Volume LIII No. 1

ISSN 1655-1516
June 2016

The Philippine Review of Economics

ARTICLES IN THIS ISSUE

A joint publication of the
University of the Philippines School of Economics
and the
Philippine Economic Society

The Philippine Review of Economics
Vol. LIII No. 1, June 2016 pp. 87-96

Measuring economic potential via the gravity model of trade

Nikkin L. Beronilla*, Patrocinio Jude H. Esguerra*, and Jamir Ocampo*

Trade is a good gauge of economic activity. Given the economic and geographical attributes of a place, one can assess the likely levels of trading activity, and this may be used as an indicator of the relative economic potential among the areas. However, trade estimates are available only at the regional level of disaggregation, despite the fact that planning by local governments and even by line agencies of the central government happens mostly at the sub-regional level. In this research, we fill this data gap by estimating trade at town or city level via a gravity model of trade. Normally sub-regional trade is estimated using Ordinary Least Regression, but this research uses Poisson Regression which is better at handling zero trade values without transformation. The research results have been applied by a number of government agencies in their respective programs.

JEL classification: C01, F17
Keywords: gravity model of trade, Poisson regression, economic potential for growth

1. Introduction

Trade is a good gauge of economic activity. Given the economic and geographical attributes of a place, one can assess the likely levels of trading activity; this may be used as an indicator of the relative economic potential among the areas. The level of trade estimates may aid in the allocation of resources and in the design of incentives. For example, if the goal is to create economic opportunities for the poor outside of the traditional growth centers, trade estimates could identify those towns[1] with high relative economic potential compared to

[1] The term "town" is used interchangeably with "municipality" or "city" in this paper.

other poor towns. The government can then move towards addressing barriers to product, finance, and labor market transactions that prevent the assessed potential from being realized. Tax incentives could be offered to areas with low trade estimates so that they can be embedded into the circuits of commerce.

Despite its important uses, the research on economic potential for the Philippine towns is patchy. One study is found in "Behind the Veil of Conflict", a paper by Judd and Esguerra [2010]. The paper estimates a proximity indicator for Mindanao towns, which is a function of the Gross Domestic Products (GDPs) of the trading towns and the distance between them. This is similar in functional form to the model used by Battersby [2006] and Yoshida and Deichmann [2009], which was originally intended to estimate the flow of labor or labor migration.

The proximity indicator formula employed by Judd and Esguerra [2010] has no weight or parameter estimation. For example, it merely assigned a weight of 1 for the GDPs of trading towns. In addition, the distance is assumed to have an exponential decay of 5, equivalent to the number of the identified growth centers in Mindanao. Using that model and assuming, for instance ₱1,320 million as an average town's GDP, the proximity indicator would decay to zero if that town is more than four kilometers from a Mindanao growth hub.

Another method to estimate the economic potential is the gravity model of trade (GMT) which was originated by Tinbergen in 1962, as cited by several authors (e.g., Linders and de Groot [2006]).[2] Recently, it was used to simulate the level of development around the world from Africa (Buys et al [2006]; Coulibaly and Fotagné [2004]) and from Latin America [Carillo and Li 2002].

In particular, GMT uses the variables found in the proximity indicator plus other variables like infrastructure level. In addition, the weights are not assumed but estimated from the data set. And as the name implies, the gravity model of trade assumes that trade is governed similarly to Newton's universal law of gravitation. With strong theoretical foundation, flexibility to accommodate other variables, and no ad hoc assumption on weights, the GMT is adapted in this paper to estimate economic potential. The succeeding sections are arranged as follows: details of the model, result and application, and summary.

2. Details of the gravity model of trade

Trade is a good indicator of potential for growth among towns, but unfortunately, trade estimates are only available at the regional level. Fortunately, the GMT can fill this gap via estimates of the weights at the regional level; the resulting formula can then be applied at the municipal level to generate municipal

[2] The Gravity Model Trade is also employed in urban planning, especially in estimating the volume of traffic, according to Prof. R. H. Racelis.

trade estimates. This is possible because the predictor variables are both found at the regional and at the municipal levels.

