NBER WORKING PAPER SERIES

THE GRAVITY MODEL

James E. Anderson

Working Paper 16576
http://www.nber.org/papers/w16576

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2010

I thank Jeffrey H. Bergstrand, Keith Head, J. Peter Neary and Yoto V. Yotov for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The Gravity Model
James E. Anderson
NBER Working Paper No. 16576
December 2010
JEL No. F10,R1

ABSTRACT

The gravity model in economics was until relatively recently an intellectual orphan, unconnected to the rich family of economic theory. This review is a tale of the orphan's reunion with its heritage and the benefits that have flowed from it. Gravity has long been one of the most successful empirical models in economics. Incorporating the theoretical foundations of gravity into recent practice has led to a richer and more accurate estimation and interpretation of the spatial relations described by gravity. Recent developments are reviewed here and suggestions are made for promising future research.

James E. Anderson
Department of Economics
Boston College
Chestnut Hill, MA 02467
and NBER
james.anderson.1@bc.edu

The gravity model in economics was until relatively recently an intellectual orphan, unconnected to the rich family of economic theory. This review is a tale of the orphan's reunion with its heritage and the benefits that continue to flow from connections to more distant relatives.

Gravity has long been one of the most successful empirical models in economics, ordering remarkably well the enormous observed variation in economic interaction across space in both trade and factor movements. The good fit and relatively tight clustering of coefficient estimates in the vast empirical literature suggested that some underlying economic law must be at work, but in the absence of an accepted connection to economic theory, most economists ignored gravity. The authoritative survey of Leamer and Levinsohn (1995) captures the mid-90's state of professional thinking: "These estimates of gravity have been both singularly successful and singularly unsuccessful. They have produced some of the clearest and most robust empirical findings in economics. But, paradoxically, they have had virtually no effect on the subject of international economics. Textbooks continue to be written and courses designed without any explicit references to distance, but with the very strange implicit assumption that countries are both infinitely far apart and infinitely close, the former referring to factors and the latter to commodities." Subsequently, gravity first appeared in textbooks in 2004 (Feenstra, 2004), following on success in connecting gravity to economic theory, the subject of Section 3.

Reviews are not intended to be surveys. My take on the gravity model, thus licensed to be idiosyncratic, scants or omits some topics that others have found important while it emphasizes some topics that others have scanted. My emphases and omissions are intended to guide the orphan to maturity. An adoptive parent's biases may have contaminated my judgment, *caveat emptor*.

My focus is on theory. Incorporating the theoretical foundations of gravity into recent practice has led to richer and more accurate estimation and interpretation of the spatial relations described by gravity, so where appropriate I will point out this benefit. The har-

vest reaped from empirical work applying the gravity model is recently surveyed elsewhere (Anderson and van Wincoop, 2004; Bergstrand and Egger, 2009).

From a modeling standpoint, gravity is distinguished by its parsimonious and tractable representation of economic interaction in a many country world. Most international economic theory is concentrated on two country cases, occasionally extended to three country cases with special features. The tractability of gravity in the many country case is due to its modularity: the distribution of goods or factors across space is determined by gravity forces conditional on the size of economic activities at each location. Modularity readily allows for disaggregation by goods or regions at any scale and permits inference about trade costs not dependent on any particular model of production and market structure in full general equilibrium. The modularity theme recurs often below, but is missing from some other prominent treatments of gravity in the literature.

1 Traditional Gravity

The story begins with setting out the traditional gravity model and noting clues to uniting it with economic theory. The traditional gravity model drew on analogy with Newton's Law of Gravitation. A mass of goods or labor or other factors of production supplied at origin i, Y_i, is attracted to a mass of demand for goods or labor at destination j, E_j, but the potential flow is reduced by distance between them, d_{ij}. Strictly applying the analogy,

$$X_{ij} = Y_i E_j / d_{ij}^2$$

gives the predicted movement of goods or labor between i and j, X_{ij}. Ravenstein pioneered the use of gravity for migration patterns in the 19th century UK (Ravenstein, 1889). Tinbergen (1962) was the first to use gravity to explain trade flows. Departing from strict analogy, traditional gravity allowed the coefficients of 1 applied to the mass variables and of 2 applied to bilateral distance to be generated by data to fit a statistically inferred relationship

between data on flows and the mass variables and distance. Generally, across many applications, the estimated coefficients on the mass variables cluster close to 1 and the distance coefficients cluster close to -1 while the estimated equation fits the data very well: most data points cluster close to the fitted line in the sense that $80-90\%$ of the variation in the flows is captured by the fitted relationship. The fit of traditional gravity improved when supplemented with other proxies for trade frictions, such as the effect of political borders, common language and the like.

Notice that bilateral frictions alone would appear to be inadequate to fully explain the effects of trade frictions on bilateral trade, because the sale from i to j is influenced by the resistance to movement on i's other alternative destinations and by the resistance on movement to j from j's alternative sources of supply. Prodded by this intuition the traditional gravity literature recently developed remoteness indexes of each country's 'average' effective distance to or from its partners ($\sum_i d_{ij}/Y_i$ was commonly defined as the remoteness of country j) and used them as further explanatory variables in the traditional gravity model, with some statistical success.

The general problem posed by the intuition behind remoteness indexes is analogous to the N-body problem in Newtonian gravitation. An economic theory of gravity is required for an adequate solution. Because there are many origins and many destinations in any application, a theory of the bilateral flows must account for the relative attractiveness of origin-destination pairs. Each sale has multiple possible destinations and each purchase has multiple possible origins: any bilateral sale interacts with all others and involves all other bilateral frictions. This general equilibrium problem is neatly solved with structural gravity models.

For expositional ease, the discussion will focus on goods movements from now on except when migration or investment are specifically treated.

2 Frictionless Gravity Lessons

Taking a step toward structure, an intuitively appealing starting point is the description of a completely smooth homogeneous world in which all frictions disappear. Developing the implications of this structure yields a number of useful insights about the pattern of world trade.

A frictionless world implies that each good has the same price everywhere. In a homogeneous world, economic agents everywhere might be predicted to purchase goods in the same proportions when faced with the same prices. In the next section the assumptions on preferences and/or technology that justify this plausible prediction will be the focus, but here the focus is on what the implications are for trade patterns. In a completely frictionless and homogeneous world, the natural benchmark prediction is that $X_{ij}/E_j = Y_i/Y$, the proportion of spending by j on goods from i is equal to the global proportion of spending on goods from i, where Y denotes world spending.

Any theory must impose adding up constraints, which for goods requires that the sum of sales to all destinations must equal Y_i , the total sales by origin i, and the sum of purchases from all origins must equal E_j, the total expenditure for each destination j. Total sales and expenditures must be equal: i.e., $\sum_i Y_i = \sum_j E_j = Y$.

One immediate payoff is an implication for inferring trade frictions. Multiplying both sides of the frictionless benchmark prediction $X_{ij}/E_j = Y_i/Y$ by E_j yields predicted frictionless trade $Y_i E_j/Y$. The ratio of observed trade X_{ij} to predicted frictionless trade $Y_i E_j/Y$ represents the effect of frictions along with random influences. (Bilateral trade data is notoriously rife with measurement error.) Fitting the statistical relationship between the ratio of observed to frictionless trade and various proxies for trade costs is justified by this simple theoretical structure as a proper focus of empirical gravity models.

Thus far, the treatment of trade flows has been of a generic good which most of the literature has implemented as an aggregate: the value of aggregate bilateral trade in goods for example. But the model applies more naturally to disaggregated goods (and factors) because

the frictions to be analyzed below are likely to differ markedly by product characteristics. The extension to disaggregated goods, indexed by k, is straightforward.

$$X_{ij}^k = \frac{Y_i^k E_j^k}{Y^k} = s_i^k b_j^k Y^k. \tag{1}$$

Here $s_i^k = Y_i^k/Y^k$ is country i's share of the world's sales of goods class k and $b_i^k = E_j^k/Y^k$ is country j's share of the world spending on k, equal to world sales of k, Y^k.

The notation and logic also readily apply to disaggregation of countries into regions, and indeed a prominent portion of the empirical literature has examined bilateral flows between city pairs or regions, motivated by the observation that much economic interaction is concentrated at very short distances. The model can interpreted to reflect individual decisions aggregated with a probability model; see section 5.1 below.

In aggregate gravity applications (i.e., most applications), it has been common to use origin and destination mass variables equal to Gross Domestic Product (GDP). This is conceptually inappropriate and leads to inaccurate modeling unless the ratio of gross shipments to GDP is constant (in which case the ratio goes into a constant term). A possible direction for aggregate modeling is to convert trade to the same value added basis as GDP, but this seems more problematic than using disaggregated gravity to explain the pattern of gross shipments and then uniting estimated gravity models within a superstructure to connect to GDP. That is the strategy of the structural gravity model research program reviewed here.

Equation (1) generates a number of useful implications.

1. Big producers have big market shares everywhere,

2. small sellers are more open in the sense of trading more with the rest of the world,

3. the world is more open the more similar in size are countries and the more specialized are countries,

4. the world is more open the greater the number of countries, and

5. world openness rises with convergence under the simplifying assumption of balanced trade.

As for implication 1 big producers have big market shares everywhere, this follows because, reverting to the generic notation and omitting the k superscript, the frictionless gravity prediction is that :

$$X_{ij}/E_j = s_i.$$

Implication 2, small sellers are more open in the sense of trading more with the rest of the world follows from

$$\sum_{i \neq j} X_{ij}/E_j = 1 - Y_j/Y = 1 - s_j$$

using $\sum_j E_j = \sum_i Y_i$, balanced trade for the world.

