Notes on the "Theoretical" Gravity Model of International Trade

Ben Shepherd*

Niehaus Center, Princeton University

& GEM, Sciences Po

This Version Dated: November 25, 2008

Abstract

I derive in detail the version of the gravity model of trade due to Anderson and Van Wincoop (2003, 2004), which has become the de facto standard in empirical work. In particular, I emphasize the model's micro-foundations in terms of Dixit-Stiglitz preferences and some basic macroeconomic identities. I briefly discuss the implications of the micro-founded gravity model's functional form for estimation and counter-factual simulations in real-world settings, and address some particular issues that arise when using sectorally disaggregated trade data.

*These notes are intended as background for participants at ARTNeT's 2008 Capacity Building Workshop for Trade Research: "Behind the Border" Gravity Modeling (Bangkok, December 15-19, 2008). They are aimed at applied/empirical researchers, and thus present the theory in way designed to highlight its implications for estimation, inference, and simulation. The material draws heavily on the many helpful comments I have received on various applied gravity papers over the last few years. I am particularly grateful to colleagues in the Development Research Group at the World Bank, and in the OECD's Trade and Agriculture Directorate. Please send comments, suggestions, and corrections to: bshepher@princeton.edu, and bashepherd@gmail.com.

1 Introduction

Although it started life based on little more than common sense and a nice analogy, the gravity model has come a long way in recent years. There are now a number of gravity-like models that can be derived from solid micro-foundations. One of the most commonly used in empirical work is the model due to Anderson and Van Wincoop (2003, 2004) ("AvW"). This note sets out in some detail the AvW gravity model, emphasizing its theoretical underpinnings and the role they play in driving estimation, inference, and simulation in applied work.

One of the more appealing features of the AvW model is that it is relatively easy to derive. There are basically four steps. First, we need to develop in detail the model's consumption side (Section 2). It relies on Dixit-Stiglitz "love of variety" preferences, and we can use that literature to state and solve the consumer's problem in order to produce a set of direct demand functions. Next, we address the model's production side (Section 3). A very large number of symmetric firms in each country engage in monopolistic competition. Since there are so many firms, strategic interactions disappear and firms adopt a simple markup pricing rule. The third step (Section 4) is to specify a price pass-through equation, which relates domestic and foreign prices. This can be done simply by introducing standard "iceberg" variable trade costs, akin to an ad valorem tariff. Finally (Section 5), we impose equality of sectoral incomes and expenditures, aggregate demand by sector, and attack the model with some additional algebra, in order to produce the AvW gravity equation. The paper then concludes with a brief discussion of some of the most important implications of the AvW model for applied trade policy work.

It is important to emphasize that although the AvW model is very commonly used and is treated by many researchers as a standard or baseline gravity model, it is by no means the only theoretically grounded gravity model. What this means, though, is that if you want to include a variable that AvW suggests is not relevant (like per capita GDP), or if you want to set up fixed effects in a way that does not sit easily with the AvW formulation, then you need to justify your approach explicitly in terms of some alternative theoretical assumptions.

2 Consumption Side

Figuring out the consumption side of the AvW model means stating the consumer's problem–maximizing utility subject to a budget constraint–then solving it to produce a set of direct demand functions. In this model, the consumer's preferences take the Dixit-Stiglitz "love of variety" form. Each sector consists of a large number of differentiated product varieties, which are substitutable in consumption with a given (constant) elasticity. The form taken by the utility function is quite standard in the post-Krugman trade literature.

Consider a world of C countries indexed by i. I assume from the start that countries can trade with each other, and thus that consumers in one country can potentially purchase varieties from any other country. For the moment, trade is costless. (We'll get to costly trade later on.)