As mentioned before, the GMT bears a strong similarity to Newton's formula of gravitation. In this model, the two trading areas could be viewed as planets and the value of trade could be viewed as the gravitational force. The trade (~gravitational force) is dependent on the GDPs (~mass) of the two trading areas and their physical distance. The bigger the GDP (~mass) between the two trading areas (~planets), the greater is the trade (~gravitational force). The trade between the two areas decays exponentially as distance increases. The similarities end there as GMT can take other variables into consideration, like infrastructure and state of electrification, which can be viewed as sources of friction or impedance. Formally, the initial GMT is shown in equation 1. The final GMT model could have fewer predictor variables depending on the statistical significance of their weights. Predictor variables that are not statistically significant from zero will not be included in the computation of town-level trade estimates.

$$T_{ij} = K \frac{M_j^a X_i^b R_j^c R_i^d E_j^e E_i^f U_j^g U_i^h W_j^k W_i^l}{D_{ij}^m} \tag{1}$$

T_{ij} = trade value, imported by municipality/town/city j from i (in million ₱)
K = constant (~ gravitational constant, a parameter/coefficient)
M = GDP of municipality j (importer) (in million ₱)
X = GDP of municipality i (exporter) (in million ₱)
R = Road quality
E = % Household with electricity
U = Unemployment rate
W = % Household with access to water
D = Distance in km
$a, b, c, d, e, f, g, h, k, l, m$ = parameters/weights
j = subscript for the importing municipality
i = subscript for the exporting municipality

In equation 1 above, the exponential decay of distance D is m which could vary depending on the data, while in Newton's formula this is assumed to be 2. The parameter m and other parameters could be interpreted as elasticities of trade with respect to the predictor variables. For example, the parameter m is the distance elasticity which measures the percentage change in trade value caused by a percent change in distance.

There are many regression methods that may be deployed for estimating the parameters and the constant for the GMT model. The simplest method is the Ordinary Least Squares, which can be applied after taking the natural logarithm of equation 1 [Buys et al. 2006]. In most trade data sets, however, some trading areas have zero trades which could become undefined if log transformation is applied. This problem could be remedied if zero values are replaced by 0.1 to make log transformation possible. However, Santos Silva and Tenreyro [2006] found that replacing zeroes with 0.1 makes the Ordinary Least Squares estimates of parameters biased. The authors investigated several regression methods and conducted simulations and found that a method called Poisson regression is less prone to bias. This type of regression is based on the Poisson distribution which is usually deployed on count data [Zeileis et al. 2010]. Its mean is identical to its variance; thus the dispersion is fixed at 1.

There are two estimators of Poisson Regression: Poisson pseudo maximum likelihood (PPML) and Quasi-Poisson. According to Root [2011], the later is the most appropriate, if the zero trade values are frequent (say, more than half of the trading areas), although both estimators yield the same parameter estimates. Quasi-Poisson gives a larger standard error to correct the bias emanating from frequent zero trade values which makes it likely to reject predictor variables that are significant under PPML.

Another set of authors, Linders and de Groot [2006] and Gómez Herrera [2010] suggested using Heckman sample selection. Gómez Herrera [2010] went further by trying all the estimation methodologies and found that the Heckman sample selection method is the best, although Linders and de Groot [2006] conceded that it is difficult to find an identification restriction, a pre-requisite in the Heckman sample selection.

From the survey of regression methods for the GMT, we find that the most appropriate is the Poisson regression. Between the two Poisson regression estimators, PPML is the most appropriate based on the regional data set to be used given that zero trades are less than half of all trades. In Poisson regression, the zero values are handled by the model without the need for transformation. The functional form is shown below:

$$\begin{aligned} T_{ij} = \exp\,(&K + aM_j + bX_i + cR_j + dR_i + eE_j + fE_i \\ &+ gU_j + hU_i + kW_j + lW_i - mD_{ij}) \end{aligned} \tag{2}$$

3. Estimating the GMT parameters

The data set needed in estimating the GMT parameters is at regional level. Exploratory analysis was done, and three data points are found to be outliers: trade within the National Capital Region, trade within Region 6, and trade between Caraga to Region 7. These outliers were removed. The summary statistics of the data set (minus the outliers) are shown below.