Implication 3 is that the world is more open the more similar in size are countries and the more specialized are countries. It is convenient to define world openness as the ratio of international shipments to total shipments, $\sum_j \sum_{i \neq j} X_{ij}/Y$. Dividing (1) through by Y^k and suppressing the goods index k, world openness is given by

$$\sum_j \sum_{i \neq j} X_{ij}/Y = \sum_j b_j(1 - s_j) = 1 - \sum_j b_j s_j.$$

Using standard statistical properties

$$\sum_j b_j s_j = N r_{bs} \sqrt{Var(s) Var(b)} + 1/N,$$

where N is the number of countries or regions, Var denotes variance, r_{bs} is the correlation coefficient between b and s and $1/N = \sum_i s_i/N = \sum_j b_j/N$, the average share.This equation is derived using standard properties of covariance and the adding up condition on shares. Here, $Var(s), Var(b)$ measures size dis-similarity while the correlation of s and b, r_{bs}, is an

inverse measure of specialization. Substituting into the expression for world openness:

$$\sum_j \sum_{i \neq j} X_{ij}/Y = 1 - 1/N - N r_{bs} \sqrt{Var(s)Var(b)} \tag{2}$$

Implication 3 follows from equation (2) because similarity of country size shrinks the variances on the right hand side while specialization shrinks the correlation r_{bs}.

The country size similarity property has been prominently stressed in the monopolistic competition and trade literature. (It is sometimes taken as evidence for monopolistic competition in a sector rather than as a consequence of gravity no matter what explains the pattern of the b's and s's.) The specialization property has also been noted in that literature as reflecting forces that make for greater net international trade, the absolute value of $s_j - b_j$. Making comparisons across goods classes, variation in the right hand side of (2) is due to variation in specialization and in the dispersion of the shipment and expenditure shares. Notice again that the cross-commodity variation in world openness arises here in a frictionless world, a reminder that measures of world home bias in a world with frictions must be evaluated relative to the frictionless world benchmark.

Country size similarity also tends to increase bilateral trade between any pair of countries, all else equal. This point (Bergstrand and Egger, 2007) is seen most clearly with aggregate trade that is also balanced, hence $s_j = b_j$. Equation (1) can be rewritten as

$$X_{ij} = s_i^{ij} s_j^{ij} \frac{(Y_i + Y_j)^2}{Y},$$

where $s_i^{ij} \equiv Y_i/(Y_i + Y_j)$, the share of i in the joint GDP of i and j. The product $s_i^{ij} s_j^{ij}$ is maximized at $s_i^{ij} = s_j^{ij} = 1/2$, so for given joint GDP size, bilateral trade is increasing in country similarity. (With unbalanced trade or specialization, an analogous similarity property holds for the bilateral similarity of income and expenditure shares. Let $\gamma_j = E_j/Y_j$. Then the same equation as before holds with the right hand side multiplied by γ_j.)

A more novel implication of equation (2) is that, implication 4, world openness is ordi-

narily increasing in the number of countries. Increasing world openness due to a rise in the number of countries reflects the property that smaller countries are more naturally open and division makes for more and smaller countries.

This effect is seen by differentiating the left hand side of $\sum_j \sum_{i \neq j} X_{ij}/Y = 1 - \sum_j b_j s_j$, yielding $-\sum_j (b_j ds_j + s_j db_j)$. Increasing the number of countries tends to imply reducing the share of each existing country while increasing the share (from zero) of the new country. The preceding differential expression should thus ordinarily be positive.

The qualification 'ordinarily' is needed because the pattern of share changes will depend on the underlying structure as revealed by the left hand side of equation (2). On the one hand, the average share $1/N$ decreases as N rises, raising world openness. On the other hand, the change in the number of countries will usually change $r_{bs}\sqrt{Var(b)Var(s)}$ in ways that depend on the type of country division (or confederation) as well as indirect effects on shares as prices change. (The apparent direct effect of N in the first on the right hand side of (2) vanishes because $1/N$ scales $\sqrt{Var(b)Var(s)}$.)

A practical implication of this discussion is that inter-temporal comparisons of ratios of world international trade to world income, to be economically meaningful, should be controlled for changes in the size distribution and the number of countries, a correction of large practical importance in the last 50 to 100 years. Alternatively, measures of openness meant to reflect the effects of trade frictions should be constructed in relation to the frictionless benchmark.

Applied to aggregate trade data, gravity yields implication 5, world openness rises with convergence under the simplifying assumption of balanced trade for each country, $b_j = s_j, \forall j$. The right hand side of equation (2) becomes $NVar(s) + 1/N$ under balanced trade, and per capita income convergence lowers $Var(s)$ toward the variance of population. Baier and Bergstrand (2001) use the convergence property to partially explain postwar growth in world trade/income, finding relatively little action, though presumably more recent data influenced by the rise of China and India might give more action.

Pointing toward a connection with economic theory, the plausible hypothesis of the frictionless model and the shares s_i and b_j must originate from an underlying structure of preferences and technology. Also, the deviation of observed X_{ij} from the frictionless prediction reflects frictions as they act on the pattern of purchase decisions of buyers and the sales decisions of sellers, which originate from an underlying structure of preferences and technology.

3 Structural Gravity

Modeling economies with trade costs works best if it moves backward from the end user. Start by evaluating all goods at user prices, applying demand side structure to determine the allocation of demand at those prices. Treat all costs incurred between production and end use as being incurred by the supply side of the market, even though there are often significant costs directly paid by the user. What matters economically in the end is the full cost between production and end use, and the incidence of that cost on producer and end user. Many of these costs are not directly observable, and the empirical gravity literature indicates the total is well in excess of the transportation and insurance costs that are observable (see Anderson and van Wincoop, 2004, for a survey of trade costs).

The supply side of the market under this approach both produces and distributes the delivered goods, incurring resource costs that are paid by end users. The factor markets for those resources must clear at equilibrium factor prices, determining costs that link to end user prices. Budget constraints require national factor incomes to pay for national expenditures plus net lending or transfers including remittances. Below the national accounts, individual economic agents also meet budget constraints. Goods markets clear when prices are found such that demand is equal to supply for each good. The full general equilibrium requires a set of bilateral factor prices and bilateral goods prices such that all markets clear and all budget constraints are met.

This standard description of general economic equilibrium is too complex to yield something like gravity. A hugely useful simplification is modularity, subordinating the economic determination of equilibrium distribution of goods within a class under the superstructure determination of distribution of production and expenditure between classes of goods. Anderson and van Wincoop (2004) call this property trade separability. Observing that goods are typically supplied from multiple locations, even within fine census commodity classes, it is natural to look for a theoretical structure that justifies grouping in this way. The structural gravity model literature has uncovered two structures that work, one on the demand side and one on the supply side, detailed in sections 3.1 and 3.2.

Modularity (trade separability) permits the analyst to focus exclusively on inference about distribution costs from the pattern of distribution of goods (or factors) without having to explain at the same time what determines the total supplies of goods to all destinations or the total demand for goods from all origins. This is a great advantage for two reasons. First, it simplifies the inference task enormously. Second, the inferences about the distribution of goods or factors is consistent with a great many plausible general equilibrium models of national (or regional) production and consumption.

Modularity also requires a restriction on trade costs, so that only the national aggregate burden of trade costs within a goods class matters for allocation between classes. The most popular way to meet this requirement is to restrict the trade costs so that the distribution of goods uses resources in the same proportion as the production of those same goods. Samuelson (1952) invented iceberg melting trade costs in which the trade costs were proportional to the volume shipped, as the amount melted from the iceberg is proportional to its volume. The iceberg metaphor still applies when allowing for a fixed cost, as if a chunk of the iceberg breaks off as it parts from the mother glacier. Mathematically, the generalized iceberg trade cost is linear in the volume shipped. Economically, distribution continues to require resources to be used in the same proportion as in production. Fixed costs are realistic and potentially play an important role in explaining why many potential bilateral flows are equal

to zero.

More general nonlinear trade cost functions continue to satisfy the production proportionality restriction and thus meet the requirements of modularity, but depart from the iceberg metaphor. Bergstrand (1985) derived a joint cost function that is homogeneous of degree one with Constant Elasticity of Transformation (CET). This setup allows for substitution effects in costs between destinations rather than the cost independence due to fixed coefficients in the iceberg model. Bilateral costs have a natural aggregator that is an iceberg cost facing monopolistically competitive firms. A nice feature of the joint cost model is its econometric tractability under the hypothesis of profit maximizing choice of destinations. While potentially more realistic, the joint cost refinement turns out to make relatively little difference empirically.

Arkolakis (2008) develops a nonlinear (in volume) trade cost function in which heterogeneous customers are obtained by firms with a marketing technology featuring a fixed cost component (running a national advertisement) and a variable cost component (leafletting or telemarketing) subject to diminishing returns as the less likely customers are encountered. Because of the Ricardian production and distribution technology, resource requirements in distribution remain proportional to production resource requirements. Arkolakis shows that the marketing technology model can rationalize features of the firm level bilateral shipments data that cannot be explained with the linear fixed costs model. His setup is not econometrically tractable but is readily applicable as a simulation model.

In all applications based on the preceding cost functions, proxies for costs are entered in some convenient functional form, usually loglinear in variables such as bilateral distance, contiguity, membership of a country, continent or regional trade agreement, common language, common legal traditions and the like. See Anderson and van Wincoop (2004) for more discussion.

More generality in trade costs that violates the production proportionality restriction comes at the price of losing modularity. See Matsuyama (2007) for recent exploration of the

implications of non-iceberg trade costs in a 2 country Ricardian model. See Deardorff (1980) for a very general treatment of the resource requirements of trade costs as a setting for his demonstration that the law of comparative advantage holds quite generally.