Consumers are identical in each country, and maximize CES utility across a continuum of varieties (index v) in K sectors (indexed by k) with the following form:

$$U_i = \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}} \tag{1}$$

The set V_i defines the range of varieties that is consumed in country i. As usual, I use $x_i^k(v)$ to indicate the quantity of variety v from sector k consumed in country i, and $p_i^k(v)$ to indicate its unit price. I use function notation because of the continuum of varieties. In the version of the model with a discrete number of varieties, v becomes a subscript, and the integrals are replaced with sums. I am using the continuum version of Dixit-Stiglitz for two reasons. First, strategic interaction on the production side only truly disappears with an uncountably large number of firms. Second, more recent versions of gravity such as HMR use continuum notation, and I would like to keep as close as possible to their formulation so that comparisons are made easier.

The utility function is simply the sum of the sectoral sub-utilities, each of which is weighted equally. That restriction can easily be relaxed by aggregating the sectoral sub-utilities via a Cobb-Douglas utility function, and allowing for different weights. So long as the shares are exogenous to the model, however, the basic results stay the same. See Chaney (2008) for an example of what the alternative expressions look like. AvW and HMR consider, in effect, a single sector so as to avoid cluttering up the algebra with additional indices. But nothing turns on this, and in the present case it is useful to retain some sectoral disaggregation so that we can examine a couple of important data implications that flow from the model in a multi-sector context.

The budget constraint in country i is

$$E_i = \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} p_i^k(v)\, x_i^k(v)\, dv \right\} \equiv \sum_{k=1}^{K} E_i^k \tag{2}$$

where E_i is total expenditure in that country, and E_i^k is country i's total expenditure in sector k.

The consumer's problem is to choose $x_i^k(v)$ for all v so as to maximize (1) subject to (2). The Lagrangian is:

$$\mathscr{L} = \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}} - \lambda \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} p_i^k(v)\, x_i^k(v)\, dv \right\} \tag{3}$$

Taking the first order condition with respect to quantity and setting it equal to zero gives:

$$\frac{\partial \mathscr{L}}{\partial x_i^k(v)} = \frac{1}{1-\frac{1}{\sigma_k}} \left\{ \int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}-1} \left(1 - \frac{1}{\sigma_k}\right) \left[x_i^k(v) \right]^{-\frac{1}{\sigma_k}} - \lambda p_i^k(v) = 0 \tag{4}$$

Define $X^k = \left\{ \int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}}$, regroup terms, and rearrange to get:

$$\frac{\left[x_i^k(v) \right]^{-\frac{1}{\sigma_k}}}{\int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv} X^k = \lambda p_i^k(v) \tag{5}$$

Now rearrange again, multiply through by prices, aggregate over all varieties in a given sector, and then solve for the Lagrangian multiplier:

$$x_i^k(v)\,p_i^k(v) = \lambda^{-\sigma_k}\left[p_i^k(v)\right]^{1-\sigma_k}\left(X^k\right)^{\sigma_k}\left\{\int_{v\in V_i^k}\left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}}dv\right\}^{-\sigma_k} \tag{6}$$

$$\int_{v\in V_i^k} x_i^k(v)\,p_i^k(v)\,dv \equiv E_i^k = \lambda^{-\sigma_k}\left(X^k\right)^{\sigma_k}\left\{\int_{v\in V_i^k}\left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}}dv\right\}^{-\sigma_k}\int_{v\in V_i^k}\left[p_i^k(v)\right]^{1-\sigma_k}dv \tag{7}$$

$$\lambda = \left\{\frac{\int_{v\in V_i^k}\left[p_i^k(v)\right]^{1-\sigma_k}dv}{E_i^k}\right\}^{\frac{1}{\sigma_k}}\frac{X^k}{\int_{v\in V_i^k}\left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}}dv} \tag{8}$$

To get the direct demand function, substitute this expression for the multiplier back into the first order condition (5):

$$\frac{\left[x_i^k(v)\right]^{-\frac{1}{\sigma_k}}}{\int_{v\in V_i^k}\left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}}dv}X^k = \left\{\frac{\int_{v\in V_i^k}\left[p_i^k(v)\right]^{1-\sigma_k}dv}{E_i^k}\right\}^{\frac{1}{\sigma_k}}\frac{X^k}{\int_{v\in V_i^k}\left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}}dv}p_i^k(v) \tag{9}$$

$$\therefore x_i^k(v) = \frac{\left[p_i^k(v)\right]^{-\sigma_k}}{\int_{v\in V_i^k}\left[p_i^k(v)\right]^{1-\sigma_k}dv}E_i^k \tag{10}$$

