Estimation of Beta and Sigma in Trade Flow Equations

February 2019

Interregional trade flows provide important information to analyze regional economic activities. While actual interregional trade flow data are not available in the U.S. and other countries, REMI utilizes the gravity model and regional supply and demand data to estimate the trade flows. The basic hypothesis is that the flows of commodities and services between regions are determined by the total output from the origin, the total demand in the destination, and the distance between two regions. This study re-estimates the coefficients in the trade flow model system with updated panel data at both state and county level from 2001 to 2016 by industry.

The Trade Flow Equation

Trade flows between two regions do not only depend on the supply and demand in each region, but also relate to the distance between the two regions. REMI uses a doubly constrained gravity model to estimate interregional trade flows. The trade flow equation is in the following form:

$$T_{ij} = A_i \cdot B_j \cdot Q_i \cdot D_j \cdot f(d_{ij}) \qquad (1)$$

where

T_{ij} = the trade flow from the region of origin i to the region of destination j;

i and j denotes the origin region and the destination region, and $1 \leq i \leq n$, $1 \leq j \leq n$ where n is the number of regions under consideration;

Q_i = the total output from region i;

D_j = the total demand in region j;

$f(d_{ij})$ is a distance decay function which is hypothesized to summarize all effects of spatial interaction:

$$f(d_{ij}) = d_{ij}^{-\beta} \tag{2}$$

and β is the distance decay parameter to be estimated;

A_i and B_j are interrelated balancing factors to ensure that all output from region i are delivered and all demand of region j are met, such that

$$Q_i = \sum_{j=1}^{n} T_{ij} \tag{3}$$

$$D_j = \sum_{i=1}^{n} T_{ij} \tag{4}$$

Plugging Equation (1) into to (3) and (4) we get

$$Q_i = A_i Q_i \sum_{j=1}^{n} B_j D_j f(d_{ij}) \tag{5}$$

$$D_j = B_j D_j \sum_{i=1}^{n} A_i Q_i f(d_{ij}) \tag{6}$$

Then Equations (5) and (6) can be simplified into

$$1 = A_i \sum_{j=1}^{n} B_j D_j f(d_{ij}) \tag{7}$$

$$1 = B_j \sum_{i=1}^{n} A_i Q_i f(d_{ij}) \tag{8}$$

or

$$A_i = (\sum_{j=1}^{n} B_j D_j f(d_{ij}))^{-1} \tag{9}$$

$$B_j = (\sum_{i=1}^{n} A_i Q_i f(d_{ij}))^{-1} \tag{10}$$

If $T_{ij},\ Q_i,\ and\ D_j$ are all known, the system could classically solved by non-linear least squares methodology. However, historical data on interregional trade flows are not available in the U.S. as well as other countries, which determines the equation cannot be solved with a classic econometric methodology. To circumvent the inadequate data problem, we have to use alternative methodology to estimate for β.

Estimating β

In the trade flow equation (1), the distance is inversely related to the trade flow between two regions. The distance decay parameter β determines how significant the effect of distance is on the flows of commodities and services. The higher the value of β is, the more the trade flows would depend on the distance between two regions. In the absence of trade flow data, we first assume $A_i = B_j = 1$, and estimates β by minimizing the sum of squares of differences between actual changes in output in the region of origin and the estimated changes in that output. The objective function is set up as

$$F(\beta) = \sum_{t=2}^{NT} \sum_{i=1}^{n} \left[\frac{Q_{i,t} - Q_{i,t-1}}{Q_{i,t} - 1} - \frac{\hat{Q}_{i,t} - \hat{Q}_{i,t-1}}{\hat{Q}_{i,t-1}} \right]^2 \tag{11}$$

where

i stands for the origin region i;

t denotes time period t;

NT is the number of time periods;

n is the total number of regions, $n = 3086$ for U.S. county level data set, while n=51 for U.S. state level data set;

Q_i^t is the actual output from region i at time t;