3.1 Demand Side Structure

The second requirement for modularity can be met by restricting the preferences and/or technology such that the cross effects in demand between classes of goods (either intermediate or final) flow only through aggregate price indexes. This demand property is satisfied when preferences or technology are homothetic and weakly separable with respect to a partition into classes whose members are defined by location, a partition structure called the Armington assumption. Thus for example steel products from all countries are members of the steel class. Notice that the assumption implies that goods are purchased from multiple sources because they are evaluated differently by end users, goods are differentiated by place of origin.

It is usual to impose identical preferences across countries. Differences in demand across countries, such as a home bias in favor of locally produced goods, can be accommodated, understanding that 'trade costs' now include the effect of a demand side home bias. In practice it is very difficult to distinguish demand side home bias from the effect of trade costs, since the proxies used in the literature (common language, former colonial ties, or internal trade dummies, etc.) plausibly pick up both demand and cost differences. Henceforth trade cost is used without quotation marks but is understood to potentially reflect demand side home bias. Declines in trade costs can be understood as reflecting homogenization of tastes.

Separability implies that each goods class has a natural quantity aggregate and a natural price aggregate, with substitution between goods classes occurring as if the quantity aggregates were goods in the standard treatment. The separability assumption implies that national origin expenditure shares within the steel class are not altered by changes in the prices of non-steel products, though of course the aggregate purchase of steel are effected by

the aggregate cross effect. Homotheticity ensures that relative demands are functions only of relative aggregate prices.

The first economic foundation for the gravity model was based on specifying the expenditure function to be a Constant Elasticity of Substitution (CES) function (Anderson, 1979). Expenditure shares in the CES case are given by

$$\frac{X_{ij}}{E_j} = \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{1-\sigma} \tag{3}$$

where P_j is the CES price index, σ is the elasticity of substitution parameter, β_i is the 'distribution parameter' for varieties shipped from i, p_i is their factory gate price and $t_{ij} > 1$ is the trade cost factor between origin i and destination j. The CES price index is given by

$$P_j = \left(\sum_i (\beta_i p_i t_{ij})^{1-\sigma}\right)^{1/(1-\sigma)}. \tag{4}$$

Notice that the same parameters characterize expenditure behavior in all locations; preferences are common across the world by assumption. Notice also that the shares are invariant to income, preferences are homothetic. With frictionless trade, $t_{ij} = 1, \forall(i,j)$ and therefore all the buyers' shares of good i must equal the sellers share of world sales (at destination prices), Y_i/Y. Thus the frictionless benchmark is justified by assuming identical homothetic preferences. For intermediate goods, the same logic works replacing expenditure shares with cost shares.

The 'distribution parameters' β_i bear several interpretations. They could be exogenous taste parameters. Alternatively, in applications to monopolistically competitive products, β_i is proportional to the number of firms from i offering distinct varieties (Bergstrand, 1989). Countries with more active firms get bigger weights. In long run monopolistic competition the number of firms is endogenous. Due to fixed entry costs, bigger countries have more active firms in equilibrium, all else equal. The number of active firms contributes to determining the Y_i's that are given in the gravity module.

The other building block in the structural gravity model is market clearance: at delivered prices $Y_i = \sum_j X_{ij}$. Multiplying both sides of (3) by E_j and summing over j yields a solution for $\beta_i p_i^{1-\sigma}$,

$$\beta_i p_i^{1-\sigma} = \frac{Y_i}{\sum_j (t_{ij}/P_j)^{1-\sigma} E_j}.$$

Define the denominator as $\Pi_i^{1-\sigma}$.

Substituting into (3) and (4) yields the structural gravity model:

$$X_{ij} = \frac{E_j Y_i}{Y} \left(\frac{t_{ij}}{P_j \Pi_i} \right)^{1-\sigma} \tag{5}$$

$$(\Pi_i)^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j} \right)^{1-\sigma} \frac{E_j}{Y} \tag{6}$$

$$(P_j)^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i}. \right)^{1-\sigma} \frac{Y_i}{Y}. \tag{7}$$

The second ratio on the right hand side of (5) is a decreasing function (under the empirically valid restriction $\sigma > 1$) of direct bilateral trade costs relative to the product of two indexes of all bilateral trade costs in the system.

Anderson and van Wincoop (2003) called the terms P_j and Π_i inward and outward multilateral resistance. Note that $\{P_j^{1-\sigma}, \Pi_i^{1-\sigma}\}$ can be solved from (6)-(7) for given $t_{ij}^{1-\sigma}$'s, E_j's and Y_i's combined with a normalization.[1] Under the assumption of bilateral trade cost symmetry $t_{ij} = t_{ji}, \forall i, j$ and balanced trade $E_j = Y_j, \forall j$, the natural normalization is $\Pi_i = P_i$. Anderson and van Wincoop estimated their gravity equation for Canada's provinces and US states with a full information estimator that utilized (7) with $\Pi_i = P_i$. Subsequent research has focused mostly on estimating (5) with directional country fixed effects to control for $E_j/P_j^{1-\sigma}$ and $Y_i/\Pi_i^{1-\sigma}$.

Multilateral resistance is on the face of it an index of inward and outward bilateral trade costs, but because of the simultaneity of the system (6)-(7), all bilateral trade costs in the

[1] For any solution to the system $\{P_j^0, \Pi_i^0\}$, $\{\lambda P_j^0, \Pi_i^0/\lambda\}$ is also a solution. Thus a normalization is needed. Anderson and Yotov find that the system (6)-(7) solves quite quickly, not surprisingly because it is quadratic in the $1-\sigma$ power transforms of the P's and Π's.

world contribute to the solution values. This somewhat mysterious structure has a simple and intuitive interpretation: inward and outward multilateral resistance measure average buyer's and sellers incidence of trade costs respectively.

The incidence interpretation follows because the uniform preferences assumption in demand implies that the seller in effect makes a single shipment at a uniform markup factor Π_i to a world market with a share determined by

$$\frac{Y_i}{Y} = \left(\frac{\beta_i p_i \Pi_i}{P_W} \right)^{1-\sigma}. \tag{8}$$

The right hand side of (8), referring to the general form (3), is interpreted as the global expenditure share on the good from i in a hypothetical unified world market, where the world price index $P_W = (\sum_i (\beta_i p_i \Pi_i)^{1-\sigma})^{1/(1-\sigma)}$ is solved from summing (8). $P_W = 1$ is a convenient normalization of this hypothetical world price. Then with given $\beta_i p_i$'s the normalization $(\sum_i (\beta_i p_i \Pi_i)^{1-\sigma})^{1/(1-\sigma)} = 1$ is a useful normalization in solving for multilateral resistances with (6)-(7). The factor Π_i is straightforwardly interpreted as the sellers' incidence of trade costs from origin i (Anderson and Yotov, 2009).

P_j is now interpreted as buyers' incidence. Solving (7) for P_j, it is a CES index of bilateral buyers' incidences $t_{ij}/\Pi_i, \forall i$, equivalent to buyers paying a uniform markup factor P_j on its entire bundle of shipments (from all i). Sellers incidence Π_i similarly is then interpretable as a CES index of the bilateral sellers' incidences t_{ij}/P_j, from (6).

The interpretation of Π and P as buyers' and sellers' incidence generalizes the elementary economics idea of incidence in the one good case. If the actual set of trade costs were to be replaced with hypothetical trade costs $\widetilde{t}_{ij} = \Pi_i P_j$, market clearance and budget constraints (6)-(7) would still hold with the initial equilibrium shares, hence the sellers' factory gate prices would remain the same and the aggregate buyers' prices would remain the same.[2] In this sense, the set of bilateral t_{ij} are equivalent to the set of $\widetilde{t}_{ij}$'s that decompose into the

[2]This property of (6)-(7) was noted by Anderson and van Wincoop (2004), foreshadowing the interpretation of multilateral resistance as incidence.

product of buyers' and sellers' incidence factors. (Unlike the one good case, it is the aggregate sales and purchases that are constant; bilateral flows would change in the hypothetical equilibrium.)

The model (5)-(7) is for a generic good. Anderson and van Wincoop (2004) argue theoretically for estimating disaggregated gravity while Anderson and Yotov (2009, 2010) demonstrate that aggregation bias is large in practice. For disaggregated gravity, all variables and parameters (5)-(7) should be understood as having superscript k's to denote the goods class in question. When accounting for substitution between goods classes, aggregate expenditure (or the cost of intermediate inputs) is given by the expenditure (or cost) function $C(P_j^1, ..., P_j^K)u_j$, where $C(\cdot)$ is the aggregate cost of living index for j and u_j is the utility of the representative agent (or quantity of aggregate output). Then, by Shephard's Lemma, $E_j^k = P_j^k \partial C(\cdot)/\partial P_j^k$. Each class of goods has expenditure shares described by (3)-(4) but amended to add superscript k to every variable and parameter.

The buyers' and sellers' incidence measures are usefully interpreted as the incidence of TFP frictions in distribution. They contrast with standard TFP-type measures of productivity in distribution. The sectoral TFP friction in distribution is defined by the uniform friction that preserves the value of sectoral shipments at destination prices: $\bar{t}_i^k = \sum_j t_{ij}^k y_{ij}^k / \sum_j y_{ij}^k$ where y_{ij}^k denotes the number of units of product class k received from i at destination j. $\bar{t}_i^k$ is a Laspeyres index of outward trade frictions facing seller i in good k.

