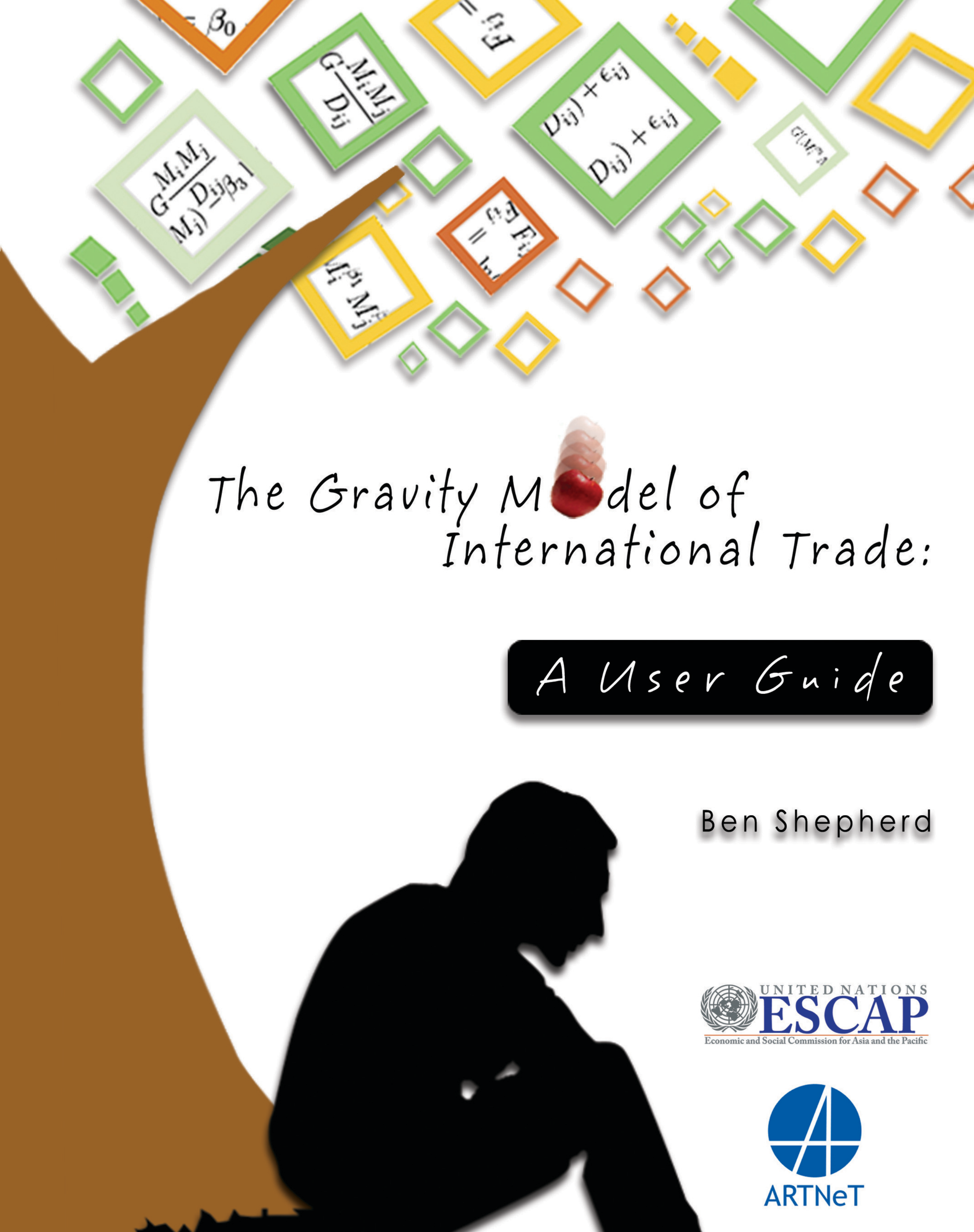


The Gravity Model of International Trade:

A User Guide

Ben Shepherd

UNITED NATIONS ESCAP
Economic and Social Commission for Asia and the Pacific

ARTNeT

The secretariat of the ESCAP is the regional development arm of the United Nations and serves as the main economic and social development centre of the United Nations in Asia and the Pacific. Its mandate is to foster cooperation between its 53 members and 9 associate members. It provides the strategic link between global and country-level programmes and issues. It supports governments of countries in the region in consolidating regional positions and advocates regional approaches to meeting the region's unique socio-economic challenges in a globalizing world. The ESCAP secretariat is located in Bangkok, Thailand. Please visit the ESCAP website at www.unescap.org for further information.

The shaded areas of the map are ESCAP Members and Associate members.

ARTNeT

Asia-Pacific Research and Training Network on Trade (ARTNeT) is an open regional network of research and academic institutions specializing in international trade policy and facilitation issues. IDRC, UNCTAD, UNDP, ESCAP and WTO, as core network partners, provide substantive and/or financial support to the network. The Trade and Investment Division of ESCAP, the regional branch of the United Nations for Asia and the Pacific, provides the Secretariat of the network and a direct regional link to trade policymakers and other international organizations. For more information, please contact the ARTNeT Secretariat at artnetontrade@un.org or visit www.artnetontrade.org.

Cover illustration: Chaveemon Sukpaibool

The Gravity Model of International Trade: A User Guide

Prepared by

Ben Shepherd

ARTNeT Gravity Modeling Initiative

2013

The Gravity Model of International Trade: A User Guide

United Nations publication

Manufactured in Thailand
First published: October 2012
Second published: June 2013
ISBN: 978-974-680-346-5
ST/ESCAP/2645

Foreword

ARTNeT's Gravity Modeling Initiative was started at the end of 2008 to enable integration of the research capacity-building and its research programme implementation. The initiative has been implemented by the ESCAP secretariat in collaboration with the Research and Information System for Developing Countries (RIS), New Delhi (which hosted two workshops) and the World Trade Organization secretariat (which provided resource persons for training and technical assistance). Financial support for ARTNeT as a whole has been provided by the International Development Research Centre, Canada.

The initiative includes the following components: (a) a series of training and research workshops; (b) post-workshop grants for research studies that apply the techniques learnt; (c) an online gravity modeling tool with several datasets; and (d) dissemination of policy-relevant results as inputs into evidence-based policymaking.

This Guide has been developed in response to requests from the participants who attended the workshops, institutional members of ARTNeT and other users of the ARTNeT online gravity tool for simple yet thorough reference material to complement the initiative. This Guide was prepared by Ben Shepherd who has been associated with the initiative from the beginning as an advisor to the ARTNeT secretariat and as a resource person working with ARTNeT researchers.

The purpose of the Guide is to provide a "hands-on" introduction to gravity modeling for applied policy researchers. It is designed to be used in conjunction with a dataset of bilateral trade in services available for free download, and readers are encouraged to replicate the results presented here, using the Stata code provided in the text. Although some basic knowledge of Stata is required, more advanced commands and techniques are introduced in the text as necessary. Once the basic techniques have been mastered, readers are encouraged to extend the results presented here by using alternative specifications and methodologies.