In the literature these days, it is common to define the "ideal" CES price index for sector k in country i as $P_i^k = \left\{\int_{v\in V_i^k}\left[p_i^k(v)\right]^{1-\sigma_k}dv\right\}^{\frac{1}{1-\sigma_k}}$. So the demand function then takes the slightly less intimidating form of:

$$x_i^k(v) = \left\{\frac{p_i^k(v)}{P_i^k}\right\}^{-\sigma_k}\frac{E_i^k}{P_i^k} \tag{11}$$

Compare (11) with equations (1) and (2) in Helpman et al. (2008), or equations (5.24) and (5.25) in Chapter 5 of Feenstra's textbook.

3 Production Side

The producer's problem in this model is to maximize profit. Assuming a continuum of firms, i.e. an uncountably large number of them, makes this problem much easier to solve. It turns out that strategic interactions disappear, and firms charge a constant markup. In terms of the overall model, this section gives us an equilibrium pricing equation which, with the equilibrium demand equation derived in the previous section, is just about all we need to generate gravity.

Each country i has a measure N_i^k of active firms in sector k. Each firm makes a unique product, so the total (worldwide) measure of products in each sector is $\sum_{i=1}^{C} N_i^k$. In order to produce one unit of its product, a firm must pay a fixed cost f_i^k and a variable cost a_i^k. A typical firm's profit function is therefore:

$$\pi_i^k(v) = p_i^k(v)\, x_i^k(v) - w a_i^k x_i^k(v) - w f_i^k \tag{12}$$

With a continuum of varieties, it does not matter at this point whether we assume Bertrand (price) or Cournot (quantity) competition. If the firms play Bertrand, the first order condition is:

$$\frac{\partial \pi_i^k(v)}{\partial p_i^k(v)} = x_i^k(v) + p_i^k(v)\frac{\partial x_i^k(v)}{\partial p_i^k(v)} - w a_i^k \frac{\partial x_i^k(v)}{\partial p_i^k(v)} = 0 \tag{13}$$

and it can usefully be solved for prices by moving the first and third terms to the right hand side, then dividing through by the partial derivative

$$p_i^k(v) = w a_i^k - \frac{x_i^k(v)}{\frac{\partial x_i^k(v)}{\partial p_i^k(v)}} \tag{14}$$

In order to do something with that expression, we need to know more about the $\frac{\partial x_i^k(v)}{\partial p_i^k(v)}$ term. The partial can be calculated by differentiating the demand function (11) with respect to prices. In doing so, it is important to note that because of the continuum assumption (an uncountably large number of firms), $\frac{\partial P_i^k}{\partial p_i^k(v)} = 0$. In this "large group" case, a small change in one firm's price does not affect the overall price level in its sector. In light of this, we can write:

$$\frac{\partial x_i^k(v)}{\partial p_i^k(v)} = -\sigma_k \left[p_i^k(v)\right]^{-\sigma_k - 1} \left\{\frac{1}{P_i^k}\right\}^{-\sigma_k} \frac{E_i^k}{P_i^k} \equiv -\frac{\sigma_k x_i^k(v)}{p_i^k(v)} \tag{15}$$

The first order condition can therefore be rewritten as:

$$p(v) = wa_i^k + x_i^k(v) \frac{p_i^k(v)}{\sigma_k x_i^k(v)}$$

then rearranged, and solved for prices:

$$p_i^k(v) - \frac{1}{\sigma_k} p_i^k(v) \equiv p_i^k(v)\left(1 - \frac{1}{\sigma_k}\right) = wa_i^k \tag{16}$$

$$\therefore p_i^k(v) = \left(\frac{\sigma_k}{\sigma_k - 1}\right) wa_i^k \tag{17}$$

The second term on the right hand side in equation (17) is simply the firm's marginal cost of production. The term in brackets is a constant (within sector) markup: since the numerator must be greater than the denominator, there is a positive wedge between the firm's factory gate ("mill") price and its marginal cost. Since the wedge only depends on the sectoral elasticity of substitution, it is constant across all firms in the sector.