$\hat{Q}_{i,t}$ is the estimated output from region *i* at time *t*. The estimated output is calculated by summing up the demand for output from region i, from different destination markets. The function is in the following form:

$$\hat{Q}_{i,t} = \sum_j^n D_{j,t} f\left(d_{ij}\right) \tag{12}$$

Then we estimate β using an iterative grid search approach. A closed interval is searched for *β* that locally minimizes the value of the objective function *F(β)*. β is selected as a damping factor, with $\beta > 1$ scaling the damping distance factor so that greater distance impede trade flow more and $\beta < 1$ providing less damping of trade flow for large distance. For example, $\beta < 1$ would be applicable in the case of equipment manufacturing where trade flow to span a greater distance than say, food services where we would expect $\beta > 1$. Also, it is reasonable to expect the values of β's change over time for some industries due to consumption pattern changes and technology development. For example, the continuing growth of e-commerce makes distance a less important factor in retail business. At the same time, the value of β also varies across regions because of differences in accessible resources between regions. Thus, it is necessary to use panel data to study trade flows and keep the estimation up to date.

Once the value of β is determined, we estimate the balancing factors A_i and B_j through another iterative searching process. We approach the problem by rewriting Equation (7) and (8) into the following objective functions:

$$f_1\left(A_i, B_j\right) = \left(A_i \sum_{j=1}^n B_j D_j f(d_{ij}) - 1\right)^2 \tag{13}$$

$$f_2\left(A_i, B_j\right) = \left(B_j \sum_{i=1}^n A_i Q_i f(d_{ij}) - 1\right)^2 \tag{14}$$

We start the estimation process from giving B_j an initial value of one, and then A_i is calculated based on Equation (9). Similarly, B_j is generated based on the newly calculated A_i according to Equation (10). This procedure of iteration is repeated until the summation of $f_1(A_i, B_j)$ and $f_2(A_i, B_j)$ reaches a threshold value, which is determined as 0.01 in our implementation. When a convergence cannot be attained for some industries, an iteration maximum of 500 is applied to terminate the loop. After A_i and B_j are determined, we estimate the price elasticity of demand σ for each industry.

Estimating σ

The price elasticity of demand σ measures the responsiveness of demand for output from a supplying region to a price change in this region. The estimates of σ is derived from interregional trade flows. For a given industry, the change in output from supplying region *i* can be expressed as

$$\frac{\Delta Q_{i,t}}{Q_{i,t-1}} = \frac{\Delta EDS_{i,t}}{EDS_{i,t-1}} + \frac{\Delta EDM_{i,t}}{EDM_{i,t-1}} \tag{15}$$

where $\frac{\Delta EDS_{i,t}}{EDS_{i,t-1}}$ is the change in demand that results from a shift of the demand curve in the demanding regions, and $\frac{\Delta EDM_{i,t}}{EDM_{i,t-1}}$ is the change in demand that results from the movement along the demand curve or by changes in price in the supplying region.

The first determinant $\frac{\Delta EDS_{i,t}}{EDS_{i,t-1}}$ can be written as

$$\frac{\Delta EDS_{i,t}}{EDS_{i,t-1}} = \frac{\sum_j^N s_{ij,t} \Delta D_{j,t}}{\sum_j^N s_{ij,t-1} D_{j,t-1}} = \frac{\sum_j^N s_{ij,t}(D_{j,t} - D_{j,t-1})}{Q_{i,t-1}} \tag{16}$$

where

$S_{ij,t}$ is the share of domestic demand in region *j* that is supplied by output from region i, and can be calculated using the trade flow and demand data:

$$S_{ij,t} = \frac{T_{ij,t}}{D_{j,t}} = A_{i,t} B_{j,t} Q_{i,t} f\left(d_{ij}\right) \tag{17}$$

The second determinant $\frac{\Delta EDM_{i,t}}{EDM_{i,t-1}}$ is formulated as

$$\frac{\Delta EDM_{i,t}}{EDM_{i,t-1}} = (1-\sigma)\left[\frac{\Delta\Omega_{i,t}}{\Omega_{i,t-1}} - \frac{\sum_j^N S_{ij,t} D_{j,t}(\sum_j^n S_{ij,t}(\frac{\Omega_{i,t}-\Omega_{i,t-1}}{\Omega_{i,t-1}}))}{Q_{i,t-1}}\right] \tag{18}$$

where

$\Omega_{i,t}$ is the cost of production in the supplying region *i* at time t;

σ is the price elasticity of demand to be estimated.

