Technical Documentation for Estimating Beta and Sigma in Trade Flow Equations

April 9, 2014

Trade Flow Equation

In their paper, "Inter and Intra-State Trade Flow Estimation in the Absence of Trade Flow Data," George Treyz, Fred Treyz, and Omar El-Gayar explain the motivation for using the double constrained gravity model as follows to model interregional commodity flow

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

where

$1 \leq i \leq n$, $1 \leq j \leq n$ where n is the number of regions under consideration;

T_{ij} is the trade flow between the region of origin i and the region of destination j;

Q_i is output in region i;

D_j is demand in region j;

$f(d_{ij}) = d_{ij}^{-\beta}$ is a distance decay function which is hypothesized to summarized all effects of spatial interaction[1], β is the distance decay parameter to be estimated;

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

In this setup of Equations when $T_{ij},\ Q_i,\ and\ D_j$ are all known, we would have n^2 Equations, one for each T_{ij}, and $2n + 1$ unknowns, $A_i,\ B_j$ and β. This type of overdetermined system is

[1] Distance between counties are calculated using the latitude and longitude of county/state centroids.

classically solved by non-linear least squares regression, which finds A_i, $B_j$and β that give a best fit Equation where Q_i and D_j are the explanatory variables and T_{ij} are the response variables[2].

However, historical data on interregional commodity flow are usually not readily available in the U.S. as well as other countries. We have to reformat the Equation to estimate β.

Estimating Beta

In the absence of real trade flow data T_{ij}, Equation (1) is not sufficient to form a model of trade flow. In both origin and destination regions, we could include balancing factors.

$$Q_i = \sum_{j=1}^{n} T_{ij} \qquad (2)$$

$$D_j = \sum_{i=1}^{n} T_{ij} \qquad (3)$$

Equation (2) ensures that all output is delivered, and Equation (3) ensures that all demand is met. Plugging in the expression for T_{ij} of Equation (1) into to the balancing Equations we get

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

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

And simplifying gives

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

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

Or

$$A_i = (\sum_{j=1}^{n} B_j D_j f(d_{ij}))^{-1} \qquad (4)$$

[2] Note that taking the product of A and B conveniently gives us n^2 coefficients with 2n variables whereas taking n coefficients C would produce and under-determined system.)

$$B_j = (\sum_{i=1}^{n} A_i Q_i f(d_{ij}))^{-1} \qquad (5)$$

From these balancing Equations we have $2n$ Equations and $2n + 1$ unknowns. Therefore, we have a degree of freedom. We can choose β such 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 ($\beta < 1$ would be applicable in the case of manufacturing for example where we would expect trade flow to span a greater distance than say, retail where we would expect $\beta > 1$).

Facing the under-determined model system composed of Equation (4) and (5), we choose to estimate β first by assuming $A_i = B_j = 1$. The objective function for estimating β is formulated as the sum of squares of differences between actual changes in output in the region of origin and the estimated changes of that output. The estimated changes of output is determined by the changes in demand in the demanding regions subject to the distance decay function, a function of distance and β. The objective function for estimating β is as follows.

$$F(\beta) = \sum_{t=2}^{NT} \sum_{i=1}^{n} [\frac{Q_i^t - Q_i^{t-1}}{Q_i^{t-1}} - \frac{QHAT_i^t - QHAT_i^{t-1}}{QHAT_i^{t-1}}]\text{^}2 \qquad (6)$$

where

i stands for region i;

t is for time period t;

NT is the number of years under consideration;

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 output for region i at time t;

$QHAT_i^T = D_j^T * f(d_{ij})$, $QHAT_i^T$ is the estimated output expressed by demand for region i at time t, where D_j^t is the demand for region j at time t and $f(d_{ij})$ is defined the same as in Equation (1).

β is estimated using an iterative grid search approach. A closed interval is searched for β that locally minimizes the objective function $F(\beta)$, by dividing the interval into points and evaluating the objective function at each point.