TABLE 1. Summary statistics of regional data set

	Minimum	1st quantile	Median	Mean	3rd quantile	Maximum
Trade (₱ million)	0.0	218.3	2214.7	6300.4	6192.6	33158.2
Distance (km)	0.0	186.0	566.0	691.9	1176.0	1582.0
Regional GDP (₱ billion)	65.7	162.0	215.1	451.7	518.3	2813.8
% Paved road	47%	66%	78%	77%	88%	100%
Unemployment	3.00%	5%	6%	6%	8%	11%
% Households with water	8%	24%	33%	32%	37%	62%
% Households with electricity	56%	78%	82%	82%	87%	99%

The data set above is collated from different sources. For example, inter-regional trade is extracted from the National Statistics Office website, while the regional GDPs are obtained from the National Statistics and Coordination Board. The raw data on road quality is obtained from the Department of Public Works and Highways. Unemployment is taken from Family Incomes and Expenditures Survey 2009. The numbers of households with electricity and access to water are taken from the National Household Targeting System. The road distance between regions is measured along the road paths connecting regional centers and obtained using Google Earth.[3]

Using the regional data set, the GMT model is estimated using an open-source software called R 2.15.3, with parameters shown below.

TABLE 2. Variables and parameter estimates

	Estimate	Standard error	z value	Pr(>|z|)	Significance
Constant (~gravitational constant)	4.9590 (K)	0.0138	358.5300	0.0000	***
Exporter area GDP [X]	0.0006 (a)	0.0000	316.0000	0.0000	***
Importer area GDP [M]	0.0005 (b)	0.0000	178.9200	0.0000	***
Distance (km) [D]	-0.0003 (m)	0.0000	-92.4800	0.0000	***
% share of paved road [R] (importer)	0.6493 (c)	0.0182	35.7800	0.0000	***
Households with electricity [E] (exporter)	0.0000 (f)	0.0000	143.8200	0.0000	***
Unemployment [U] (exporter)	6.0830 (h)	0.1298	46.8700	0.0000	***

[3] Extraction was automated using a script.

The variable that is not significant, like the percentage of households with access to water, is removed in the final estimation of the model. All the signs of parameters are positive, except for distance "*m*," which is negative to indicate that trade between the municipality and growth center decreases with distance.

4. Estimating economic potential

With the parameters of the GMT estimated, the next step is to estimate the economic potential. This is done by plugging the municipal level variables that are identified as statistically significant as shown in Table 2. The final form of the GMT model is in exponential form specified below.

$$T_{ij} = \exp^{K + aX + bM - mD + cR_j + fE_i + hU_i} \tag{3}$$

In the regional data set, a region may trade with all other regions. This could be mirrored at the municipal level. However, this may lead to complexity as extraction of distance among 1,600 municipalities could be very tedious. To manage the computational complexity, a particular municipality is assumed to trade only with the three closest regional centers or other growth centers.[4] Mathematically, this could be summarized as

$$\sum_{j=1}^{3} T_{ij} \tag{4}$$

Both the import and the export values of municipal *j* from *i* are computed using equation 4. The average of the import and export values is the economic potential indicator.

Some of the datasets at the municipal level, such as the proportion of household with electricity and access to water, to be plugged into equation 3 are readily available as primary data. However, other datasets like municipal GDP and unemployment are not available; hence, they must be estimated from secondary data sources. To address this constraint, the Structure Preserving Estimation (SPREE) technique pioneered by Purcell and Kish [1979] is used. This technique is also used by Judd and Esguerra [2010] in estimating proximity indicator.

The SPREE technique uses ratio and proportions to estimate the missing variables at the lower level of geographic disaggregation. For example, to get the municipal GDP, the regional GDP is first divided into provincial outputs using

[4] Other growth centers are 60 towns or cities that appear both in the NEDA list of 101 highly urbanized areas based on 125,000 population cut-off and in the National Anti-Poverty Commission (NAPC) list of 99 high growth towns based on estimated municipal GDPs.

the number of establishments as weights. Finally, the provincial output is divided into municipal outputs based on the share of municipal population. To get the unemployment at the municipal level, the share of rural and urban unemployment at the provincial level are multiplied to urban and rural population to get the count. The estimated count of the unemployed at the municipal level is then expressed as share of unemployment. The description of the data sets and their origin are summarized in the table below.