4 Trade Costs

Thus far, we have not considered the conditions under which international trade takes place. At the moment, the model simply consists of a set of demand functions and pricing conditions for all countries and sectors. As it is, the model describes trade in a frictionless world, in which goods produced in country i can be shipped to country j at no charge. Arbitrage ensures, therefore, identical prices in both countries.

To introduce trade frictions, we can use the common "iceberg" formulation. When a firm ships goods from country i to country j, it must send $\tau_{ij} \geq 1$ units in order for a single unit to arrive. The difference can be thought of as "melting" (like an iceberg) en route to the destination. Equivalently, the marginal cost of producing in country i a unit of a good subsequently consumed in the same country i is wa_i^k, but if the same product is consumed in country j then the marginal cost is instead $\tau_{ij} wa_i^k$. Using this definition, costless trade corresponds to $\tau_{ij} = 1$, and τ_{ij} corresponds to one plus the ad valorem tariff rate. Since the size of the trade friction associated with a given iceberg coefficient does not depend on the quantity of goods shipped, we can treat iceberg costs as being variable (not fixed) in nature.

Taking any two countries i and j, the presence of iceberg trade costs means that the price in country j of goods produced in country i is (from equation (17) above):

$$p_j^k(v) = \left(\frac{\sigma_k}{\sigma_k - 1} \right) \tau_{ij}^k wa_i^k = \tau_{ij}^k p_i^k(v) \tag{18}$$

This result allows us to rewrite the country price index in a more useful (and general) form:

$$P_j^k = \left\{ \int_{v \in V_j^k} \left[\tau_{ij}^k p_i^k(v) \right]^{1-\sigma_k} dv \right\}^{\frac{1}{1-\sigma_k}} \tag{19}$$

Note that this index includes varieties that are produced and consumed in the same country: all the τ_{ii}^k terms are simply set to unity, so as to reflect the absence of internal trade barriers.

5 Gravity With Gravitas

These are all the ingredients required to put together a gravity model with gravitas. The trick is in combining them in the right way.

The gravity model is usually concerned with the value of bilateral trade (x_{ij}^{k}), i.e. exports from country i to country j of a particular product variety. Combining the price equation (18) with the demand function (11) gives:

$$x_{ij}^{k}(v) = p_{ij}^{k}(v)\, x_{j}^{k}(v) \equiv \tau_{ij}^{k} p_{i}^{k}(v) \left\{ \frac{\tau_{ij}^{k} p_{i}^{k}(v)}{P_{j}^{k}} \right\}^{-\sigma_k} \frac{E_{j}^{k}}{P_{j}^{k}} \tag{20}$$

$$\equiv \left\{ \frac{\tau_{ij}^{k} p_{i}^{k}(v)}{P_{j}^{k}} \right\}^{1-\sigma_k} E_{j}^{k} \tag{21}$$

The above expression gives us bilateral exports of a single product variety. Compare equation (6) in AvW (2003), and equation (6) in HMR. To derive something that looks more obviously like a gravity equation, we need to aggregate this expression to give total sectoral exports from i to j, i.e. X_{ij}^{k}. From the production side of the model, it is clear that all firms in a given country-sector are symmetrical in terms of marginal cost, sales, price, etc. Using the measure $N_i$of firms active in country i, we can therefore write total sectoral exports very simply:

$$X_{ij}^{k} = N_i \left\{ \frac{\tau_{ij}^{k} p_{i}^{k}(v)}{P_{j}^{k}} \right\}^{1-\sigma_k} E_{j}^{k} \tag{22}$$