The publication of this book also marks half a century since the gravity model was first introduced to empirical trade research by Tinbergen in 1962. Despite initial scepticism and a slow uptake of this research method, today it is one of the most preferred models among applied trade researchers. It is the ARTNeT secretariat's hope that this Guide will provide useful assistance to the users of gravity modeling and contribute towards improving the quality of research based on use of the model. Enhanced provision of a high-quality and locally produced applied research for policymaking in all trade related areas, including trade facilitation, investment and regional integration, remains one of the most important pillars of ARTNeT's operation.

Ravi Ratnayake
Director TID

Contents

Executive Summary

The gravity model is the workhorse of the applied international trade literature. It has been used in literally thousands of research papers and published articles covering all areas of trade. It is of particular interest to policy researchers because it makes it possible to estimate the trade impacts of various trade-related policies, from traditional tariffs to new "behind-the-border" measures. With data increasingly available for developing, as well as developed countries, the gravity model has come to be the starting point for a wide variety of research questions with a policy component.

Although first put forward as an intuitive explanation of bilateral trade flows, the gravity model has more recently acquired a range of micro-founded theoretical bases. These approaches are important to policy researchers because they affect the data, specification, and econometric technique used to estimate the gravity model. Use of a theoretically-grounded gravity model can lead to substantially different results and interpretations from those obtained via a "naive" formulation, and high quality policy research and advice increasingly needs to be based on a rigorously established methodology.

The purpose of this User Guide is to provide a "hands-on" introduction to gravity modeling for applied policy researchers. It is designed to be used in conjunction with a dataset of bilateral trade in services available for free download, and readers are encouraged to replicate the results presented here using the Stata code provided in the text. Although some basic knowledge of Stata is required as a pre-requisite, more advanced commands and techniques are introduced in the text as necessary. Once the basic techniques have been mastered, readers are encouraged to extend the results presented here using alternative specifications and methodologies.

The User Guide first presents the intuition behind the gravity model, relying on descriptive statistics and graphical methods in addition to simple regressions. The next section of the Guide then presents the theory behind recent, more rigorously specified gravity models, focusing on the "gravity with gravitas" model that is now a benchmark in the applied international trade literature. Section 3 focuses on the basic econometrics of the gravity model, including estimation and testing using theoretically grounded gravity models. Section 4 discusses alternative econometric estimators that have been proposed in the recent literature, and the rationale behind them. Finally, the User Guide concludes with a discussion of how the gravity model can be used in applied trade policy research.

1

Introduction

1 Introduction

Over the last half-century, the gravity model has become the workhorse of the applied international trade literature. Starting with Tinbergen (1962),[1] the gravity model has given rise to literally thousands of publications and working papers covering a wide variety of regions, time periods, and sectors. For example, Disdier and Head (2008) in their meta-analysis of the effect of distance on trade cover 1,052 separate estimates in 78 papers. By linking trade flows directly with economic size and inversely with trade costs, usually proxied by geographical distance as an indicator of transport costs, the gravity model captures some deep regularities in the pattern of international trade and production. Indeed, Leamer and Levinsohn (1995) have argued that the gravity model has produced "some of the clearest and most robust findings in empirical economics".

The gravity model is a key tool for researchers interested in the effects of trade-related policies. It provides a convenient testing bed on which to assess the trade impacts of different policies. Gravity models now routinely include variables far beyond those such as tariffs, which are imposed at the border, to cover behind-the-border barriers as well. Regulatory policies, as well as deep political and institutional characteristics of countries, have been shown to influence trade as modeled in the gravity framework. Moreover, the gravity model is no longer concerned only with trade in goods, but has recently been applied with success to trade in services (e.g., Kimura and Lee, 2006). Indeed, the exercises presented in this user guide will concentrate on the emerging area of trade in services, where increased data availability is making it increasingly feasible to apply the most up-to-date estimation methods and models.

Although the gravity model is an attractive platform for applied international trade researchers, its use does not come without some potential pitfalls. Chief among these is the choice of exactly which model to estimate (specification). Traditionally, gravity models have been based largely on intuitive ideas as to which variables are likely to influence trade. More recently, however, a number of "theoretical" gravity models have been developed, which use various micro-founded theories of international trade to develop gravity-like models. Indeed, Deardorff (1995) has argued that an equation that looks something like gravity must emerge from "just about any sensible trade model".

[1] See De Benedictis and Taglioni (2011) for a review of the development of the gravity model and its early implementations. That paper also contains information on many examples of the successful application of gravity models in policy contexts, and is strongly recommended as complementary reading for this User Guide.

At the same time, econometricians have set out a number of alternative methodologies for estimating the models themselves. The challenge for applied researchers in the current environment is to make best use of these recent theoretical and empirical advances in answering interesting and relevant policy questions. This user guide is envisaged with that challenge in mind.

The purpose of this user guide is to provide policy researchers with an applied introduction to the gravity model and its applications. With the exception of the worked Stata examples, all the material presented here draws on published papers and existing research. The user guide thus serves as a kind of compendium – at an introductory level – of recent advances in gravity model practice. The material is presented in a largely non-technical way that should be accessible to anyone with a grounding in graduate-level microeconomics and econometrics. To keep the presentation simple and uncluttered, proofs of basic propositions are generally omitted, and readers are referred to standard sources – particularly econometrics textbooks – for more detailed information.

In light of its purpose and intended audience, this user guide is a complement to existing sources on the gravity model, its foundations, and its applications. Applied policy researchers can use it as a starting point for their own research, or as a ready reference for Stata code and other technical details. However, each research application of the gravity model is highly specific, and needs to be careful crafted and related to the previous literature. Moreover, the literature in this area is changing rapidly, and is currently unsettled in a number of areas, particularly when it comes to econometric methods for gravity. Researchers therefore need to ensure that they are up to date with the latest developments in the field when putting together their own applications of the model.

With these points in mind, the user guide proceeds as follows. Section 2 first presents the traditional gravity model, which we refer to as the "intuitive" model. It also introduces the dataset used throughout the guide, and shows how we can use Stata to calculate descriptive statistics and produce graphs that allow us to analyze the way in which the model captures some important stylized facts of international trade. In the second part of Section 2, we move on to consider "theoretical" gravity models, namely those with sound microeconomic foundations. We focus in particular on the famous "gravity with gravitas" model of Anderson and Van Wincoop (2003), which has become one of the standard formulations used in applied work. The Appendix provides a full derivation of the model, whereas the text of the section focuses on a general description of its features. In Section 3, our attention turns to estimation of the gravity model using econometric methods. We discuss estimation and testing of the intuitive model using ordinary least squares, and the limitations of that approach. We then discuss two approaches for estimating the theoretical gravity model, including the use of fixed effects panel data methods. Section 4 continues with the econometric analysis by introducing two recent advances in methodology that account for potential problems with the ordinary least squares estimator. In both cases, the presentation is intuitive and application-focused. Readers are referred to the original research for full technical details of the two new estimators we consider, namely Poisson and the Heckman sample selection model. Sections 3 and 4 make extensive use of the sample dataset provided with this user guide, and include all Stata code necessary for producing the results discussed in the text. Finally, Section 5 concludes with a summary of current issues in gravity modeling practice. It also discusses the ways in which applied policy researchers can best make use of the gravity model in producing relevant and technically-sound research.