The TFP measure $\bar{t}_i^k$ is useful for analyzing distribution productivity of the world economy as a whole, but it is misleading for purposes of understanding comparative economic performance and the national patterns of production and trade. $\bar{t}_i^k$ gives the sellers' incidence only under the partial equilibrium and inconsistent assumption that all incidence falls on the seller i. Anderson and Yotov (2009, 2010) show that in practice these differences are significant: Laspeyres TFP measures and the incidence of TFP in distribution differ in magnitude and in the case of inward measures the correlation between them is low.[3]

[3] An alternative measure proposed by Redding and Venables (2004) resembles multilateral resistance but does not measure incidence. Their measure of 'market access' uses essentially the same formula as (6) while

For consistency of the gravity modules with full general equilibrium, involving allocation across the sectors k in each country, the Π's are normalized in each sector k for given parameters and 'factory gate' price p_i^k by

$$\sum_j (\beta_i^k p_i^k \Pi_i^k)^{1-\sigma_k} = 1. \tag{9}$$

In practice, when analyzing a gravity module, it is often convenient to normalize one of the P's to one. The choice of normalization is irrelevant to distribution of the goods because only relative incidence matters.

Now return to the interpretation of the gravity equation (5), reproduced below for convenience.

$$X_{ij} = \frac{Y_i E_j}{Y} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma}.$$

The right hand side is the product of two ratios. The first ratio is the predicted frictionless trade flow given the E's and Y's, $Y_i E_j / Y$. The second ratio is thus interpreted as the ratio of predicted (given the t's) to predicted frictionless trade.

The useful measure of Constructed Home Bias (Anderson and Yotov, 2009) is interpreted as the predicted value of internal trade of i with itself to the predicted value of internal trade in the frictionless equilibrium. Constructed Home Bias is thus given by

$$CHB_i \equiv \left(\frac{t_{ii}}{\Pi_i P_i} \right)^{1-\sigma}. \tag{10}$$

CHB varies substantially by country, product and time due to changing expenditure and supply shares, even when gravity coefficients are constant (Anderson and Yotov; 2009, 2010).

Policy makers are often focused on overall import penetration ratios such as $\sum_{i \neq j} X_{ij}/E_j$ and the analogous ratio $\sum_{i \neq j} X_{ji}/Y_j$ for exports. These concerns are acute for certain goods

their measure of 'supplier access' uses the CES price index formula $P_j^k = [\sum_i (\beta_i^k p_i^k t_{ij}^k)^{1-\sigma_k}]^{1/(1-\sigma_k)}$. These variables are constructed without taking account of the simultaneous determination of the two variables, so they do not measure incidence.

classes. The import and export penetration ratios are a linear function of CHB for any goods class k:

$$\sum_{i \neq j} X_{ij}^k / E_j^k = 1 - (t_{jj}^k / P_j^k \Pi_j^k)^{1-\sigma_k} Y_j^k / Y^k. \tag{11}$$

$$\sum_{i \neq j} X_{ji}^k / Y_j^k = 1 - (t_{jj}^k / P_j^k \Pi_j^k)^{1-\sigma_k} E_j^k / Y^k. \tag{12}$$

Anderson and Yotov show that CHB's vary a lot across goods and more importantly for policy concerns, they exhibit a lot of intertemporal movement due to changing world shipment shares at constant t_{ij}^k's, implying a lot of explanatory power over the import and export ratios.

The interpretation of the second ratio in (5) applies straightforwardly to any bilateral flow: it is equal to the ratio of predicted bilateral trade to predicted frictionless trade, hence $(t_{ij}/\Pi_i P_j)^{1-\sigma}$ is the 'constructed trade bias' on the link from i to j due to the buyer's bilateral incidence from i relative to the average buyer's incidence for country j. Alternatively, the same statistic viewed from the exporter's viewpoint is due to the bilateral seller's incidence relative to the average seller's incidence. Bilateral trade flows shift about due to changes in production and expenditure shares of world shipments, as implied by the frictionless gravity model, but also due to the general equilibrium force of share changes that alters incidence even when trade costs $\{t_{ij}\}$ are constant (Anderson and Yotov, 2009).

The gravity model also readily disaggregates within countries, allowing useful investigations of inter-regional vs. international trade costs. Indeed, the development of the structural gravity model (Anderson and van Wincoop, 2003) was provoked to solve a puzzle posed by one of the most provocative and useful empirical findings of the traditional gravity literature. McCallum (1995) found that crossing the Canadian border had an enormous trade destroying effect on the trade flows of Canada's provinces. Canada's provinces were found by McCallum to trade 22 times more with each other than with US states, all else equal. This was too large to make sense as a component of bilateral trade costs t_{ij}.

Structural gravity solved the puzzle by showing that the border dummy variable in McCallum's traditional model reflected the effect of multilateral resistance. The border dummy in the McCallum regression shifts the ratio of inter-provincial trade to province-state trade. Because it is a traditional gravity regression it does not control for multilateral resistance. Using (5) to form this ratio for a pair of such flows in the structural gravity model and rearranging terms yields, for British Columbia's exports to adjacent Alberta and across the US border to adjacent Washington

$$\frac{X_{BC,AB}}{X_{BC,WA}} = \left(\frac{t_{BC,WA}}{t_{BC,AB}}\frac{P_{AB}}{P_{WA}}\right)^{\sigma-1}.$$

The expression on the right hand side of the equation reflects not only the direct trade cost increase at the US border that raises $t_{BC,WA}/t_{BC,AB}$, but the effect of the ratio of multilateral resistances for a province and a state, in this case Alberta and Washington, P_{AB}/P_{WA}. Since Canada's provinces must do far more of their trade with the outside world than do US states (Canada is about one tenth the size of the US in GDP), the provinces naturally have higher multilateral resistance than the states, thereby greatly increasing inter-provincial trade. In McCallum's traditional gravity regression the border dummy variable has a regression coefficient that is an average of such terms, though a biased estimate of it due to the omission of the multilateral resistance controls from his regression. Estimating the structural gravity model, Anderson and van Wincoop (2003) found a more plausible border cost component of t_{ij}, in the range of 20% to 50%.

Inter-regional vs. international trade cost implications of structural gravity were further developed by Anderson and Yotov (2009). They offer a decomposition of incidence into domestic and international components and calculate sellers' incidence for Canada's provinces on trade within Canada as compared to trade with the rest of the world. They find that while incidence overall declined substantially from 1990-2002, it was entirely on the external trade; sellers' incidence on domestic trade remained constant. Similar investigations are likely to

provide a useful context for regional integration policy in many countries and economic areas around the world where separatism and economic integration are important concerns.

Notice that the trade flows in (5) are invariant to a uniform rise in trade costs (including costs of internal shipment). This follows because (6)-(7) imply that raising all t_{ij}'s by the factor $\lambda > 1$ will raise each Π and P by the factor $\lambda^{1/2}$. This formal homogeneity property has useful empirical content: if the world really were getting smaller uniformly, the gravity model would be unable to reveal it. The empirical literature tends to indicate little change in gravity coefficients (see especially Anderson and Yotov, 2009 and 2010), contrary to intuition about globalization driven by falling communications costs and improving quality of transport but consistent with uniform shrinkage of resistance to trade.

Anderson (1979) was the first to derive gravity from the Armington/CES preference structure, noting that Armington preferences implied a bilateral trade flow gravity equation of the form of (5) that would require controlling for the importer and exporter trade cost indexes. By using a units choice to set all equilibrium factory gate prices equal to 1, Anderson's 1979 derivation concealed how (5)-(7) formed a conditional general equilibrium module that would be the foundation for the very useful comparative statics to come a generation later. The comparative statics of inward and outward multilateral resistance were first used by Anderson and van Wincoop (2003). Recognition that multilateral resistance is interpreted as incidence is in Anderson and Yotov (2009).

3.2 Supply Side Structure

An alternative derivation of a mathematically equivalent structural gravity model was proposed by Eaton and Kortum (2002), based on homogeneous goods on the demand side, iceberg trade costs, and Ricardian technology with heterogeneous productivity for each country and good due to random productivity draws from a Frechet distribution. Despite CES structure for the intermediate goods demand, in equilibrium the share of goods demanded from i by country j is determined only on the supply side; the influence of σ disappears into

a constant term. In equilibrium each country will be assigned a subset of the goods, and except for knife-edge cases it is the only supplier of these goods. The bilateral trade flows obey the same equations as (5)-(7). $1-\sigma$ is interpreted as $-\theta$ where θ is the dispersion parameter of the Frechet distribution. In contrast to the Armington/CES model, all action is on the extensive margin of trade. Eaton and Kortum derive their model for one 'sector' only, a specification generalized by Costinot and Komunjer (2008), so that θ_k is the dispersion parameter for the distribution describing productivity draws in sector k.

The Ricardian structure of supply leads to a very simple general equilibrium superstructure, an appealing feature that has led to a growing literature combining estimation and simulation. General equilibrium superstructure is discussed below in section 6.

Chaney (2008) derives a similar supply side gravity structure based on Ricardian productivity draws from a Pareto distribution where the dispersion parameter of the Pareto distribution plays essentially the same role as θ in the Eaton-Kortum model. For each firm, changes in variable trade costs act on the intensive margin, but for the total sectoral bilateral trade flow these effects disappear and the aggregate effect is effectively on the extensive margin of trade. Chaney's model includes a fixed cost of export for monopolistically competitive firms, and in equilibrium the elasticity of substitution affects the pattern of trade by being part of the elasticity of equilibrium trade volume with respect to the fixed cost.

