey input-output tables, *RSRI Discussion Paper Series No. 119*, Regional Science Research Institute, Amherst, MA.

Stevens, Benjamin H., George I. Treyz, David. J. Ehrlich, and James R. Bower. (1983) A new technique for the construction of non-survey regional input-output models and comparison with two survey-based models, *International Regional Science Review*, 8, 271–286.

Stevens, Benjamin H., George I. Treyz, and Michael L. Lahr. (1989) On the comparative accuracy of RPC estimating techniques, in Ronald E. Miller, Karen R. Polenske, and Adam Z. Rose (eds.), *Frontiers of Input-Output Analysis*. New York: Oxford University Press, pp. 258–278.

Sun, Yue. (2018) *Commodity Based Freight Demand Modeling Framework using Structural Regression Model*. Unpublished PhD dissertation, University of California, Irvine.

Szyrmer, Janusz M. (1989) Trade-off between error and information in the RAS procedure, in: Ronald E. Miller, Karen R. Polenske and Adam Z. Rose (eds), *Frontiers of Input–Output Analysis*. New York: Oxford University Press, pp. 258–279

Timmer, Marcel. P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaitzen J. De Vries. (2015) An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics*, 23, 575–605.

Többen, Johannes and Tobias H. Kronenberg. (2015) Construction of multi-regional input–output tables using the CHARM method. *Economic Systems Research*, 27, 487–507

Treyz, George I. and Benjamin H. Stevens. (1981) Location analysis for multiregional modeling. In F.G. Adams and N.M. Glickman (eds.), *Modeling the Multiregional Economic System*. Lexington, MA: Lexington Books, pp. 75–87.

Treyz, George I. and Benjamin H. Stevens. (1985) The TFS regional modelling methodology. *Regional Studies*, 19, 547–562.

Wood, Richard, Konstantin Stadler, Tatyana Bulavskaya, Stephan Lutter, Stefan Giljum,Arjan de Koning, Jeroen Kuenen, Helmut Schütz, José Acosta-Fernández, Arkaitz Usubiaga, Moana Simas, Olga Ivanova, Jan Weinzettel, Jannick H. Schmidt, Stefano Merciai, and Arnold Tukker. (2015) Global sustainability accounting—developing EXIOBASE for multi-regional footprint analysis. *Sustainability*, 7, 138–163.

Zhang, Lianfa, Jianquan Cheng, and Cheng Jin. (2019) Spatial interaction modeling of OD flow data: Comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR). *ISPRS International Journal of Geo-Information*, 8, 220. https://www.mdpi.com/2220-9964/8/5/220/pdf.

Appendix

Table A.1: Descriptive Statistics of Basic Study Variables across the European Union, 2014

	RPC	Land area km^2	Hotel room-nights/km^2	Weight/value ton/$1,000	SLQ	S/D	Share of EU jobs
Mean	0.4097	159,898.7	1,508.3	1.0908	0.7449	0.7821	0.0357
Median	0.3901	81,372.5	518.0	0.3050	0.8868	0.9300	0.0181
St. Error	0.2936	167,665.4	4,673.8	2.2444	0.2964	0.2849	0.0469
Min	0.0000	316.0	54.9	0.0000	0.0000	0.0000	0.0009
Max	0.9844	640,679.0	25,614.9	18.9000	1.0000	1.0000	0.1882

Table A.2: Correlation Matrix

	RPC	Land area km^2	Hotel room-nights/km^2	Weight/Value ton/$1,000	SLQ	S/D	Share of EU jobs
RPC	1.00	.29	-.07	.02	.37	.57	.24
Land area		1.00	-.20	-.02	.09	.21	.68
hotel room-nights/km^2			1.00	-.08	-.13	-.17	-.13
Weight/value				1.00	.00	-.12	-.06
SLQ					1.00	.61	.03
S/D						1.00	.20
Share of EU jobs							1.00