2

The Basis for the Gravity Model: From Intuition to Theory

2 The Basis for the Gravity Model: From Intuition to Theory

This section provides a general grounding in the gravity model and its relationship to the data. We start the analysis intuitively, and state the basic ideas lying behind the gravity model. We then use descriptive statistics and graphical techniques to investigate whether those ideas are in fact reflected in a dataset of bilateral trade in services. In the second part of the section, we discuss some of the limitations of the traditional, or intuitive, gravity model. This discussion leads into the third part of the section, in which we discuss gravity theory. Over the last decade, theory has become an increasingly important part of gravity modeling, and applied researchers now have the choice among a number of commonly-used theoretically-grounded gravity models. We provide the intuition behind such models in the main text, and then present full derivations of the most commonly used one – the Anderson and Van Wincoop (2003) "gravity with gravitas" model – in an appendix.

2.1 The Gravity Intuition

As noted above, the gravity model was initially presented as an intuitive way of understanding trade flows. In its most basic form, the gravity model can be written as follows:

$$\log X_{ij} = c + b_1 \log GDP_i + b_2 \log GDP_j + b_3 \log \tau_{ij} + e_{ij} \tag{1a}$$

$$\log \tau_{ij} = \log(distance_{ij}) \tag{1b}$$

$X_{ij} = c + \frac{GDP_1 GDP_2}{\tau}$

where X_{ij} indicates exports from country i to country j, GDP is each country's gross domestic product, τ_{ij} represents trade costs between the two countries, distance is the geographical distance between them – as an observable proxy for trade costs – and e_{ij} is a random error term. The c term is a regression constant, and the b terms are coefficients to be estimated. The name "gravity" comes from the fact that the nonlinear form of equation 1a resembles Newton's law of gravity: exports are directly proportional to the exporting and importing countries' economic "mass" (GDP), and inversely proportional to the distance between them (not the square of the distance between them, as in physics). In other words, gravity says that we expect larger country pairs to trade more, but we expect countries that are further apart to trade less, perhaps because transport costs between them are higher.

Before moving to the details of gravity modeling using econometric methods, we can use graphical techniques to examine the basic intuition underlying the model. For all empirical examples in this user guide, we use a dataset on bilateral trade in services compiled by Francois et al. (2009). Their original data has been re-aggregated using a programme supplied by the authors so that the sectoral definition follows the one used by the Global Trade Analysis Project (GTAP). We will usually be interested in total (aggregate) services trade, which is in the sector SER. Readers are free to use the remaining sectors to extend the examples presented here. To facilitate the replication of examples, this user guide provides all necessary Stata code, and it can be used jointly with the dataset accompanying it, which also includes standard gravity model control variables from standard sources such as CEPII, as well as additional variables explained later in the text.

A first step in examining the intuition behind the gravity model is to examine the correlations among the variables. To do that, we first use Stata's *generate* command to transform the dataset into logarithms. We then use the *correlate* command to produce a correlation matrix, as in Table 1. By adding the *if* option, we limit the calculation to the SER sector, i.e. total services trade.

Table 1: Correlation matrix for basic gravity variables

```
. gen ln_trade = ln(trade)
(19012 missing values generated)

. gen ln_distance = ln(dist)
(606 missing values generated)

. gen ln_gdp_exp = ln(gdp_exp)
(991 missing values generated)

. gen ln_gdp_imp = ln(gdp_imp)
(992 missing values generated)

. correlate ln_trade ln_gdp_exp ln_gdp_imp ln_distance if sector == "SER"
(obs=3884)

             | ln_trade ln_gd~xp ln_gd~mp ln_dis~e
-------------+------------------------------------
    ln_trade |   1.0000
  ln_gdp_exp |   0.3643   1.0000
  ln_gdp_imp |   0.3731  -0.3103   1.0000
 ln_distance |  -0.2648   0.0518   0.0431   1.0000
```

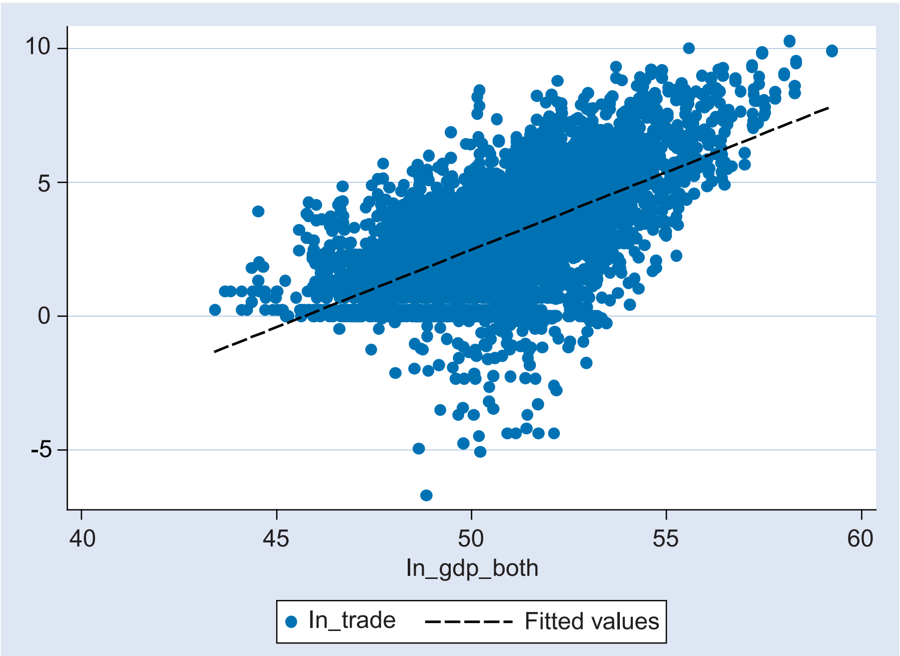
The correlations of interest are contained in the non-diagonal elements of the matrix. We see that trade and GDP are strongly positively correlated, and that the correlation is approximately the same for exporter and importer GDP. This finding supports the basic intuition that bigger countries tend to trade more. By contrast, we find a strong negative correlation between trade and distance: country pairs that are further apart tend to trade less. Again, this finding is in line with the basic intuition of the gravity model.