Once we determine a value for , we can then choose A_i and B_j such that all output is delivered and all demand is met. The problem now is to iteratively find a solution for Equation (4) and (5). We approach this problem by minimizing Equation (7) and (8).

$$f_1(A_i, B_j) = \left(A_i \sum_{j=1}^{n} B_j D_j f(d_{ij}) - 1\right)^2 \qquad (7)$$

$$f_2(A_i, B_j) = \left(B_j \sum_{i=1}^{n} A_i Q_i f(d_{ij}) - 1\right)^2 \qquad (8)$$

We start from an initial value of 1 for all B_j and then calculate A_i based on Equation (4). New B_j is generated based on the newly calculated A_i according to Equation (5). This procedure of iteration is run until the summation of $f_1(A_i, B_j)$ and $f_2(A_i, B_j)$ is less than a threshold value, which is determined as 0.01in our implementation. For some industries at the county level, we cannot reach a convergence such that the objective function go below the threshold. In this situation, we set an upper limit for the iterations counts (usually we use 600 as the iteration limit) to terminate the loop. After we have A_i, B_j and β ready, we can estimate σ for each industry.

Data

The data set used for the estimation is a panel data set of output and demand for 3086 counties and 51 states for the year from 1990 through 2011. This data set is generated by the REMI model building system. There are 66 sectors included in the data set. The 66 sectors are private sectors used in PI$^+$ 1.4 and newer version as listed in Appendix 1. Output and demand for each sector are normalized to the U.S. national level to make sure the county/state level summation equals to the national total. In addition, output are adjusted for exports to get domestic output, and demand are adjusted for imports to get demand that is satisfied by domestic supply. The detailed procedure to adjust output and demand is elaborated in Appendix 2.

Estimating Sigma

The price elasticity of demand (Sigma) is derived from interregional trade flows. For a given industry, it is hypothesized that the change in output in the supplying region is determined by two factors[3]: (1) the change in demand that results from a shift of the demand curve in the demanding regions, which is identified as $\frac{\Delta EDS_{i,t}}{EDS_{i,t}}$, and (2) the change in demand that results from the movement along the demand curve or by change in price in the supplying region, which is expressed as $\frac{\Delta EDM_{i,t}}{EDM_{i,t}}$. The change in output for industry i at time t is formulated as Equation (9)

$$\frac{\Delta Q_{i,t}}{Q_{i,t}} = \frac{\Delta EDS_{i,t}}{EDS_{i,t}} + \frac{\Delta EDM_{i,t}}{EDM_{i,t}} \qquad (9)$$

The rate of change in demand due to a shift of the demand curve for the region i can be expressed as

$$\frac{\Delta EDS_{i,t}}{EDS_{i,t}} = \frac{\sum_j^N (S_{ij,t}(\Delta D_{j,t}))}{\sum_j^N (S_{ij,t}(D_{j,t}))} = \frac{\sum_j^N (S_{ij,t}(D_{j,t} - D_{j,t-1}))}{Q_{i,t}} \qquad (10)$$

Where $S_{ij,t} = A_{i,t}B_{j,t}Q_{i,t}F(d_{ij}) = \frac{T_{ij,t}}{D_{j,t}}$ is the share of domestic demand in region j supplied by region i.

The rate of change in demand resulting from price change or from the movement along the demand curve

$$\frac{\Delta EDM_{i,t}}{EDM_{i,t}} = (1-\sigma)\left[\frac{\Delta \Omega_{i,t}}{\Omega_{i,t}} - \frac{\sum_j^N S_{ij,t} D_{j,t} (\sum_j^n S_{ij,t} (\frac{\Omega_{i,t} - \Omega_{i,t-1}}{\Omega_{i,t}}))}{Q_{i,t}}\right] \qquad (11)$$

where

σ is the price elasticity of demand we need to estimate;

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

all other variables are the same as defined in Equation (1).

By plugging Equation (10) and (11) back into Equation (9), we have the function to estimate sigma expressed as

[3] The following discussion on Equation (9) through Equation (12) is based on notes from Omar El-Gayar's binder of computer runs and on the Fortran computer program that implemented the formulas from these notes.

$$\frac{\Delta Q_{i,t}}{Q_{i,t}} - \frac{\sum_j^N (S_{ij,t}(D_{j,t} - D_{j,t-1}))}{Q_{i,t}} = (1-\sigma) * [\frac{\Delta \Omega_{i,t}}{\Omega_{i,t}} - \frac{\sum_j^N S_{ij,t} D_{j,t} (\sum_j^n S_{ij,t} (\frac{\Omega_{i,t} - \Omega_{i,t-1}}{\Omega_{i,t}}))}{Q_{i,t}}] \quad (12)$$

An OLS regression approach is used to estimate the unknown (1- σ).