3.3 Zeroes

In practice, many potential bilateral trade flows are not active. The data presented to the analyst may record a zero that is a true zero or it may reflect shipments that fall below a threshold above zero. In addition there may be missing observations that may or may not reflect true zeroes. The prevalence of zeroes rises with disaggregation, so that in finely grained data a large majority of bilateral flows appear to be inactive. Finally, over time, the small bilateral flows in finely disaggregated data appear to wink on and off. The zeroes present two distinct issues for the analyst: appropriate specification of the economic model

and appropriate specification of the error term on which to base econometric inference. Discussion of the specification of the error term is deferred to the section on estimation.

In specifying the economic model, zero trade flows present a problem for the CES/Armington model of demand and the Eaton-Kortum supply side structure. With elasticities of substitution greater than one (or the equivalent dispersion/comparative advantage parameter restriction for the Eaton-Kortum model), the empirically relevant case, some volume will be purchased no matter how high the price. One way to rationalize zeroes is to modify the demand specification so as to allow 'choke prices' above which all demand is choked off. A start is made by Novy (2010) who derives gravity in a highly restricted one slope parameter translog expenditure function case that allows for zeroes in demand.[4] More general translog treatments are feasible and desirable. Anderson and Neary (2005) present a general homothetic preferences structure, showing that multilateral resistance is defined and solved from a similar equation system once the functional form and its parameters are specified, along with data on shipment and expenditure shares.

An alternative economic specification explanation retains CES/Armington preferences and rationalizes zeroes as due to fixed costs of export facing monopolistic competitive firms. If no firm in i is productive enough to make incurring the fixed cost of exporting to j profitable (given the cost of production in i, variable trade cost t_{ij} and willingness to pay in j), then zero trade results. Helpman, Melitz and Rubinstein (HMR, 2008) develop this idea. The selection effect determines which markets are active and also determines a volume effect V_{ij} due to productivity heterogeneity among firms whereby markets that are active have a greater or lesser numbers of firms active depending on the same selection mechanism. The gravity model becomes

$$X_{ij}^k = \frac{E_j^k Y_i^k}{Y^k} V_{ij}^k \left(\frac{t_{ij}^k}{P_j^k \Pi_i^k} \right)^{1-\sigma_k}$$

[4]Novy's aggregate bilateral OECD trade flow data contain no zeroes, so this feature is not exploited yet.

$$
\begin{aligned}
(\Pi_i^k)^{1-\sigma_k} &= \sum_j \left(\frac{t_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{V_{ij}^k E_j^k}{Y^k} \\
(P_j^k)^{1-\sigma_k} &= \sum_i \left(\frac{t_{ij}^k}{\Pi_i^k}. \right)^{1-\sigma_k} \frac{V_{ij}^k Y_i^k}{Y^k}.
\end{aligned}
$$

HMR report results suggesting that this mechanism is indeed potent, and that inference without accounting for it biases estimates of the variable trade costs downward.

The key mechanism is a Pareto productivity distribution of potential trading firms. The Pareto distribution is capable of capturing the empirical observation that the largest and most productive firms export the most and to the most destinations. The Pareto distribution allows a tractable estimation procedure that requires only aggregate bilateral trade data, an important advantage because firm level trade data is not widely available. In practice, identification of the parameters in estimating the HMR model requires a plausible exclusion restriction — a proxy for the fixed cost of export that is not also a proxy for the variable cost of trade. HMR use common religion, a specification that many find dubious.

An important challenge for the future is combining the HMR mechanism with the translog expenditure system. A potent objection to the CES demand structure in monopolistic competition is that it implies constant markups. The translog allows variable markups. And it is apparently far more tractably manipulated into a gravity representation.

3.4 Discrete Choice Structure

The third alternative model of structural gravity is based on modeling individual discrete choice in a setting where the individual trader faces costs or receives benefits not observable to the econometrician. Of all possible bilateral pairs, the trader chooses one because it yields the greatest gain. A population of such traders has observable characteristics such as bilateral distance that condition the probability of each choice, the econometrician observes the resulting masses allocated and uses a probability model to structure statistical inference An early attempt on these lines was made by Savage and Deutsch (1960) and followed by

Leamer and Stern (1970). Several problems with the model limited its appeal. It did not offer a rationale for the linear homogeneity of the mass variables in gravity and its characterization of cross effects did not have a sound rationale.

Discrete choice modeling was greatly advanced by McFadden (1973), who proved that under plausible restrictions in this setting (the random variable, to the econometrician, results in the observed choices following the Type 1 extreme value distribution), the resulting probability model is the multinomial logit. Building on the multinomial logit, it is easy to generate a structural gravity model. This reasoning has rationalized recent work on models of migration (e.g., Grogger and Hanson, 2008, and Beine, Docquier and Ozden, 2009).

It is straightforward to combine the discrete choice setup with the market clearance conditions to derive the buyers' and sellers' incidence of trade costs exactly as in the preceding models. The development is postponed to the next section, but is noted here because exactly the same reasoning applies to goods traders making discrete choices where to sell or buy their goods. Thus the discrete choice probability model rationalizes structural gravity equally well.

It may be fruitful to explore the applicability of two-sided matching models in the trade context as well as the job market context.

4 Estimation

As an empirical model, gravity is fundamentally about inferring trade costs in a setting where much of what impedes trade is not observable to the econometrician. What is observable are the trade flows and a set of proxies for various types of trade costs, along with direct measures of some components of trade costs. Most issues with modeling trade costs are discussed in Anderson and van Wincoop (2004). Since that time there have been several notable advances in modeling and inferring trade costs.

Two of the advances deal with the implications of zeroes in the bilateral trade flow data. One view of zeroes is that they stand for flows too small to report, an interpretation that

indeed represents reporting practices of government trade ministries. Interpreting zeroes in this way, it is legitimate to drop the zero observations from estimation because there is no economic significance to the zeroes relative to the non-zero observations.

In the presence of heteroskedastic errors, Santos-Silva and Tenreyro (2006) point out that inconsistent estimation arises from the usual econometric gravity practice using logarithmic transforms of (5) augmented with a normal disturbance term and estimated with Ordinary Least Squares (OLS). Since the data has a lot of zeroes, the disturbance term must have a substantial mass at very small values, violating the normal distribution assumption. They propose instead to model the disturbance term as generated from a Poisson distribution, leading to estimation with a Poisson Pseudo-Maximum Likelihood (PPML) technique. Their results show that PPML leads to smaller estimates of trade costs compared to OLS.

The heteroskedastic error problem identified by Santos-Silva and Tenreyro is important, but their solution has not convinced all researchers. Martin and Pham (2008) argue based on Monte Carlo simulations that when heteroskedasticity is properly controlled, Tobit estimators outperform PPML when zeroes are common. Heteroskedasticity is likely to be attenuated using size-adjusted trade X_{ij}/Y_iE_j as the dependent variable, as advocated by Anderson and van Wincoop (2003, 2004).

An alternative view of zeroes, encountered above, is that economically meaningful selection generates the zeroes. All firms in origin i face fixed costs of entering exporting to any particular destination j, and only the sufficiently productive ones can afford to pay the fixed cost. When a destination j is so expensive to reach that no firm in i can afford the fixed cost, zeroes are generated in the data. In this case, OLS estimation without accounting for selection is biased for two reasons; the standard left censored selection reason and also because, for bilateral pairs with positive flows there is a volume effect due to selection of firms along with the bilateral trade cost t_{ij} that is the object of investigation in OLS or PPML estimation. HMR find that their technique also results in lower cost estimates than does OLS. (They report that estimation with Poisson error terms as opposed to normal ones

does not alter their findings.)

In principle, an economic model of zeroes is attractive, but many researchers are suspicious of the exclusion restriction used by HMR to identify their volume effect. They assume that common religion affects the fixed cost of export but not the variable cost t_{ij}. Moreover, the tractability of the HMR model depends on a restrictive distributional assumption on the productivity draws distribution of firms, which in turn is a specialization of a particular model of monopolistic competition that is not applicable to all sectors.

Anderson and Yotov (2010) report that estimation with PPML, HMR or OLS leads to essentially identical results for buyers' and sellers' resistance and Constructed Home Bias because it leads to gravity coefficients that are almost perfectly correlated. The homogeneity property of (5)-(7) implies that only relative trade costs can be inferred by gravity, hence the differences in techniques effectively amount to different implicit normalizations. Anderson and Yotov report this near perfect correlation finding based on estimation with the three techniques over 18 3 digit manufacturing sectors, 76 countries and 13 years of data.

4.1 Traditional

Some researchers continue to use a traditional form of the gravity model, presumably in the belief that the structural model featured above is not sufficiently well established. It seems useful to review a generic traditional model along with my objections.

A typical traditional gravity model regresses the log of bilateral trade on log trade costs proxied by a vector of bilateral variables that are not at issue here, log GDP for origin and destination, and log population for origin and destination. In addition, a number of authors include remoteness indexes of each countries distance from its partners, atheoretic measures that are inadequate attempts to control for multilateral resistance. (Anderson and van Wincoop, 2003, report significant differences between gravity estimated with remoteness and with multilateral resistance.)

The first objection to the traditional model is its aggregation, which causes two prob-

lems. There is aggregation bias due to sectorally varying trade costs and sectorally varying elasticities of trade with respect costs (see Anderson and van Wincoop, 2004, for analysis and Anderson and Yotov, 2009 and 2010 for evidence on downward bias). The second aggregation problem is specification bias because GDP is a value added concept with a variable relationship to gross trade flows. Much recent attention to the vertical disintegration of production and its international aspect emphasizes the variable intertemporal relationship of gross trade to GDP and its variation across countries is also significant. Disaggregation and use of the appropriate sectoral output and expenditure variables fixes both problems.