→ so far everything we've seen in class

We can use Stata's *twoway* command to present the same information graphically. That command allows us to overlay graphs one on top of the other, and we use it to combine a scatter plot of the variables of interest with a linear line of best fit, which reflects the correlation between them. Figure 1 shows the result using the combined economic mass of the exporting and importing countries, i.e. the product of their GDPs, as the explanatory variable. The scatter plot shows a clear positive association between the two variables, in line with the correlation analysis. Similarly, the line of best fit is strongly upward sloping. Graphical evidence therefore also supports the basic intuition that larger country pairs tend to trade more than smaller ones.

Figure 1: Scatter plot and line of best fit for trade versus combined GDP

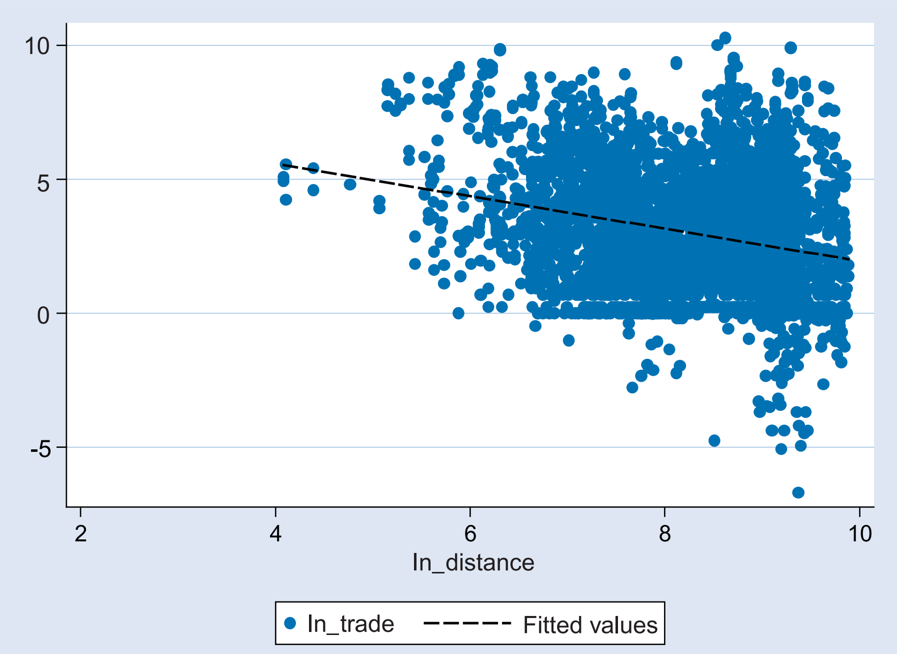
```
. gen ln_gdp_both = ln(gdp_exp*gdp_imp)
(1983 missing values generated)

. twoway (scatter ln_trade ln_gdp_both if sector == "SER") (lfit ln_trade
ln_gdp_both if sector == "SER")
```



We can also use graphical methods to investigate the association between trade and distance. Figure 2 presents results using the same approach as for Figure 1. In this case, the scatter plot is suggestive of a negative association. That impression is reinforced by the line of best fit, which is strongly downward sloping. Graphical evidence again confirms the basic gravity intuition that country pairs that are further apart tend to trade less.

Figure 2: Scatter plot and line of best fit for trade versus distance

```
. twoway (scatter ln_trade ln_distance if sector == "SER") (lfit ln_trade
ln_distance if sector == "SER")
```



2.2 Problems with the Intuitive Gravity Model

The previous section shows that the basic gravity model picks up two important stylized facts of international trade: bigger countries trade more, and more distant countries trade less. These regularities are almost uniformly reflected in the early gravity literature, which applies the model to all regions of the world, covering both developed and developing economies, and various products and time periods. The model is clearly a useful starting point in applied international trade research.

However, the intuitive gravity model is not free of difficulties once more advanced concepts from the trade literature are introduced. As one example, consider the impact on trade between countries i and j of a change in trade costs between countries i and k. An example of such a change might be that countries i and k enter into a preferential trade agreement that lowers tariffs on their respective goods. Basic economic theory suggests that such a move may well impact the trade of country j, even though it is not party to the agreement. The well-known concepts of trade creation and trade diversion are examples of such effects. However, the intuitive gravity model does not account for this issue at all. As is clear from equation 1a, $\frac{\partial \log X_{ij}}{\partial \log t_{ik}} = 0$. Reducing trade costs on one bilateral route therefore does not affect trade on other routes in the basic model, which is at odds with standard economic theory.

A change in transport cost but not in consumer prices, GDP doesn't change so even though costs decrease (which is supposed to increase trade), there's no reflection of that in the model

A second problem with the basic model arises if we consider equal decreases in trade costs across all routes, including domestic trade (goods that a country sells internally, or X_{ii}). An example could be a fall in the price of oil, which lowers transport costs everywhere, including within countries. In the basic model, this move would result in proportional increases in trade across all bilateral routes, including domestic trade. However, such an outcome sits ill with the observation that despite the change in trade costs, relative prices have not changed at all. In the absence of a change in relative prices, we would expect consumption patterns to remain constant for a given amount of total production (GDP). This is a second instance in which the basic gravity model makes predictions that are at odds with standard economic theory.

2.3 "Theoretical" Gravity Models

Issues such as those identified in the previous section led some researchers to turn to theory to provide a basis for a gravity-like model of trade. Clearly, the basic model needs to be altered in some fundamental dimensions if it is to deal with the issues raised in the previous section. The first example of such an approach is Anderson (1979). However, the applied literature has only paid serious attention to theoretically-grounded gravity models since the famous "gravity with gravitas" model of Anderson and Van Wincoop (2003). In this section, we present the basic intuition behind that model, with the mathematical details left to the Appendix.

At its most basic, the Anderson and Van Wincoop (2003) gravity model is essentially a demand function. It owes much of its final form to the constant elasticity of substitution structure chosen for consumer preferences. Consumers have "love of variety" preferences, which means that their utility increases both from consuming more of a given product variety, or from consuming a wider range of varieties without consuming more of any one. → interesting but fair

On the production side, the model makes assumptions that are standard following Krugman (1979). Each firm produces a single, unique product variety under increasing returns to scale. By assuming a large number of firms, competitive interactions disappear and firms engage in constant markup pricing: in equilibrium, the difference between price and marginal cost is just enough to cover the fixed cost of market entry.

A producer in one country can sell goods in any country, either the one where it is located, or an overseas country. To simplify the model, selling goods locally is assumed to involve no transport reflected in distance?

costs. However, selling goods internationally does involve transport costs. Consumers therefore consume product varieties from all countries, but the prices of non-domestically produced varieties are adjusted upwards to take account of the cost of moving goods between countries.