The Dependent variable is

$$Y = \frac{\Delta Q_i}{Q_i} - \frac{\sum_j^N (S_{ij,t}(D_{j,t} - D_{j,t-1}))}{Q_{i,t}}$$

The Independent variable is

$$X = \frac{\Delta \Omega_{i,t}}{\Omega_{i,t}} - \frac{\sum_j^N S_{ij,t} D_{j,t} (\sum_j^n S_{ij,t} (\frac{\Omega_{i,t} - \Omega_{i,t-1}}{\Omega_{i,t}}))}{Q_{i,t}}$$

The dataset used in the regression approach is the same as from what we use in estimating β. The results for β and σ estimations are presented in Table 1. Using the estimation for β and σ, we will calculate distance deterrence elasticity for region i using the identity $\eta_i = \frac{\beta_i}{\sigma_i - 1}$, which will be used in the Intermediate Input Access Index.

Table 1: Beta and Sigma Estimation Results

Index	Industry	Beta	Sigma
1	Forestry et al.	2.211	2.138589
2	Agriculture	2.211	2.138589
3	Oil, gas extraction	3.37	1.471356
4	Mining (except oil, gas)	3.37	1.471356
5	Support activities for mining	3.37	1.471356
6	Utilities	2.71	2.375559
7	Construction	2.911	3.078605
8	Wood product mfg	2.05	2.291857
9	Nonmetallic mineral prod mfg	1.48	1.674972
10	Primary metal mfg	1.085	1.663196
11	Fabricated metal prod mfg	1.04	1.993006
12	Machinery mfg	1.34	2.154506
13	Computer, electronic prod mfg	1.881	4.623089
14	Electrical equip, appliance mfg	0.794	2.208143
15	Motor vehicle mfg	1.711	2.411177
16	Transp equip mfg. exc. motor veh	0.753	1.786376
17	Furniture, related prod mfg	1.726	2.176149

18	Miscellaneous mfg	0.922	1.724387
19	Food mfg	1.26	2.545166
20	Beverage, tobacco prod mfg	3.151	1.638189
21	Textile mills; Textile product mills	1.881	4.623089
22	Apparel manufacturing; Leather and allied product manufacturing	1.881	4.623089
23	Paper manufacturing	1.102	2.132616
24	Printing and related support activities	1.39	2.117563
25	Petroleum and coal products manufacturing	1.881	4.623089
26	Chemical manufacturing	0.783	1.997443
27	Plastics and rubber products manufacturing	1.47	1.844665
28	Wholesale trade	1.36	1.641848
29	Retail trade	2.54	3.33007
30	Air transportation	1.125	1.643114
31	Rail transportation	0.9	2.292143
32	Water transportation	2.28	2.790666
33	Truck transportation	1.3	3.610976
34	Couriers and messengers	1.343	3.161831
35	Transit and ground passenger transportation	2.28	2.790666
36	Pipeline transportation	2.28	2.790666
37	Scenic and sightseeing transportation; Support activities for transportation	1.912	1.850184
38	Warehousing and storage	2.28	2.790666
39	Publishing industries, except Internet	0.968	2.388799
40	Motion picture and sound recording industries	0.82	2.929498
41	Internet publishing and broadcasting; ISPs, search portals, and data processing; Other information services	2.13	1.172518
42	Broadcasting, except Internet	1.58	1.308631
43	Telecommunications	1.44	1.57901
44	Monetary authorities – central bank; Credit intermediation and related activities; Funds, trusts, & other financial vehicles	1.83	1.92267
45	Securities, commodity contracts, investments	1.37	3.517835
46	Insurance carriers and related activities	1.09	1.221006
47	Real estate	1.33	1.238562
48	Rental and leasing services; Lessors of nonfinancial intangible assets	1.59	1.298617
49	Professional, scientific, and technical services	1.87	1.618773
50	Management of companies and enterprises	2.83	2.925342
51	Administrative and support services	1.59	1.464455
52	Waste management and remediation services	1.62	1.826976
53	Educational services	1.03	1.303824
54	Ambulatory health care services	1.97	1.96524
55	Hospitals	1.11	1.464305
56	Nursing and residential care facilities	1.23	2.135119
57	Social assistance	2.11	2.20673
58	Performing arts and spectator sports	1.68	2.533518

59	Museums, historical sites, zoos, and parks	0.812	1.758708
60	Amusement, gambling, and recreation	1.04	1.654837
61	Accommodation	1.93	2.859222
62	Food services and drinking places	1.93	2.859222
63	Repair and maintenance	1.74	2.520985
64	Personal and laundry services	1.74	2.520985
65	Membership associations and organizations	1.74	2.520985
66	Private households	1.74	2.520985