The second objection is omitted variable bias from the perspective of the structural gravity model — the traditional model leaves out multilateral resistance. Multilateral resistance has only low correlation with remoteness indexes, and the omitted variable will be correlated with the other right hand side variables and thus bias estimation. The traditional model's inclusion of mass variables such as GDP and population presumably picks up a part of the missing explanatory power of multilateral resistance, since Anderson and Yotov's work shows that multilateral resistance is associated with country size. Estimation with country fixed effects controls appropriately for all these issues.

4.2 Structural

Anderson and van Wincoop (2003) combine (6)-(7) with the stochastic version of (5) to form a full information estimator of the coefficients of the proxies for trade cost such as distance and international borders. Utilizing the unitary elasticities on the E's and Y's, their dependent variable is X_{ij}/Y_iE_j, size-adjusted trade.

An alternative fixed effects estimator controls for the unobservable multilateral resistances and activity variables

$$X_{ij} = x_i m_j t_{ij}^{1-\sigma} \epsilon_{ij}, \tag{13}$$

where ϵ_{ij} is the random error term, x_i is the fixed effect for country i as an exporter and m_j

is the fixed effect for country j as an importer. (13) is less efficient than a full information estimator but seems preferable to most subsequent investigators. Feenstra (2005) argues for the fixed effects estimator because it does not require custom coding, but another and perhaps better reason is that researchers should be suspicious that there may be other country specific unobservables that the fixed effects pick up, but which full information estimation would drive toward spurious results.

A major drawback to fixed effects estimation is its demolition of structure: the econometrician blows up the building to get at the safe inside containing the inferred bilateral trade costs. Fortunately, in the case of structural gravity, it is feasible to reconstruct the building like an archeologist, using structural principles in the form of (6)-(7). Thus Anderson and Yotov (2009, 2010) use fixed effects to estimate (5) in its stochastic form, but then calculate the multilateral resistances by calculating the fitted t_{ij}'s and plugging them into (6)-(7).

This technique is used to 'test' the structural gravity model by comparing the estimates of fixed effects ($x_i m_j$) with the structural gravity term ($Y_i E_j \Pi_i^{\sigma-1} P_j^{\sigma-1}$). The results are remarkably close in an economic sense (the fitted regression line has an estimated elasticity around 0.96, compared to the theoretical value of 1.0) across 76 countries and 18 manufacturing sectors over 13 years. While this result suggests that the constraints that legitimize full information methods are very close to being valid, fixed effects estimation still seems the better, more cautious practice to follow.

Baier and Bergstrand (2009) propose an alternative direct estimator of multilateral resistance based on a Taylor's series approximation of (5). They report reasonably good results, but I suspect that many researchers will be wary of the approximation error. In contrast, the method of Anderson and Yotov avoids approximation error. As Baier and Bergstrand emphasize, the advantage of their method relative to panel estimation with fixed effects is that it avoids the upper bound on the number of fixed effects imposed by typical econometric packages at this writing. (STATA currently imposes a limit of 11,000 independent variables, while 100 countries over 10 years require approximately 200,000 fixed effects and even yearly

estimation requires 20,000.) In principle the Baier and Bergstrand estimator could be used to construct t_{ij}'s and then combined with data on the Y's and E's using (6)-(7) in order to obtain the incidence measures and perform comparative statics with them. The constructed multilateral resistances in Baier and Bergstrand's method can be compared to the point estimates, differences being attributed to random error and approximation error.

4.3 Foreign Affiliate Sales

A large share of international trade is sales by foreign affiliates of multinational firms. Standard trade gravity models include this trade along with that of domestically owned firms. If the trade costs are the same for both types of firms, this treatment is entirely appropriate.

There is reason to believe, however, that the trade cost structure facing foreign affiliate sales differs from that facing domestic firms. For trade in intermediate inputs, information and other transactions costs are reduced for intra-firm trade, but even for horizontal trade there are likely to be transactions cost advantages when a foreign affiliate sells into its 'home' country. This reasoning suggests splitting the home and foreign firms into separate 'sectors' for more accurate and informative inference about trade costs.

This approach to gravity with multinationals follows the conditional general equilibrium strategy, treating total sales as exogenous. It avoids taking a stand on determinants of the location of production. A significant literature that is at least loosely related to gravity attempts to explain this location decision along with the volume of foreign affiliate sales. It is treated below in the discussion of Foreign Direct Investment.

5 Gravity and Factor Flows

Gravity has long been applied to empirically model factor movements. As with trade flows, the model always fits well. But, in contrast to the recent development of an economic structural gravity model of trade, there has been little progress in building a theoretical

foundation. This section sets out a structural model of migration, reviews promising steps toward a structural model of Foreign Direct Investment (FDI) and closes by pointing to the unsolved puzzle of modeling international portfolio capital movements.

5.1 Migration

The decision to migrate is a discrete choice from a menu of locations. Each worker that migrates faces a flow cost common to all workers who migrate in a particular bilateral link, but each worker also has an idiosyncratic component of cost or utility from the move. We may think of an idiosyncratic cost component as plausibly associated with a fixed cost, but in the migration decision the distinction between fixed and variable cost plays no important role because the decision to migrate has no volume decision accompanying it. This stands in contrast to the export selection model of Helpman, Melitz and Rubinstein (2008) where the decision to export and the decision how much to export are distinct.

Let w^i denote the wage at location $i, \forall i$. The worker h who migrates from origin j to destination i faces a cost of migration modeled with iceberg cost factor $\delta^{ji} > 1$, receiving net wage (w^i/δ^{ji}). Worker h's idiosyncratic utility from migration is represented by ϵ^{jih}, private information to him. He chooses to migrate if $(w^i/\delta^{ji})\epsilon^{jih} \geq w^j$ for at least some i. Among alternative destinations he chooses the one with the largest surplus. Suppose that the worker has logarithmic utility. Then his observable component of utility of migration from j to i is $u^{ji} = \ln w^i - \ln \delta^{ji} - \ln w^j$. In this sort of setting, McFadden (1973) showed that if $\ln \epsilon$ had the type-1 extreme value distribution, the probability that a randomly drawn individual would pick any particular migration destination is given by the multinomial logit form.

Building on this insight, migration models subsequently used the multinomial logit to model bilateral migration flows. For two recent examples, see Beine, Docquier and Ozden (2009) and Grogger and Hanson (2008). This section develops a novel gravity model representation of the migration model by making use of the market clearing conditions to derive the appropriate multilateral resistance variables.

At the aggregate level the probability is equal to the proportion of migrants from j (assumed to be identical except for their values of ϵ, that pick destination i. Let N^j denote the population of natives of j. The predicted migration flow from j to i that results from the setup is

$$M^{ij} = G(u^{ji})N^j. \tag{14}$$

where

$$G(u^{ji}) = \frac{\exp(u^{ji})}{\sum_k \exp(u^{jk})}.$$

With logarithmic utility, the migration equation is

$$M^{ji} = \frac{w^i/\delta^{ji}}{\sum_k w^k/\delta^{jk}} N^j. \tag{15}$$

(15) is a structure analogous to the CES demand (in the Armington model) or Ricardian supply (in the Eaton-Kortum model) shares that underpin the trade gravity equation. The connection of the share equation (15) to the structural gravity form of the model is completed by using the labor market balance equations to solve for and substitute out the equilibrium w's.

Define $W^j \equiv \sum_k w^k/\delta^{jk}$ and define the labor force supplied to i from all origins

$$L^i \equiv \sum_j M^{ji}. \tag{16}$$

Also, $N \equiv \sum_j N^j = \sum_i L^i$, the world labor supply N. The labor market clearance equation is

$$L^i = w^i \sum_j ((1/\delta^{ji})/W^j)N^j.$$

Then

$$w^i = \frac{L^i}{\Omega^i N} \tag{17}$$

where

$$\Omega^i = \sum_j \frac{1/\delta^{ji}}{W^j} \frac{N^j}{N}. \tag{18}$$

Using (17) to substitute for the wage in W^j,

$$W^j = \sum_k \frac{1/\delta^{jk}}{\Omega^k} \frac{L^k}{N}. \tag{19}$$

Substituting for the wage in (15) using (17) yields the structural gravity equation of migration:

$$M^{ji} = \frac{L^i N^j}{N} \frac{1/\delta^{ji}}{\Omega^i W^j}. \tag{20}$$

The first ratio represents the migration pattern of a frictionless world. The implication is that in a frictionless world, populations originating in j would be found in equal proportions to their share of world population in all destinations: $M^{ji}/L^i = N^j/N$. The second term represents the effect of migration frictions. The bilateral migration friction δ^{ji} reduces migration. It is divided by the the product of weighted averages of the inverse of migration frictions, one for inward migration to i from all origins and one for outward migration from j to all destinations. The system (18)-(19) can be solved for the Ω's and W's (subject to a normalization). Their interpretation and their connection to multilateral resistance in the more familiar trade gravity model is easier to see in the case where utility is generalized to the log of a Constant Relative Risk Aversion function.[5]

Let the coefficient of relative risk aversion be θ. In this case (20) becomes

$$M^{ji} = \frac{L^i N^j}{N} \left(\frac{\delta^{ji}}{\bar{\Omega}^i \bar{W}^j} \right)^{1-\theta}$$

[5]A tractable gravity equation results from (14) by using a restriction on utility to convert $\exp u^{ji}$ into a tractable form. When utility is given by the log of any power function of the wage net of migration costs, the CES-type form of gravity results, with consequent ease of estimation and resemblance to the trade flow structural gravity model.

where

$$\bar{\Omega}^i \equiv \left[\sum_j \frac{(\delta^{ji})^{1-\theta}}{\bar{W}^j} \frac{N^j}{N} \right]^{1/(1-\theta)}$$

and

$$\bar{W}^j \equiv \left[\sum_i \frac{(\delta^{ji})^{1-\theta}}{\bar{\Omega}^i} \frac{L^i}{N} \right]^{1/(1-\theta)} .$$

Here, $\bar{\Omega}^i$ and $\bar{W}^j$ are CES price indexes of migration frictions, one for inward ($\bar{\Omega}^i$) and one for outward ($\bar{W}^j$) migration frictions. These equations are exactly analogous to Anderson and van Wincoop's gravity, inward and outward multilateral resistance equations for trade, but applied to migration. As with the trade gravity model, outward multilateral resistance gives the sellers' incidence of the migration costs on average while the inward multilateral resistance gives the 'buyers' incidence of migration costs. (18)-(20) results from the special case $\theta = 2$.