These building blocks make it possible to derive an equilibrium in which firms both produce for the local market and engage in international trade, and in which consumers consume accordingly. The basic model provides expressions for the volume of exports by each firm. Aggregating across firms within an economy then makes it possible to derive an expression for the total value of a country's exports, which is the dependent variable in the gravity model.

so some more relationships

All that is required to produce a gravity-like model from these foundations is to impose some macroeconomic accounting identities. Essentially, these identities flow from the fact that in a single sector economy like the one being modeled, where there are no input-output relationship, the sum of all production must be equal to GDP. Performing the aggregation in an appropriate way makes it possible to derive the "gravity with gravitas" model:

exports indexed over countries (so from one to another?)

elasticity of substitution between the countries?

$$\log X_{ij}^{k} = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] \qquad (2a)$$

GDP

Expenditure (Idk what it does)

world GDP in that sector

$$\Pi_i^k = \sum_{j=1}^{C} \left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k} \frac{E_j^k}{Y^k} \qquad (2b)$$

$$P_j^k = \sum_{i=1}^{C} \left\{\frac{\tau_{ij}^k}{\Pi_i^k}\right\}^{1-\sigma_k} \frac{Y_i^k}{Y^k} \qquad (2c)$$

where X is exports indexed over countries (i and j) and sectors (k), Y is GDP, E is expenditure (which is not necessarily the same as GDP on a sectoral basis), $Y^k = \sum_{i=1}^{C} Y_i^k$ (i.e., world GDP), σ_k is the intra-sectoral elasticity of substitution (between varieties), and τ_{ij}^k is trade costs.

The first notable feature of the Anderson and Van Wincoop (2003) model is its inclusion of two additional variables, Π_i^k and P_j^k. The first is called outward multilateral resistance, and it essentially captures the fact that exports from country i to country j depend on trade costs across all possible export markets. The second is called inward multilateral resistance, and it likewise captures the dependence of imports into country i from country j on trade costs across all possible suppliers. Together, these terms are the key to the model, and they resolve both issues identified as problems with the intuitive gravity model in the previous section. In particular, it is immediately apparent that because the multilateral resistance terms involve trade costs across all bilateral

Think about this more (no idea how to interpret it yet) routes, $\frac{\partial \log X_{ij}}{\partial \log t_{ik}} \neq 0$. In other words, this model picks up the fact that changes in trade cost on one bilateral route can affect trade flows on all other routes because of relative price effects. Because the intuitive model does not include these two multilateral resistance variables but they are, by construction, correlated with trade costs, there is a classic case of omitted variables bias in the intuitive model. Finding a way to correct for this problem will be the main thrust of the estimation approaches discussed in the remainder of this user guide.

Here

The second point to note is that the theoretical gravity model has a number of implications for the way in which a gravity model should be set up, and the types of data that should be used. In the early gravity model literature, some authors used dependent variables such as the logarithm of total trade for a country pair (the sum of exports and imports) or the average of exports in both directions. The theoretical gravity model suggests that such an approach is likely to produce misleading results. The model applies to unidirectional export flows, which means that each line in a gravity database should represent a single flow. Thus, exports from Australia to New Zealand are recorded in one line of the database, and exports from New Zealand to Australia are recorded in a separate one.

Another question that has arisen in the literature is whether trade values should be expressed in nominal or real terms. For the standard cross-sectional gravity model, of course, nothing turns on this issue: data for a single year will give equivalent results regardless of any uniform scaling factor applied. In a time series context, however, this question can be important. The answer given by theory is clear: trade flows should be in nominal, not real, terms. The reason is that exports are effectively deflated by the two multilateral resistance terms, which are special price indices. Deflating exports using different price indices, such as the CPI or the GDP deflator, would not adequately capture the unobserved multilateral resistance terms, and could produce misleading results.

A similar analysis applies to the GDP data used in the model: they too should be in nominal, not real, terms. The reason is again that they are effectively deflated by the multilateral resistance terms, which are unobserved price indices. Deflating by some other factor, such as a readily observable price index, is likely to be misleading. In addition, the sectoral version of the gravity model makes clear that ideally we would like to include data on sectoral expenditure and output rather than GDP as such. However, this is usually impossible in an empirical context – particularly when developing countries are included in the sample – so GDP is used as an acceptable proxy. Finally, the model makes clear that it is aggregate GDP, not per capita GDP, which should be used. Expedients sometimes encountered in the older gravity literature, such as using population and per capita GDP as separate regressors, should therefore be avoided.

The last part of the model that needs to be specified for estimation purposes is the trade costs function τ_{ij}^{k}. The literature typically specifies this function in terms of observable variables that are believed to influence trade costs, using a simple log-linear specification. In the examples in this user guide, we generally specify the trade costs function as follows:

$$\log \tau_{ij}^{k} = b_1 \log distance_{ij} + b_2 contig + b_3 comlang\_off + b_4 colony + b_5 comcol \qquad (3)$$

Distance is the geographical distance between countries i and j, contig is a dummy variable equal to unity for countries that share a common land border, comlang_off is a dummy variable equal to unity for country pairs that share a common official language, colony is a dummy variable equal to unity if countries i and j were once in a colonial relationship, and comcol is a dummy variable equal to unity for country pairs that were colonized by the same power. This formulation is typical of the gravity model literature, in which each of these factors has been found to be a significant determinant of bilateral trade. However, this specification is by no means exhaustive. We use it as a baseline for the examples in this user guide, but researchers will generally need to augment it in applied work to include other variables, especially trade-related policies.

One point that needs to be emphasized in relation to the trade costs function is that it is impossible to separate the elasticity of substitution σ_k from the trade cost elasticities (the b terms) when estimating the model. The two are always multiplied together. This fact has important implications for the way in which gravity models with multiple sectors are estimated, which we discuss in the next section. More broadly, it suggests that we should be wary of interpreting differences in estimated coefficients across sectors as evidence of different levels of sensitivity to particular trade cost factors: we might simply be observing differences in the elasticities of substitution. In order to obtain pure trade cost elasticities, we would need to interact the trade cost variables with estimates of the substitution elasticities, either using a general assumption, or model-based estimates such as those in Broda and Weinstein (2006). However, this path is generally not followed in applied work using the gravity model.

Although we have focused here on the Anderson and Van Wincoop (2003) gravity model, the literature provides a variety of other theoretically-grounded models. For example, Chaney (2008) and Helpman et al. (2008) develop gravity-like equations based on underlying models of trade in which firms are heterogeneous in productivity. Although there are of course important differences among the exact forms of gravity produced by these models, they all retain some fundamental similarities with the basic model set out at the beginning of this section. The same is true of the Ricardian model due to Eaton and Kortum (2002). Applied researchers are therefore free to choose from among a number of theoretical gravity models when developing an estimating model for particular purposes. In the current literature, however, it is increasingly important that a model be as theoretically grounded as possible. It is becoming more and more difficult to justify atheoretical gravity models, based on ad hoc specifications.

Appendix 2.1: Derivation of the Anderson and Van Wincoop (2003) Gravity Model

Consumption Side

Consider a world of C countries indexed by i. We 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. 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:

$$U_i = \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} \left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}} \tag{4}$$

The set Vi defines the range of varieties that is consumed in country i. As usual, we 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. We 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.