Now comes the important part: introducing a general equilibrium accounting identity. It must be the case that sectoral income in country i, Y_{i}^{k}, is the income earned from total worldwide sales of all locally made varieties in that sector. Thus:

$$Y_{i}^{k} = \sum_{j=1}^{C} X_{ij}^{k} = N_i \left[p_{i}^{k}(v) \right]^{1-\sigma_k} \sum_{j=1}^{C} \left\{ \frac{\tau_{ij}^{k}}{P_{j}^{k}} \right\}^{1-\sigma_k} E_{j}^{k} \tag{23}$$

Solving for $N_i \left[p_{i}^{k}(v) \right]^{1-\sigma_k}$ gives:

$$N_i \left[p_{i}^{k}(v) \right]^{1-\sigma_k} = \frac{Y_{i}^{k}}{\sum_{j=1}^{C} \left\{ \frac{\tau_{ij}^{k}}{P_{j}^{k}} \right\}^{1-\sigma_k} E_{j}^{k}} \tag{24}$$

Next, substitute the above expression back into the sectoral exports equation (22):

$$X_{ij}^k = \frac{Y_i^k E_j^k}{\sum_{j=1}^C \left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k} E_j^k} \left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k} \tag{25}$$

For convenience, define $\left(\Pi_i^k\right)^{1-\sigma_k} = \sum_{j=1}^C \left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k} \frac{E_j^k}{Y^k}$ where Y^k is total world output in sector k. Dividing the above expression through by Y^k and substituting for $\left(\Pi_i^k\right)^{1-\sigma_k}$ gives the AvW gravity equation (compare AvW (2003) equations 9-11, AvW (2004) equations 5-7, and Feenstra Chapter 5 equation 5.35):

$$X_{ij}^k = \frac{Y_i^k E_j^k}{Y^k} \left\{\frac{\tau_{ij}^k}{\Pi_i^k P_j^k}\right\}^{1-\sigma_k} \tag{26}$$

or in the more common log-linearized form:

$$\log\left(X_{ij}^k\right) = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij} - \log \Pi_i^k - \log P_j^k\right] \tag{27}$$

The basic logic of this equation accords with that of the simple gravity model, in the sense that economic mass in the origin and destination markets is positively associated with bilateral trade, but trade costs are negatively associated (since $\sigma_k > 1$ by assumption). However, there are many other interesting and important things going on here, and it useful to address some of them explicitly. (What follows draws in particular on Baldwin and Taglioni, 2007.)

5.1 Relative Prices Matter

One of the main shortcomings of traditional (non-theoretical) gravity models is that they do not take account of reallocations of trade flows across trading partners in general equilibrium. For instance, if country i lowers trade barriers affecting imports from country j only, the traditional gravity model predicts an effect on bilateral trade between i and j, but not on any other trading relationships. This is strongly counter-intuitive: think “trade diversion”.

When you do gravity with gravitas, things start to look more like they should. From (26) and the definitions of Π_i^k and P_j^k, it is immediately apparent that bilateral trade between i and j depends on trade barriers on all worldwide routes, not just those acting directly on ij trade. AvW refer to Π_i^k as outward multilateral resistance, since it aggregates all trade barriers facing exports from i (i.e., across all destination markets). They call P_j^k inward multilateral resistance because it aggregates all trade barriers facing imports into j (i.e., across all origin markets).

As AvW (2003, 2004) show, leaving these terms out of an empirical gravity model has two nasty results. First, the econometric estimates are biased, due to omitted variables. Second, counter-factual simulations produce biased results because they do not take into account the general equilibrium effects captured by the two multilateral resistance terms. (This bias is in addition to the bias resulting from poorly estimated parameters.)

From an estimation point of view, it does not overly matter which approach is taken to account for multilateral resistance (fixed effects, nonlinear estimation as in AvW, or a Taylor series approximation as in Baier and Bergstrand, 2007). When it comes to counter-factual simulations, however, the choice is important. Fixed effects will give unbiased parameter estimates, but counter-factuals will not take account of general equilibrium effects transmitted through the multilateral resistance terms. Either the nonlinear approach of AvW, or a Taylor series approximation, should be used.