Appendix 1: Summary of Industries for PI+ v1.4

Index	Industry name	NAICS	Index	Industry name	NAICS
1	Forestry and logging; Fishing, hunting, and trapping	113,114	34	Couriers and messengers	492
2	Agriculture and forestry support activities	115	35	Transit and ground passenger transportation	485
3	Oil and gas extraction	211	36	Pipeline transportation	486
4	Mining (except oil and gas)	212	37	Scenic and sightseeing transportation; Support activities for transportation	487,488
5	Support activities for mining	213	38	Warehousing and storage	493
6	Utilities	22	39	Publishing industries, except Internet	511
7	Construction	23	40	Motion picture and sound recording industries	512
8	Wood product manufacturing	321	41	Internet publishing and broadcasting; ISPs, search portals, and data processing; Other information services	518,519
9	Nonmetallic mineral product manufacturing	327	42	Broadcasting, except Internet	515
10	Primary metal manufacturing	331	43	Telecommunications	517
11	Fabricated metal product manufacturing	332	44	Monetary authorities – central bank; Credit intermediation and related activities; Funds, trusts, & other financial vehicles	521,522,525
12	Machinery manufacturing	333	45	Securities, commodity contracts, investments	523
13	Computer and electronic product manufacturing	334	46	Insurance carriers and related activities	524
14	Electrical equipment and appliance manufacturing	335	47	Real estate	531
15	Motor vehicles, bodies and trailers, and parts manufacturing	3361-3363	48	Rental and leasing services; Lessors of nonfinancial intangible assets	532,533

16	Other transportation equipment manufacturing	3364-3369	49	Professional, scientific, and technical services	54
17	Furniture and related product manufacturing	337	50	Management of companies and enterprises	55
18	Miscellaneous manufacturing	339	51	Administrative and support services	561
19	Food manufacturing	311	52	Waste management and remediation services	562
20	Beverage and tobacco product manufacturing	312	53	Educational services	61
21	Textile mills; Textile product mills	313,314	54	Ambulatory health care services	621
22	Apparel manufacturing; Leather and allied product manufacturing	315,316	55	Hospitals	622
23	Paper manufacturing	322	56	Nursing and residential care facilities	623
24	Printing and related support activities	323	57	Social assistance	624
25	Petroleum and coal products manufacturing	324	58	Performing arts and spectator sports	711
26	Chemical manufacturing	325	59	Museums, historical sites, zoos, and parks	712
27	Plastics and rubber products manufacturing	326	60	Amusement, gambling, and recreation	713
28	Wholesale trade	42	61	Accommodation	721
29	Retail trade	44,45	62	Food services and drinking places	722
30	Air transportation	481	63	Repair and maintenance	811
31	Rail transportation	482	64	Personal and laundry services	812
32	Water transportation	483	65	Membership associations and organizations	813
33	Truck transportation	484	66	Private households	814

Appendix 2: Adjustment for Output and Demand

- **Variable Specification**

  DJ_i : demand for region i

  Q_i: output for region i

  DU: national demand

  MU: national import

QU: national supply

XU: national export

$IDEM_i$: Intermediate demand

$IDEM1_i$: Intermediate demand not including industry's demand for itself

- **Normalize Q and D with respect to national level.**

$$T_D = \sum_{i=1}^{NR} DJ_i$$

$$T_Q = \sum_{i=1}^{NR} Q_i$$

$$Q_i = Q_i * QU/T_Q$$

$$DJ_i = DJ_i * DU/T_D$$

$$IDEM_i = IDEM_i * DU/T_D$$

$$IDEM1_i = IDEM1_i * DU/T_D$$

- **Adjust for Import and Export**

$$Q_i = \left(1 - \frac{XU}{QU}\right) * Q_i$$

$$DJ_i = \left(1 - \frac{MU}{DU}\right) * DJ_i$$

$$IDEM_i = \left(1 - \frac{MU}{DU}\right) * IDEM_i$$

$$IDEM1_i = \left(1 - \frac{MU}{DU}\right) * IDEM1_i$$

$$DJ1_i = DJ_i - (IDEM_i - IDEM1_i)$$

- **Replace D with demand not supplying itself, and replace output with the proportion of output going into intermediate demand**

$$T_D = \sum_{i=1}^{NR} DJ_i$$

$$T_Q = \sum_{i=1}^{NR} Q_i$$

$$Q_i = Q_i * T_D / T_Q$$

$$DJ_i = DJ1_i$$