Ω and W are general equilibrium concepts as is clear because their solution in the simultaneous systems above involves every bilateral migration cost in the world. They are conditional general equilibrium concepts because the L's are endogenous in a full general equilibrium. It is possible in a Ricardian production setting to combined the migration system with the trade gravity model to derive equilibrium labor supplies that are functions of the incidence of both migration frictions and trade frictions.

As with trade gravity models, $\bar{\Omega}$'s and $\bar{W}$'s can be computed once the δ's are econometrically constructed and the labor supplies L^i and population stocks N^j are observed. A normalization is needed. (See Anderson and Yotov, 2009, for details.)

A similar model has been applied to services trade by Head, Mayer and Ries (2009). Instead of actually changing locations, the foreign worker does the job in his home location. The cost of migration becomes the cost of monitoring the distant worker. Worker productivities in each location have the Frechet distribution, as in the Eaton-Kortum model. The firm selects workers so as to minimize the log of the delivered unit labor cost. Then the distribution of log productivities takes the Gumbel form. The fraction of service jobs in origin i

going to workers in location j has the multinomial logit form. The total numbers of workers and of jobs in each location enter the model in the same way as in the migration model. I suspect that the choice between off-shoring the service job and migrating the worker can be fruitfully addressed with some combination of the two models.

The preceding treatment applies to a stationary equilibrium where the L's are the result of M's fully adjusting labor supplied at each location to its equilibrium value given the initial stocks of labor $\{N^j\}$ and the set of migration frictions, the δ's. In adapting the model to fit actual data, the N's, L's and M's are observed at points in time, and with panel data the observations are linked over time.

If the sequence of observations is regarded as reaching the static equilibrium each period, the observed migration is just that amount needed to reach the equilibrium in each period. This model would be consistent with naive expectations about future wages, or with a pure guest worker model in which migration is determined by contemporaneous variables only. So in principle under this interpretation the preceding model could be applied at each date, all variables now having a time subscript.

The alternative is a dynamic model in which the migrants form expectations about the sequence of future wages based on underlying expectations about the future evolution of the distribution of trade frictions, the populations, and, as we will see following the development of the integrated trade and migration model, variables that predict the demand for labor at all locations. This sophistication requires a big increase in complexity, with dubious applicability of rational expectations to unskilled workers.

The other issue raised by thinking of dynamics is the issue of partial adjustment — migration in any one year may not suffice to reach the static equilibrium of the preceding section. In this case, the standard ad hoc approach of partial adjustment due to quadratic adjustment costs might be applied without too large an increase in complexity.

5.2 Foreign Direct Investment

Foreign direct investment has been successfully explained by gravity structures without a theoretical foundation. More recent work has made progress on foundations. Satisfactory foundations are more difficult to find for at least two reasons. First, the location of production question must be answered in an upper level general equilibrium model, which requires taking a stand on one of many possible production, preference and market structures restricted so as to produce tractable results. Second, the determinants of location depend on whether the good in question is vertically or horizontally linked to other sources of firm profits.

A key element in explaining the location of horizontally linked production is the proximity-concentration tradeoff: a firm with fixed cost reduces per unit production cost by concentrating production at one location but can save distribution costs by allocating production in proximity to markets. Even under strong restrictions, the models obtained so far are nonlinear and require approximation to be taken to data.

Helpman, Melitz and Yeaple (2004) model interaction between horizontally linked exports and foreign affiliate sales, where the firm chooses between exporting from home or investing abroad and selling from a foreign plant. They are able to draw inferences from aggregate data by modeling heterogeneous productivity of firms with a Pareto distribution. Fixed costs of export and of investing abroad serve to select firms into non-traders, exporters and multinationals with ratios that vary market by market due to trade costs modeled as transport costs and tariffs only, omitting the usual gravity variables. Their empirical application with US data obtains fairly good results in explaining the ratio of exports to foreign affiliate sales with a linear approximation to their underlying nonlinear model. The model fits much less well than standard export gravity equations, not surprising because the dependent variable is different and the question addressed is more difficult to answer.

Kleinert and Toubal (2010) extend Helpman, Melitz and Yeaple to allow for fixed setup costs that rise with distance, a wrinkle that can explain why foreign affiliate sales can fall rather than rise with distance as the earlier proximity-concentration tradeoff suggested. They

also derive a gravity-type relationship from two other structures, a vertical integration model and a two country factor proportions model of fragmentation.

Bergstrand and Egger (2007) offer a gravity model of FDI derived from the knowledge-based capital theory of horizontal multi-national enterprises. Their objective is a full general equilibrium model that can explain trade, foreign affiliate sales and foreign direct investment. They simulate a theoretical model that generates nonlinear relationships between exports, affiliate sales and their exogenous determinants. Then they fit an approximate 'empirical' relationship to the generated data and take the same relationship to actual data, with some success. A limitation of their model is that, though the factor proportions model with 3 factors is used to explain simultaneous exports and affiliate sales, the countries in their simulation setup have identical endowment proportions and differ only in size.

Keller and Yeaple (2009) develop a gravity model of vertically integrated intra-firm trade featuring trade costs with two elements, a standard iceberg trade cost and a communication cost that rises with the complexity of the firm's technology. Input complexity raises technology transfer costs while the costs of embodied technology transfer are independent of complexity and increasing in trade costs. An increase in trade costs reduces foreign affiliate sales and this effect is strongest in the most complex sectors. In contrast, an increase in trade costs reduces the imports of foreign affiliates and this effect is weakest in the most complex sectors. Like the standard trade gravity model, Keller and Yeaple's model of foreign affiliate sales permits inference about trade costs from observable trade flows.

Keller and Yeaple report fairly good results estimating the model using confidential data on U.S. multinational firm activity from the Bureau of Economic Analysis. The role played by communication cost interacting with technological complexity thus appears likely to be helpful in explaining the rising share (in total trade) of intra-firm trade and also the rising share of trade in intermediates.

An alternative strategy along the lines of the conditional general equilibrium approach outlined above for migration appears useful. The migration decision model of section 5.1

could apply to FDI since the location decision for a plant is similar to the location decision of a migrant. (Unlike migration but like trade, FDI involves a volume decision along with a participation decision.) The rate of return on investment could be taken as exogenous in a conditional general equilibrium approach just as wages are taken as exogenous in the migration gravity model, while market clearing conditions apply just as in the migration model. Idiosyncratic cost factors would apply to the various investment projects, just as they do to the individual migrants. The Keller-Yeaple model of vertically integrated intra-firm trade offers a structure for identifying one type of cost. A weakness in the extension by analogy is in risk diversification. Migrants cannot diversify their risks, but firms can, though with limited possibilities that may be very limited for FDI. The potential risk diversification would modify the utility derived from each location choice. The discrete choice approach faces truly formidable modeling challenges in endogenizing the investment rates of return, unlike the wage equation suggested by the migration model.

A promising start on these lines is by Head and Ries (2008). Potential acquisitions go to the highest bidder, who bids based on his anticipated return net of monitoring costs that rise with distance and other standard gravity variables. The probability of the winning bid going to source country i takes the multinomial logit form. The mass variables are the stocks of projects in each host country and each source country's share of world bidders.

5.3 Portfolio Investment

Martin and Rey (2004) offer the first gravity type model of international portfolio investment. The coefficient of relative risk aversion plays the role, in equilibrium, of the elasticity of substitution in the CES demand specification. While appealing as a rationale for the gravity application of Portes and Rey (2005), the Martin and Rey model does not provide a fully satisfactory foundation for gravity models of investment flows because (i) trade is assumed to be frictionless, (ii) investment costs are uniform, and (iii) most important, the analysis is restricted to two countries. The third party effects that play a big role in the gravity model

of trade (and of migration) cannot be treated.

6 Integrated Superstructure

The gravity model nests inside a general equilibrium superstructure. As pointed out in Anderson and van Wincoop (2004), modularity implies that the problem of resource and expenditure allocation across sectors in the general equilibrium superstructure can be treated separably from the gravity module problem of distribution within sectors to destinations or from origins. Consistency between the two levels of the problem requires fixed point calculations in general, but the economy of thought and computation due to separability is extremely useful, and in particular makes it possible to integrate gravity with a wide class of general equilibrium production models. So far, only very simple production models have been used for full general equilibrium comparative statics, but I anticipate that this situation will change.

The simplest production structure is an endowments economy. Anderson and van Wincoop (2003) use the endowments model to calculate the effect of eradicating the US-Canada border on their estimated gravity model of trade between US states, Canadian provinces and the aggregated rest of the world.