5.2 Fixing Fixed Effects

For applied researchers, fixed effects are a very appealing alternative to the relatively complex estimation procedure of AvW (2003). Care needs to be exercised, however, in choosing appropriate dimensions. Following the pattern of subscripts in (26), it is possible to formulate some simple recommendations:

1. In single year datasets using total (aggregate) trade data, multilateral resistance can be captured using exporter and importer fixed effects. These terms capture in each case the sum of income or expenditure and inward or outward multilateral resistance.

2. In single year datasets using sectoral trade data, multilateral resistance can be captured using exporter-sector, importer-sector, and sector fixed effects. These terms capture in each case the sum of sectoral income or expenditure and inward or outward multilateral resistance.

3. In multi-year datasets, interact the fixed effects specified in 1. or 2. above with year effects, and include a full set of year effects. These terms capture the fact that income/expenditure and multilateral resistance are likely to change over time.

5.3 GDP, Prices, and All That

Gravity models usually use GDP in the origin and destination countries as proxies for their economic mass. Alternatively, some authors include population and GDP per capita separately. Some people prefer nominal data, others use real (deflated) data.

It can be important to bring some gravitas to bear on these choices as well. First, the AvW model does not suggest any particular role for per capita GDP in driving bilateral trade. That does not mean that it should never be included in any self-respecting gravity model. But it does mean that researchers interested in the role of income levels should go back to the theory, and find an explicit rationale (and functional form) for its inclusion.

Second, the AvW model suggests that GDP (and trade flow data) should be entered in nominal, not real, terms. To see why, consider an extension of the above model to include a number of years. On the assumption that trade costs, expenditures, or incomes change over time–for whatever reason– the multilateral resistance terms will also change over time. Since these terms are already akin to trade price indices, there is no reason for any additional deflators.

Third, it is important to point out in light of the previous paragraph that although the multilateral resistance terms look and function like price indices, they are not equivalent to commonly available price data like CPIs, PPIs, etc. The multilateral resistance terms depend on world prices, trade barriers, and the intra-sectoral elasticity of substitution. They are aggregated in a very particular way. As a result, they cannot adequately be proxied by "over the counter" price data.

Fourth, and for exactly the same reasons as in the previous paragraph, multilateral resistance terms cannot usually be captured by atheoretical "remoteness" indices. Although they look like remoteness and play a similar role in the model, they have a precise definition that must be respected.

Fifth, it is important to pay particular attention to equation (26) and the associated definitions when using disaggregated (sectoral) data. Aggregate GDP can be a decent proxy for economy-wide income and expenditure. But in a sectoral context, what the model actually needs is sectoral incomes and expenditures. Direct data can be hard to come by, so it will sometimes be appropriate to interact GDP with fixed effects in order to allow these shares to change across sectors and countries. Alternatively, the model can be estimated separately for each sector (i.e., without pooling).

5.4 Exports, Imports, and Bilateral Trade

The model derived above is expressed in terms of exports from i to j. This would tend to suggest that gravity empirics should use export data rather than imports or total bilateral trade (exports+imports). In reality, there can in fact be good reasons for using import data instead, at the price of sacrificing a little theoretical purity. Most countries track imports much more closely than they track exports, because they represent a tax base. Value data, in particular, are likely to be more accurate in the import data than in the export data. There are also usually fewer zeros. A common technique is therefore to use either all import data, or export data for countries that are assumed to be good reporters combined with import ("mirror") data for countries that have more serious data issues.

In any case, there is no particular reason for using total bilateral trade, or the average of exports and imports.