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. Anderson and Van Wincoop (2003) and Helpman et al. (2008) 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:

$$E_i = \sum_{k=1}^{K} \left\{ \int_{v \in V_i^k} p_i^k(v) x_i^k(v) dv \right\} \equiv \sum_{k=1}^{K} E_i^k \tag{5}$$

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 (4) subject to (5). The Lagrangian is:

$$\mathcal{L} = \sum_{k=1}^{K} \left\{ \int\limits_{v \in V_i^k} \left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}} - \lambda \sum_{k=1}^{K} \left\{ \int\limits_{v \in V_i^k} p_i^k(v) x_i^k(v) dv \right\} \tag{6}$$

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

$$\frac{\partial \mathcal{L}}{\partial x_i^k(v)} = \frac{1}{1-\frac{1}{\sigma_k}} \left\{ \int\limits_{v \in V_i^k} \left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}} dv \right\}^{\frac{1}{1-\frac{1}{\sigma_k}}-1} \left(1-\frac{1}{\sigma_k}\right) \left[x_i^k(v)\right]^{-\frac{1}{\sigma_k}} - \lambda p_i^k(v) = 0 \tag{7}$$

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:

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

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

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

$$\int\limits_{v \in V_i^k} p_i^k(v) x_i^k(v) dv \equiv E_i^k$$

$$= \lambda^{-\sigma_k} \left[p_i^k(v)\right]^{1-\sigma_k} \left(X^k\right)^{\sigma_k} \left\{ \int\limits_{v \in V_i^k} \left[x_i^k(v)\right]^{1-\frac{1}{\sigma_k}} dv \right\}^{-\sigma_k} \int\limits_{v \in V_i^k} \left[p_i^k(v)\right]^{1-\sigma_k} dv \tag{10}$$

$$\lambda = \left\{ \frac{\int_{v \in V_i^k} \left[p_i^k(v) \right]^{1-\sigma_k} dv}{E_i^k} \right\}^{\frac{1}{\sigma_k}} \frac{X_k}{\int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv} \tag{11}$$

To get the direct demand function, substitute this expression for the Lagrangian multiplier back into the first order condition (8):

$$\frac{\left[x_i^k(v) \right]^{-\frac{1}{\sigma_k}}}{\int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv} X^k = \left\{ \frac{\int_{v \in V_i^k} \left[p_i^k(v) \right]^{1-\sigma_k} dv}{E_i^k} \right\}^{\frac{1}{\sigma_k}} \frac{X_k}{\int_{v \in V_i^k} \left[x_i^k(v) \right]^{1-\frac{1}{\sigma_k}} dv} p_i^k(v) \tag{12}$$

$$\therefore x_i^k(v) = \frac{\left[p_i^k(v) \right]^{-\sigma_k}}{\int_{v \in V_i^k} \left[p_i^k(v) \right]^{1-\sigma_k} dv} E_i^k \equiv \left\{ \frac{p_i^k(v)}{P_i^k} \right\}^{-\sigma_k} \frac{E_i^k}{P_i^k} \tag{13}$$

where $P_i^k = \left\{ \int_{v \in V_i^k} \left[p_i^k(v) \right]^{1-\sigma_k} dv \right\}^{\frac{1}{1-\sigma_k}}$ is the ideal CES price index for sector k in country i.

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 $\sum_{i=1}^{C} N_i^k$. To produce one unit of its product, a firm must pay a fixed cost f_i^k and a variable cost a_i^k . With the wage rate equal to w, a typical firm's profit function is therefore:

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

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:

$$\frac{\partial \pi_i^k(v)}{\partial p_i^k(v)} = x_i^k(v) + p_i^k(v) \frac{\partial x_i^k(v)}{\partial p_i^k(v)} - w a_i^k \frac{\partial x_i^k(v)}{\partial p_i^k(v)} = 0 \tag{15}$$

Solving for prices gives:

$$p_i^k(v) = wa_i^k - \frac{x_i^k(v)}{\frac{\partial x_i^k(v)}{\partial p_i^k(v)}} \tag{16}$$

To do something with that expression, we need to know more about the partial $\frac{\partial x_i^k(v)}{\partial p_i^k(v)}$. In fact, it can be directly evaluated using the demand function (13) and noting that due to the large group assumption (a continuum of firms) $\frac{\partial P_i^k}{\partial p_i^k(v)}$ = 0. In other words, a small change in one firm's price does not affect the overall level of prices in the sector because so many firms are competing. In light of this, we can write:

$$\frac{\partial x_i^k(v)}{\partial p_i^k(v)} = -\sigma_k \left[p_i^k(v)\right]^{-\sigma_k - 1} \left\{\frac{1}{P_i^k}\right\}^{-\sigma_k} \frac{E_i^k}{P_i^k} = -\frac{\sigma_k x_i^k(v)}{p_i^k(v)} \tag{17}$$

The first order condition for profit maximization can therefore be rewritten as:

$$p_i^k(v) = wa_i^k + x_i^k(v) \frac{p_i^k(v)}{\sigma_k x_i^k(v)} \tag{18}$$

then rearranged and solved for prices to give:

$$p_i^k(v) - \frac{1}{\sigma_k} p_i^k(v) \equiv p_i^k(v)\left(1 - \frac{1}{\sigma_k}\right) = wa_i^k \tag{19}$$

$$\therefore p_i^k(v) = \left(\frac{\sigma_k}{\sigma_k - 1}\right) wa_i^k \tag{20}$$

The second term on the right hand side in equation (20) 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.

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 $\tau_{ij}^{k} \geq 1$ 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 $\tau_{ij}^{k} wa_{i}^{k}$. Using this definition, costless trade corresponds to $\tau_{ij}^{k} = 1$, and τ_{ij}^{k} 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 (20) above):

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

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

$$P_{j}^{k} = \left\{ \int_{v \in V_{j}^{k}} \left[\tau_{ij}^{k} p_{i}^{k}(v)\right]^{1-\sigma_k} dv \right\}^{\frac{1}{1-\sigma_k}} \tag{22}$$

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.

Model Closure: 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 (##) with the demand function (##) gives:

$$x_{ij}^{k}(v) = p_{ij}^{k}(v)x_{j}^{k}(v) = \tau_{ij}^{k}p_{i}^{k}(v)\left\{\frac{\tau_{ij}^{k}p_{i}^{k}(v)}{P_{i}^{k}}\right\}^{-\sigma_k}\frac{E_{i}^{k}}{P_{i}^{k}} \equiv \left\{\frac{\tau_{ij}^{k}p_{i}^{k}(v)}{P_{i}^{k}}\right\}^{1-\sigma_k}E_{i}^{k} \tag{23}$$

The above expression gives us bilateral exports of a single product variety. 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:

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

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:

$$Y_{i}^{k} = \sum_{j=1}^{C}X_{ij}^{k} = N_i\left[p_{i}^{k}(v)\right]^{1-\sigma_k}\sum_{j=1}^{C}\left\{\frac{\tau_{ij}^{k}}{P_{j}^{k}}\right\}^{1-\sigma_k}E_{j}^{k} \tag{25}$$