TABLE 3. Municipal level variables and their sources

Variables	Level	Sources	Indirect	Method	Allocation variables
GDP	Municipal	NSCB[a], NSO[b]	Yes	SPREE[e]	Number of establishments, population
Distance	Municipal	Google Maps	No	NA	
% Share of paved road	Municipal	DPWH[c]	No	NA	
Households with electricity	Municipal	NHTS[d]	No	NA	
Unemployment	Muncipal	NSCB, NSO	Yes	SPREE	Rural and urban population

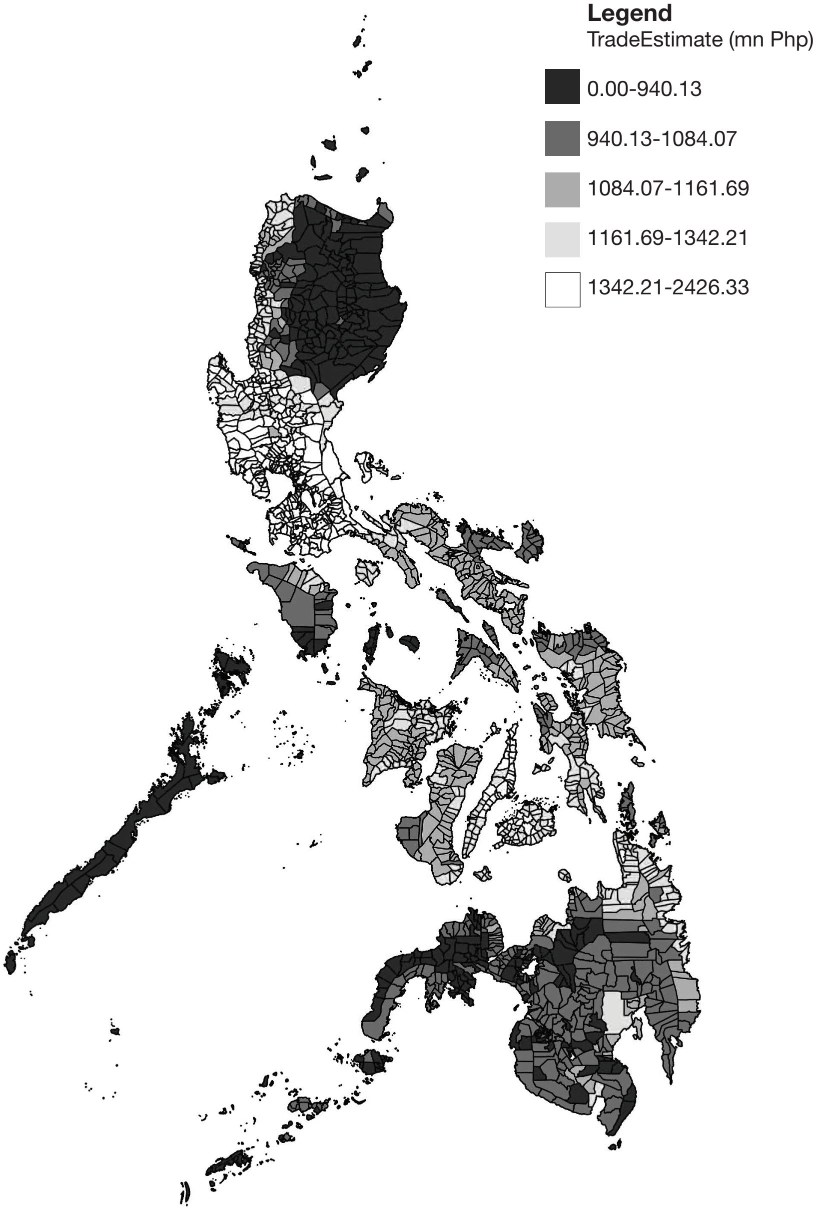
a NSCB: National Statistics Coordination Board
b NSO: National Statistics Office
c DPWH: Department of Public Works and Highways
d NHTS: National Household Targeting System
e SPREE: Structure Preserving Estimation, in which the allocation variables are used to distribute the values at the lower levels of disaggregation. For example, regional GDP is distributed to province based on the number of establishment, and to municipalities based on population. For detailed information on SPREE, see Purcell and Kish [1979].

On road quality, the lowest disaggregation is at the engineering district level. With this constraint, it is assumed in this exercise that municipal level road quality is uniform within each engineering district. On the other hand, the road distance from a municipality to the growth center is taken from Google Earth, but in cases where a road is not existent, a straight line is used. Given that there is no road, the default road quality assigned for that particular municipality is the lowest in the country.

5. Results and application

The result of economic potential is mapped using a heat map color scheme as shown in Figure 1 with QGIS software. The high economic potential areas are shown in green, while low economic potential areas are shown in red. The result confirms the common knowledge that the National Capital Region (together with its neighboring provinces) and Cebu are areas that have high economic potential. The map also shows the areas with low trades which are located mostly at the peripheries, like Cagayan Valley, and central to western portion of Mindanao.