5.5 Cross-Sectoral Heterogeneity and Trade Cost Elasticities

Most gravity models these days are estimated using data disaggregated at the sectoral level. This is good in the sense that it can result in more precise parameter estimates. But it is important to note a feature of gravity models like (26): the reduced form elasticity of trade with respect to trade costs can vary across sectors due to changes in the intra-sectoral elasticity of substitution σ^k. A coefficient of -1.2 on distance in fact represents the product of two structural parameters: the elasticity of trade costs with respect to distance, and $(1-\sigma_k)$. It is thus not entirely correct to interpret a reduced form coefficient of -1.2 as a "distance elasticity". To do that, it would be necessary to divide through by $(1-\sigma_k)$.

In addition, this discussion makes clear that sectoral gravity models should usually allow reduced form trade elasticities to vary by sector. Why? Because the $(1-\sigma_k)$ term varies by sector even if the structural trade cost elasticity does not. This is another good reason for using extensive sectoral interaction terms, or estimating the model separately for each sector.

5.6 Zeros in the Bilateral Trade Matrix

This is not the place for a detailed discussion of "the zero problem". See HMR.

It is important to note, however, that estimating a log-linear gravity model when the trade matrix includes zeros might introduce some serious bias to the extent that there is useful information in the part of the matrix that is dropped. The model set out in these notes does not say anything in particular about zero entries in the bilateral trade matrix, but the heterogeneous firms extension in HMR does. Possible approaches to dealing with zeros include:

1. Ad hoc fixes, such as adding a small positive constant to all zero flows.

2. Econometric fixes that allow the zero entries to be included in the estimation sample, such as the Tobit or Poisson estimators.

3. Theoretical and econometric fixes that 1) provide a rationale for the existence of zeros, and 2) correct for their presence, for instance through a modified version of the Heckman sample selection model.

6 Conclusion

The theoretical structure set out by AvW, while very general, is only one of many that can produce something looking like a gravity model. However, their formulation is currently one of the more important benchmarks. As a result, empirical work should keep as closely as possible to the constraints implied by this model in terms of the estimating equation. Deviations from that benchmark need to be explained, and preferably justified by reference to an explicit (and credible) theory.

The AvW model has important implications for how applied researchers estimate and interpret gravity models, particularly in panel data contexts (multiple years and/or multiple sectors). The selection of dependent and independent variables, the configuration of fixed effects, the decision whether or not to pool data, and the interpretation of reduced-form elasticities, all depend intimately on the theoretical results derived above. Moreover, the model's general equilibrium basis, as captured through the multilateral resistance terms, has important implications for running counter-factual simulations–a particularly important point for policy-oriented work.

References and Further Reading

Anderson, James E., and Eric Van Wincoop, 2003, "Gravity with Gravitas: A Solution to the Border Puzzle", *The American Economic Review*, 93(1), 170-192. ("AvW 2003")

Anderson, James E. and Eric Van Wincoop, 2004, "Trade Costs", *Journal of Economic Literature*, 42(3), 691-751. ("AvW 2004")

Baldwin, Richard E. and Daria Taglioni, 2007, "Trade Effects of the Euro: A Comparison of Different Estimators", *Journal of Economic Integration*, 22(4), 780-818.

Baldwin, Richard E., et al., 2005, "All You Wanted to Know About Dixit-Stiglitz but were Afraid to Ask",
http://team.univ-paris1.fr/teamperso/martinp/dixit-stiglitz.pdf.

Baier, Scott, and Jeffrey H. Bergstrand, 2007, "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects using the Gravity Equation", Working Paper,
http://www.nd.edu/%7Ejbergstr/Working_Papers/BVOLSOctober2007.pdf.

Chaney, Thomas, 2008, "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review*, 98(4), 1707-1721.

Dingel, Jonathan I, 2008, "The Basics of Dixit-Stiglitz Lite",
http://www.columbia.edu/~jid2106/td/dixitstiglitzbasics.pdf.

Feenstra, Robert, *Advanced International Trade*, Chapter 5.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, 2008, "Estimating Trade Flows: Trading Partners and Trading Volumes", *Quarterly Journal of Economics*, 123(2), 441-487. ("HMR")