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

$$N_i\left[p_{i}^{k}(v)\right]^{1-\sigma_k} = \frac{Y_{i}^{k}}{\sum_{j=1}^{C}\left\{\frac{\tau_{ij}^{k}}{P_{j}^{k}}\right\}^{1-\sigma_k}E_{j}^{k}} \tag{26}$$

Next, substitute that expression back into the sectoral exports equation (24):

$$X_{ij}^{k} = \frac{Y_{i}^{k}E_{j}^{k}}{\sum_{j=1}^{C}\left\{\frac{\tau_{ij}^{k}}{P_{j}^{k}}\right\}^{1-\sigma_k}E_{j}^{k}}\left\{\frac{\tau_{ij}^{k}}{P_{j}^{k}}\right\}^{1-\sigma_k} \tag{27}$$

For convenience, define $\Pi_i^k = \sum_{j=1}^{C}\left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k}\frac{E_j^k}{Y^k}$ where Y^k is total world output in sector k. Dividing the above expression through by Y^k and substituting Π_i^k gives the Anderson and Van Wincoop (2003) gravity model:

$$X_{ij}^k = \frac{Y_i^k E_j^k}{Y^k}\left\{\frac{\tau_{ij}^k}{\Pi_i^k P_j^k}\right\}^{1-\sigma_k} \tag{28}$$

or in the more common log-linearized form:

$$\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] \tag{29}$$

3

Estimating the Gravity Model

3 Estimating the Gravity Model

This section addresses some of the basic econometric issues that arise when estimating gravity models in practice. It first uses the intuitive gravity model presented in Section 1, and discusses estimation via ordinary least squares and interpretation of results. The next part addresses estimation issues that arise in the context of the "theoretical" gravity model, focusing on the Anderson and Van Wincoop (2003) model discussed in the previous section. The crucial difference between the two approaches is the way in which econometric techniques can be used to account for multilateral resistance, even though the price indices included in the theoretical model are not observable. We discuss two sets of techniques that have been applied in the literature: fixed effects estimation, and the use of a Taylor-series approximation of the multilateral resistance terms. Finally, we address an important issue in the use of gravity models for applied trade policy research, namely possible endogeneity of some explanatory variables.

3.1 Estimating the Intuitive Gravity Model

3.1.1 Ordinary Least Squares: Estimation and Testing

At its most basic, the intuitive gravity model takes the following log-linearized form:

$$\log X_{ij} = b_0 + b_1 \log GDP_i + b_2 \log GDP_j + b_3 \log \tau_{ij} + e_{ij} \qquad (30a)$$

$$\log \tau_{ij} = \log distance_{ij} \qquad (30b)$$

where e_{ij} has been added as a random disturbance term (error). As an econometric problem, the objective is to obtain estimates of the unknown b parameters. The logical place to start is with ordinary least squares (OLS), which is the econometric equivalent of the lines of best fit used to show the connection between trade and GDP or trade and distance in Section 2. As the name suggests, OLS minimizes the sum of squared errors e. Under certain assumptions as to the error term e_{ij} , OLS gives parameter estimates that are not only intuitively appealing but have useful statistical properties that enable us to conduct hypothesis tests and draw inferences.

Under what conditions will OLS estimates of the gravity model be statistically useful? Basic econometric theory lays down three necessary and sufficient conditions:

1. The errors e_{ij} must have mean zero and be uncorrelated with each of the explanatory variables (the orthogonality assumption).

2. The errors e_{ij} must be independently drawn from a normal distribution with a given (fixed) variance (the homoskedasticity assumption).

3. None of the explanatory variables is a linear combination of other explanatory variables (the full rank assumption).

If all three properties hold, then OLS estimates are consistent, unbiased, and efficient within the class of linear models. By consistent, we mean that the OLS coefficient estimates converge to the population values as the sample size increases. By unbiased, we mean that the OLS coefficient estimates are not systematically different from the population values even though they are based on a sample rather than the full population. By efficient, we mean that there is no other linear, unbiased estimator that produces smaller standard errors for the estimated coefficients.

Once we have OLS coefficient estimates that satisfy assumptions one through three, we can use them to test hypotheses using the data and model. To test a hypothesis that involves a single parameter only – for example that the distance elasticity is -1 – we use the t-statistic. To test a compound hypothesis that involves more than one variable – for example that both GDP coefficients are equal to unity – we use the F-statistic. The details of such tests and their statistical properties are fully set out in standard econometric texts. We focus in the next section on their implementation in Stata, and interpretation.

3.1.2 Estimating the Intuitive Gravity Model in Stata

OLS is implemented in Stata in the *regress* command. It takes the following format:

regress dependent_variable independent_variable1 independent_variable2 ... [if ...], [options]

The *if* statement can be used to limit the estimation sample to a particular set of observations. If no *if* command is specified, then the entire sample is used for estimation. Stata automatically handles issues such as missing observations of either the dependent or independent variables – they are dropped from the sample – so there is no need to drop those observations from the dataset prior to estimation.

Among the various options that can be specified with the *regress* command, two are of particular interest in the gravity context. Indeed, they are so widely used in applied work that researchers should not usually present results that do not include these two estimation options. The first is *robust*, which produces standard errors that are *robust* to arbitrary patterns of heteroskedasticity in the data. The *robust* option is therefore a simple and effective way of fixing violations of the second OLS assumption. The second option that is commonly used by gravity modelers is *cluster(variable)*, which allows for correlation of the error terms within groups defined by *variable*. Failure to account for clustering in data with multiple levels of aggregation can result in greatly understated standard errors (e.g., Moulton, 1990). For example, errors are likely to be correlated by country pair in the gravity model context, so it is important to allow for clustering by country pair. To do this, it is necessary to specify a clustering variable that separately identifies each country pair independently of the direction of trade. An example is distance, which is unique to each country pair but is identical for both directions of trade. A common option specification is therefore *cluster(distance)*.

Table 2 presents results for OLS estimation of an intuitive gravity model using the services data. The *if* command is used to limit the estimation sample to total services trade (aggregating across all sectors). In addition to distance, we include a number of other trade cost observables as control variables. Specifically, we include a dummy variable equal to unity for countries that share a common land border (contig), another dummy equal to unity for those countries that share a common official language (comlang_off), a dummy equal to unity for those country pairs that were ever in a colonial relationship, and finally a dummy equal to unity for those countries that were colonized by the same power. There is evidence from the gravity model literature that each of these factors can sometimes exert a significant impact on trade flows, presumably because they increase or decrease the costs of moving goods internationally.