Therefore, Equation (15) can be expanded into:

$$\frac{\Delta Q_{i,t}}{Q_{i,t-1}} - \frac{\sum_j^N S_{ij,t}(D_{j,t}-D_{j,t-1})}{Q_{i,t-1}} = (1-\sigma)\left[\frac{\Delta\Omega_{i,t}}{\Omega_{i,t-1}} - \frac{\sum_j^N S_{ij,t} D_{j,t}(\sum_j^n S_{ij,t}(\frac{\Omega_{i,t}-\Omega_{i,t-1}}{\Omega_{i,t-1}}))}{Q_{i,t-1}}\right] \tag{19}$$

in which all variables are known or can be calculated except σ. The left-hand-side of the equation serves as the dependent variable, and $\left[\frac{\Delta\Omega_{i,t}}{\Omega_{i,t-1}} - \frac{\sum_j^N S_{ij,t} D_{j,t}(\sum_j^n S_{ij,t}(\frac{\Omega_{i,t}-\Omega_{i,t-1}}{\Omega_{i,t-1}}))}{Q_{i,t-1}}\right]$ on the right-hand-side can be treated as the independent variable. An OLS regression is applied to estimate the unknown coefficient $(1-\sigma)$.

Once σ is determined, it will be used with β to calculate the distance deterrence elasticity in region i as $\eta_i = \frac{\beta_i}{\sigma_i - 1}$ which is used in the Intermediate Input Access Index. Thus, σ needs to be greater than 1 to keep η>0.

Data

We use panel data of 51 states and 3086 counties from 2001 to 2016 to estimate trade flows for each industry sector. The estimation is performed for the 66 private industry sectors from the REMI model system. The regional output and demand data are extracted from the REMI PI+ V2.2 model and are normalized to the U.S. level to make sure the regional level summation equals to the national total in each sector. On the other hand, the output are adjusted for exports and the demand are adjusted for imports to keep the supply and demand in this study domestic. We calculated the distance between two regions using the haversine formula, which gives the great circle distance between two points on earth using their latitudes and longitudes. The cost of production data is also from the REMI PI+ V2.2 model. Due to data anomalies, we have eliminated the outlier observations where the annual change in demand and output exceeds a certain threshold.

Result

Table 1 presents the estimation results of β and σ for the 66 REMI private sectors with comparison to the current model parameter values from PI+ V2.2. The results confirm our hypothesis that distance do matter. The change in output from a region is affected by the change in demand in the respective markets. The extent of dependence depends on the distance between the location of the supplier and the markets being served, and varies by industry.

A higher β value indicates that greater distance impede trade flow more for the industry. The highest β values are found in Construction, Repair and Maintenance, and Food Services and Drinking Places, which are all service industries where trade flows tend to span shorter distances. Management of Companies and Enterprises has the lowest β value, which could be attributed to the nature of the industry. For establishments that hold securities or other equity interest in and make administrative decisions for companies and enterprises, the effect of distance on the trade flows would be negligible.

The value of σ measures the responsiveness of demand for output from a region to a price change in the supplying region. Industries with more stable demand, higher transportation costs or less price fluctuations are likely to have lower σ values. For example, Oil and Gas Extraction provides non-substitutable resources for production and residential use. This industry is highly restricted by the location of found reserves and production facilities. In addition, the oil and gas price often follows the same trend. Limited supply and less price differences between regions make the demand for oil and gas relatively inelastic. By contrast, Apparel Manufacturing has a higher σ value as expected. As its output is neither indispensable nor expensive to ship, the demand would more relatively more elastic to price changes.