Another attractive candidate is the Ricardian production model. Eaton and Kortum (2002) nest gravity inside a Ricardian model of production, a choice followed by a host of subsequent researchers such as Arkolakis (2008). An important feature of these models is the action on the extensive margin, as industries arise or disappear. In the Eaton-Kortum model of 2002, the extensive margin is the only margin. Arkolakis and others have variants in which both extensive and intensive margins are active. This is an important feature because disaggregated trade data and especially firm level data indicate that both margins are active.

Between the two extremes of zero and infinite elasticity of transformation of the endowments and Ricardian models lie a host of more complex production structures in which action

occurs on the intensive margin of production when relative prices change, leading to another channel of interaction between the gravity modules in each sector (and resulting buyers' and sellers' incidences) and the pattern of production. Consistency between the modules is achieved by using (9) to normalize the Π's in each sector. I think the future will see work with these more complex general equilibrium features.

Migration of labor and capital in the form of FDI has been given a complete gravity representation in this essay. In the integrated superstructure it can be treated simultaneously with the trade modules. In this setting, multilateral resistance in trade has significant effects on migration and vice versa. I anticipate that development of this link will be useful.

A number of authors have constructed integrated models that motivate econometric work aimed at discriminating between one or another specification of the upper level production and market structure. A summary of work on these lines is in Feenstra (2004), chapter 5, where the main focus is on the link between gravity and increasing returns to scale. Research has continued on these lines, but I will not review it here.

I think the gravity model is a poor vehicle for inferences about returns to scale, market structure and the global general equilibrium links between economies. This essay and my previous work argue that gravity is about the distribution of given amounts of goods in each origin drawn by given amounts of expenditure in each destination, enabling inference about trade costs from the deviation of observed distribution from the frictionless equilibrium. The determinants of total shipments and total expenditures are irrelevant to this inference because country fixed effects are a consistent control that does not require taking a stand on any particular production or market structure model. Conversely, the cross section variation of bilateral trade does not seem likely to have much useful information about the determination of national total shipments or expenditure. Interdependence is so deeply wound between these variables in the full general equilibrium model that inference about structure seems implausible. In contrast, simulation models look reasonably promising as a source of insight.

7 Conclusion

This idiosyncratic review of work on the gravity model suggests that the story is not over, so a conclusion can only point to potential future chapters. Distribution broadly defined consumes a very large share of the world's resources and gravity has proven to be the most generally useful empirical model for understanding the distribution of goods and factors of production. It appears to work well at almost any scale.

The progress in structural modeling of gravity has yielded three distinct rationales for the same observationally equivalent model of the distribution of economic flows between origins and destinations, one based on the demand side (the CES/Armington model), one based on the supply side (the Eaton-Kortum model), and one based on a discrete choice model of the individual actor transferring the goods or factors. Further work may suggest ways to discriminate between these.

The structural modeling of gravity imposes trade separability, permitting gravity modules to be nested inside a wide range of general equilibrium superstructures. Future work with simulation models may suggest which of many candidate general equilibrium production models do better.

The problem of zeroes in the trade and factor flows data has been addressed with some success, particularly by Helpman, Melitz and Rubinstein. But I expect future work to do better. The CES framework (with elasticity of substitution greater than one) is unsuitable for describing small amounts of trade. The translog cost function, in particular, seems likely to yield better descriptions and better understanding of why so many potential flows are equal to zero. This is so even if, as in HMR, fixed export costs play an important role in selecting firms to export.

Incidence measures produced by Anderson and Yotov have been featured in this review. If the profession agrees that they are as interesting and useful as they appear to me, more work is needed to see how believable the measures are. As it stands, they are completely reliant on CES structure. How well does the CES do in representing the world economy?

This is an especially important question in light of the zeroes question in the preceding paragraph. I look forward to development of the translog case to help answer this question.

References

[1] Anderson, James E. 1979. "A theoretical foundation for the gravity equation", American Economic Review 69, 106-116.

[2] Anderson, James E. and J. Peter Neary. 2005. *Measuring the Restrictiveness of International Trade Policy*, Cambridge: MIT Press.

[3] Anderson, James E. and Eric van Wincoop. 2004."Trade Costs", *Journal of Economic Literature*, 42, 691-751.

[4] Anderson, James E. and Eric van Wincoop. 2003. "Gravity with Gravitas", *American Economic Review*, 93, 170-92.

[5] Anderson, James E. and Yoto V. Yotov. 2009. "The Changing Incidence of Geography," *American Economic Review*, forthcoming.

[6] Anderson, James E. and Yoto V. Yotov. 2010. "Specialization: Pro- and Anti-globalizing, 1990-2002", Boston College.

[7] Arkolakis, Costas. 2008. "Market Penetration Costs and the New Consumers Margin in International Trade", NBER Working Paper No. 14214.

[8] Baier, Scott L. and Jeffrey H. Bergstrand. 2007. "Do free trade agreements actually increase members' international trade?," *Journal of International Economics*, 71(1), 72-95.

[9] Baier, Scott L. and Jeffrey H. Bergstrand, 2009. "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects using the Gravity Equation", *Journal of International Economics*, vol. 77, no. 1, .

[10] Baier, Scott L., and Jeffrey H. Bergstrand, 2001. "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity", *Journal of International Economics*, vol. 53, no. 1, 1-27.

[11] Beine, Michel, Frederic Docquier and Caglar Ozden. 2009. "Diasporas", World Bank.

[12] Bergstrand, Jeffrey H., 1985. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence", *Review of Economics and Statistics*, Vol. 67, No. 3, August 1985, 474-481.

[13] Bergstrand, Jeffrey H., 1989, "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade," *Review of Economics and Statistics*, Vol. 71, No. 1, February 1989, 143-153.

[14] Bergrstrand, Jeffrey H. and Peter Egger. 2007. "A Knowledge- and Physical-capital Model of International Trade Flows, Foreign Direct Investment and Multinational Enterprises", *Journal of International Economics*, 73, 278-308.

[15] Bergrstrand, Jeffrey H. and Peter Egger. 2009. "Gravity Equations and Economic Frictions in the World Economy", in Daniel Bernhofen, Rodney Falvey, David Greenaway and Udo Krieckemeier, eds., *Palgrave Handbook of International Trade*, Palgrave-Macmillan Press, forthcoming.

[16] Chaney, Thomas,. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review*, 98, 1707-21.

[17] Deardorff, Alan, 1980. "The General Validity of the Law of Comparative Advantage", *Journal of Political Economy*, ¨Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade", *Econometrica*, 70: 1741-1779.

[18] Feenstra, Robert. 2004. *Advanced International Trade: Theory and Evidence*, Princeton, NJ: Princeton University Press.

[19] Grogger, Jeffrey and Gordon H. Hanson. 2008. "Income Maximization and the Selection and Sorting of International Migrants", NBER Working Paper No. 13821.

[20] Head, Keith, Thierry Mayer and John Ries. 2009. "How Remote is the Off-shoring Threat?", *European Economic Review*, 53, 429-44.

[21] Head, Keith and John Ries. 2008. "FDI as an Outcome of the Market for Corporate Control: Theory and Evidence", *Journal of International Economics*, 74, 2-20.

[22] Helpman, Elhanan, Marc J. Melitz and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes", Harvard University, *Quarterly Journal of Economics*, 123: 441-487.

[23] Helpman, Elhanan, Marc J. Melitz and Stephen R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms", *American Economic Review*, 94, 300-316.

[24] Keller, Wolfgang and Stephen R. Yeaple. 2009. "Gravity in the Weightless Economy", NBER Working Paper No. 15509.

[25] Kleinert, Jörn and Farid Toubal, 2010. "Gravity for FDI", *Review of International Economics*, 18 (1), 1-13.

[26] Leamer, Edward E. and Robert M. Stern, 1970. *Quantitative International Economics*, Chicago: Aldine Press.

LeamerLevinsohn, Leamer, Edward E. and James Levinsohn, 1995. "International Trade Theory: the Evidence" in Gene M. Grossman and Kenneth Rogoff, eds., *Handbook of International Economics, vol. 3*, Amsterdam: Elsevier Science B.V.

[27] Martin, Philippe and Helene Rey, (2004), "Financial Super-markets: Size Matters for Asset Trade", *Journal of International Economics*, 64, 335-61.

[28] Martin, Will and Pham S. Cong (2008), "Estimating the Gravity Equation when Zero Trade Flows Are Frequent", World Bank.

[29] Matsuyama, Kiminori, 2007, "Beyond Icebergs: Towards a Theory of Biased Globalization", *Review of Economic Studies*, 74, 237-53.

[30] McCallum, John (1995) "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 1995, 85(3), pp. 615-623.

Novy10, Novy, Dennis. 2010. "International Trade and Monopolistic Competition without CES: Estimating Translog Gravity", Warwick University.

[31] Redding, Stephen and Anthony J. Venables. 2004. "Economic Geography and International Inequality", *Journal of International Economics*, 62(1), 53-82.

[32] Portes, Richard and Helene Rey, (2005), "The Determinants of Cross-Border Equity Flows", *Journal of International Economics*, 65, 269-96.

[33] Ravenstein, Edward George, 1889, "The Laws of Migration." *Journal of the Royal Statistical Society*, Vol. 52, No. 2. (June, 1889), pp. 241305.

[34] Samuelson, Paul A. (1952), "The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent", *The Economic Journal*, 62, 278-304.

[35] Santos Silva, Jorge and Sylvana Tenreyro. 2006. "The Log of Gravity", *Review of Economics and Statistics*, Vol. 88, No. 4: 641-658.

[36] Savage, I. Richard, and Karl W. Deutsch, 1960. "A Statistical Model of the Gross Analysis of Transactions Flows", *Econometrica*, 28, 551-72.

[37] Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.