FIGURE 1. Municipal level estimated average trade or economic potential for growth



The estimates of economic potential generated in this research have been used in the two policy applications, and in both instances in conjunction with other variables. For example, in the Bottom Up Budget planning for 2014, in which the goal is to shift the growth to poor areas, economic potential is used in conjunction with poverty count. This was done by selecting the top 650 towns with high economic potential (or high trade) on one side, and another top 650 poor towns with high poor count on the other side. Established urban centers were not considered (see footnote 7). The towns and cities that appear in both interim listings are considered poor towns or cities with high economic potential. To ensure actual economic potential, the resulting list was further trimmed down to town/city with in-migration above 1,000 (NSO 2000) as in-migration below 1,000 indicates minimal economic activity.

In another policy application, in which the goal of the Board of Investments is to grant tax incentives to new businesses that operate in less developed areas, the economic potential estimates were used in conjunction with Municipal Income Class classification.[5] Similar to the method above, two interim lists were created: a list of less developed areas as proxied by low economic potential (i.e., less developed areas); and a list of low income class towns (4th to 6th class). The towns that appear in both these lists were the ones granted with tax incentives.

6. Summary

This research is able to fill the gap of the policy question as to which among the towns/municipalities have the potential for economic growth. It uses the gravity model of trade, which is estimated using Poisson regression and calibrated using regional data sets. The economic potential estimates were computed using municipal level data sets that are either collected from the ground or generated using secondary sources of data. Aside from filling the gap, the more important thing this research has achieved is to aid and to clarify the selection of the municipalities based on differing policy objectives, but in conjunction with other variables, like poverty count, municipal income class, and in-migration.

**National Anti-Poverty Commission*

The authors are grateful to the participants of the 12th National Convention on Statistics and the 51st Philippine Economic Society Annual Meeting both held in Metro Manila for their helpful comments and suggestions. The usual disclaimer applies.

[5] Income Class is an indicator of revenue in a town; towns with lower income class have lower revenues and a smaller tax base. A business that will locate in lower income class town will help broaden the town's tax base.

References

Battersby, B. [2006] "Does distance matter? The effect of geographic isolation on productivity levels", *OECD Economic Studies* **2006**(1): 205-225.

Buys, P., U. Deichmann, and D. Wheeler [2006] "Road network upgrading and overland trade expansion in Sub-Saharan Africa", Development Research Group, Washington, DC: World Bank.

Carrillo, C. and C. Li [2002] "Trade blocks and the gravity model: evidence from Latin American countries", *Discussion Paper*, UK: University of Essex.

Coulibaly, S. and L. Fontagné, [2004] "South-South trade: geography matters", CEPII *Working Paper No. 2004-08*, Paris: CEPII Research Center.

Herrera, E.G. [2010] "Comparing alternative methods to estimate gravity models of bilateral trade", *Working Paper*. Department of Economic Theory, Spain: University of Granada.

Judd, M. and J. Esguerra [2010]. "Behind the veil of conflict: moving towards economic integration for sustained development and peace in Mindanao Philippines", Washington, DC: World Bank.

Linders, G.M. and H. L.F. de Groot [2006] "Estimation of the gravity equation in the presence of zero flows", *Discussion Paper No 2006-072/3*, The Netherlands: Tinbergen Institute.

Purcell, N.J. and L. Kish [1979]. "Estimation of small domains", *Biometrics* **35**: 365-384.

QGIS Development Team [2012] "QGIS Geographic Information System", Open Source Geospatial Foundation Project. http://qgis.osgeo.org. Accessed 28 April 2012.

Redding, S. and A.J. Venables [2002] "The economics of isolation and distance", *Nordic Journal of Political Economy* **28**: 93-108.

Root, E.D. [2011] "Poisson regression using count data", http://www.colorado.edu/geography/class_homepages/geog_4023_s11/Lecture07b_PoissReg.pdf. Accessed 27 April 2012.

R Core Team [2013] "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/. Accessed 28 April 2012.

Santos Silva, J.M.C. and S. Tenreyro [2006]. "The log of gravity", *The Review of Economics and Statistics* **88**: 641-658.

Yoshida, N., and U. Deichmann [2009] "Measurement of accessibility and its applications", *Journal of Infrastructure Development* **1**(1):1-18.

Zeileis, A., C. Kleiber, and S. Jackman [2010] "Regression models for count data in R", http://www. cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf. Accessed 27 April 2012.