Table 1: Beta and Sigma Estimation Results

#	Industry	New Estimates		Current Values	
		Beta	Sigma	Beta	Sigma
1	Forestry and Logging; Fishing, hunting, and trapping	1.73	2.61	2.21	2.14
2	Support activities for agriculture and forestry	1.81	1.85	2.21	2.14
3	Oil and gas extraction	1.15	1.41	3.37	1.47
4	Mining (except oil and gas)	2.07	1.80	3.37	1.47
5	Support activities for mining	2.01	1.51	3.37	1.47
6	Utilities	2.71	2.05	2.71	2.38
7	Construction	4.72	2.79	2.91	3.08
8	Wood product manufacturing	2.68	3.08	2.05	2.29
9	Nonmetallic mineral product manufacturing	1.20	1.93	1.48	1.67
10	Primary metal manufacturing	1.69	1.66	1.09	1.66
11	Fabricated metal product manufacturing	1.23	1.91	1.04	1.99
12	Machinery manufacturing	1.15	1.97	1.34	2.15
13	Computer and electronic product manufacturing	1.63	3.25	1.88	4.62
14	Electrical equipment, appliance, and component manufacturing	0.53	2.17	0.79	2.21
15	Motor vehicles, bodies and trailers, and parts manufacturing	1.81	2.51	1.71	2.41
16	Other transportation equipment manufacturing	1.86	1.88	0.75	1.79
17	Furniture and related product manufacturing	2.77	2.38	1.73	2.18
18	Miscellaneous manufacturing	1.03	2.11	0.92	1.72
19	Food manufacturing	1.88	3.67	1.26	2.55
20	Beverage and tobacco product manufacturing	2.83	1.73	3.15	1.64
21	Textile mills; Textile product mills	1.90	4.30	1.88	4.62
22	Apparel manufacturing; Leather and allied product manufacturing	1.77	4.67	1.88	4.62
23	Paper manufacturing	1.49	2.31	1.10	2.13
24	Printing and related support activities	1.22	2.02	1.39	2.12
25	Petroleum and coal products manufacturing	1.43	3.86	1.88	4.62
26	Chemical manufacturing	1.01	1.91	0.78	2.00
27	Plastics and rubber products manufacturing	1.95	2.01	1.47	1.84
28	Wholesale trade	1.68	2.38	1.36	1.64
29	Retail trade	2.94	3.82	2.54	3.33
30	Air transportation	1.91	1.84	1.13	1.64
31	Rail transportation	0.87	2.47	0.90	2.29
32	Water transportation	1.59	2.57	2.28	2.79
33	Truck transportation	1.69	2.38	1.30	3.61
34	Couriers and messengers	1.28	1.93	1.34	3.16
35	Transit and ground passenger transportation	2.50	2.51	2.28	2.79
36	Pipeline transportation	2.89	2.51	2.28	2.79
37	Scenic and sightseeing transportation; Support activities for transportation	1.57	2.53	1.91	1.85
38	Warehousing and storage	2.89	2.51	2.28	2.79
39	Publishing industries, except Internet	0.71	2.02	0.97	2.39
40	Motion picture and sound recording industries	1.91	2.66	0.82	2.93

		New Estimates		Current Values	
#	**Industry**	Beta	Sigma	Beta	Sigma
41	Data processing, hosting, and related services; Other information services	1.76	1.35	2.13	1.17
42	Broadcasting, except Internet	1.27	1.27	1.58	1.31
43	Telecommunications	1.10	1.71	1.44	1.58
44	Monetary authorities – central bank; Credit intermediation and related activities	0.87	2.12	1.83	1.92
45	Securities, commodity contracts, other investments; Funds, trusts, other vehicles	1.36	3.15	1.37	3.52
46	Insurance carriers and related activities	1.08	1.55	1.09	1.22
47	Real estate	2.69	2.07	1.33	1.24
48	Rental and leasing services; Lessors of nonfinancial intangible assets	3.43	2.76	1.59	1.30
49	Professional, scientific, and technical services	2.21	1.73	1.87	1.62
50	Management of companies and enterprises	0.51	1.46	2.83	2.93
51	Administrative and support services	1.92	1.84	1.59	1.46
52	Waste management and remediation services	1.50	1.82	1.62	1.83
53	Educational services; private	1.13	1.38	1.03	1.30
54	Ambulatory health care services	2.18	1.82	1.97	1.97
55	Hospitals; private	1.63	1.79	1.11	1.46
56	Nursing and residential care facilities	1.35	1.82	1.23	2.14
57	Social assistance	1.89	2.04	2.11	2.21
58	Performing arts, spectator sports, and related industries	1.42	2.10	1.68	2.53
59	Museums, historical sites, and similar institutions	0.68	1.97	0.81	1.76
60	Amusement, gambling, and recreation industries	2.63	2.13	1.04	1.65
61	Accommodation	3.52	3.54	1.93	2.86
62	Food services and drinking places	4.35	2.22	1.93	2.86
63	Repair and maintenance	4.40	2.64	1.74	2.52
64	Personal and laundry services	3.00	2.23	1.74	2.52
65	Religious, grantmaking, civic, professional, and similar organizations	2.19	1.76	1.74	2.52
66	Private households	1.45	2.43	1.74	2.52