Table 2: OLS estimates of the intuitive gravity model using Stata

```
. regress ln_trade ln_gdp_exp ln_gdp_imp ln_distance contig comlang_off colony
comcol if sector=="SER", robust cluster(dist)

Linear regression                                      Number of obs =    3884
                                                       F(  7,  2151) =  442.01
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5431
                                                       Root MSE      =  1.5281

                                 (Std. Err. adjusted for 2152 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ln_gdp_exp |   .6011672   .0132209    45.47   0.000     .5752401    .6270942
  ln_gdp_imp |   .6176062   .0142666    43.29   0.000     .5896284    .6455839
 ln_distance |  -.7385146     .03536   -20.89   0.000    -.8078579   -.6691714
      contig |   .3989524   .1829276     2.18   0.029     .0402191    .7576858
 comlang_off |   .8861328   .0993078     8.92   0.000     .6913835    1.080882
      colony |   1.202965   .1201971    10.01   0.000     .9672503     1.43868
      comcol |  -.0245067   .2018195    -0.12   0.903    -.4202883     .371275
       _cons |  -22.03706    .671738   -32.81   0.000    -23.35438   -20.71974
------------------------------------------------------------------------------
```

A number of interesting features are apparent from these first estimates. The first is that the model fits the data relatively well: its R2 is 0.54, which means that the explanatory variables account for over 50 per cent of the observed variation in trade in the data. This figure will increase as we add more variables to the model, and in particular once we apply panel data techniques in the next section. A second indication that the model is performing relatively well is that the model F-test is highly statistically significant: it rejects the hypothesis that all coefficients are jointly zero at the 1 per cent level.

To interpret the model results further, we need to look more closely at the estimated coefficients and their corresponding t-tests. Taking the GDP terms first, we see that importer and exporter GDP are both positively associated with trade, as we would expect: a 1 per cent increase in exporter or importer GDP tends to increase services trade by about 0.6 per cent, and this effect is statistically significant at the 1 per cent level (indicated by a p-value in the fifth column of less than 0.01). The coefficient on distance, on the other hand, is negative and 1 per cent statistically significant: a 1 per cent increase in distance tends to reduce trade by about 0.7 per cent. This effect is weaker than in goods trade, where the estimated elasticity tends to be around -0.1. This finding is perhaps in line with the fact that cross-border services trade does not directly engage transport costs, which tends to reduce the impact of geographical distance as a source of trade costs. However, the fact that distance significantly affects trade in services suggests that the world is still far from “flat” in the sense that services do not move costlessly across borders.

Of the remaining geographical and historical variables, all except the common colonizer dummy have the expected positively signed coefficient and are statistically significant at the 5 per cent level or better. Quantifying the effect of each of these types of link on trade is straightforward. For geographical contiguity, for example, we find that countries that share a common border trade 49 per cent more than those that do not (exp[0.4] – 1 = 1.49). Dummy variables can therefore be given a quantitative interpretation in much the same way as continuous variables, although the calculation is different in each case.

By interpreting the coefficient t-statistics, we have already used the model to test a number of simple hypotheses. We can also use it to conduct tests of compound hypotheses. For example, GDP coefficients in the goods trade literature are frequently found to be close to unity – and some theories suggest they should be exactly unity – so we can test whether that is in fact the case in our services data. Table 3 contains results. It shows that the null hypothesis of equality is strongly rejected by the data: the p-value of the F-statistic is less than 0.01, which means that we can reject the hypothesis at the 1 per cent level.

Table 3: A test of the hypothesis that both GDP coefficients are equal to unity

```
. test (ln_gdp_exp = ln_gdp_imp = 1)

 ( 1)  ln_gdp_exp - ln_gdp_imp = 0
 ( 2)  ln_gdp_exp = 1

       F(  2,  2151) =  467.56
            Prob > F =    0.0000
```

Using the same approach, we can test the compound hypothesis that historical and cultural links do not matter for trade in services, i.e. that the coefficients on all such variables are jointly equal to zero. Table 4 presents results. Again, the null hypothesis is strongly rejected: the p-value associated with the F-test is less than 0.01, which means we can reject the null hypothesis at the 1 per cent level. Based on these results, we conclude that historical and cultural links are important determinants of trade in services.

Table 4: A test of the hypothesis that all historical and cultural coefficients are equal to zero

```
. test (contig = comlang_off = colony = comcol = 0)

 ( 1)  contig - comlang_off = 0
 ( 2)  contig - colony = 0
 ( 3)  contig - comcol = 0
 ( 4)  contig = 0

       F(  4,  2151) =  100.16
            Prob > F =    0.0000
```

As a final example of how to estimate the intuitive gravity model, we can augment the basic formulation to include policy variables. The OECD's ETCR indicators are commonly used as measures of the restrictiveness of services sector policies, which cannot be easily quantified in the way that tariffs can be for goods. The OECD dataset only covers 30 countries in our dataset, which greatly reduces the estimation sample. Nonetheless, including measures of exporter and importer policies allows us to get a first idea of the extent to which policy restrictiveness matters as a determinant of the pattern of services trade.

Results for the augmented gravity model are in Table 5. The two variables of primary interest – the exporter and importer ETCR scores – both have negative and 1 per cent statistically significant coefficients of very similar magnitude. In both cases, a one point increase in a country's ETCR score – which equates to a more restrictive regulatory environment, as measured on a scale of zero to six – is associated with a 36 per cent or 37 per cent decrease in trade. Based on these results, we would conclude that policy in exporting and importing countries has the potential to greatly affect the observed pattern of services trade around the world.

In terms of the control variables, results are qualitatively similar to those for the baseline model, although there are some differences in the magnitudes of some coefficients. The only notable differences are for the contiguity dummy, which has an unexpected negative and 5 per cent significant coefficient, and the colony dummy, which has a positive sign, as expected, but is statistically insignificant. It is important to note that the common colonizer dummy has been dropped automatically by Stata because of a lack of within-sample variation for this reduced estimation sample: for the countries for which all data are available, the common colonizer dummy is always equal to zero, which means that it cannot be identified separately from the constant term and must be dropped from the regression.

Table 5: OLS estimates of an augmented gravity model

```
. regress ln_trade etcr_exp etcr_imp ln_gdp_exp ln_gdp_imp ln_distance contig
comlang_off colony comcol if sector == "SER", robust cluster(dist)
note: comcol omitted because of collinearity

Linear regression                                      Number of obs =     816
                                                       F(  8,   413) =  139.24
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6833
                                                       Root MSE      =  1.3835

                              (Std. Err. adjusted for 414 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    etcr_exp |  -.3605257   .0910402    -3.96   0.000    -.5394858   -.1815657
    etcr_imp |  -.3721994   .0796389    -4.67   0.000    -.5287475   -.2156512
  ln_gdp_exp |   .7736852   .0349451    22.14   0.000     .7049927    .8423777
  ln_gdp_imp |   .8223475   .0349431    23.53   0.000      .753659     .891036
 ln_distance |  -1.114939   .0626474   -17.80   0.000    -1.238087   -.9917915
      contig |  -.5579995   .2452544    -2.28   0.023    -1.040102    -.075897
 comlang_off |   1.378263   .2090961     6.59   0.000     .9672377    1.789289
      colony |   .1771059   .2077632     0.85   0.394    -.2312993    .5855111
      comcol |          0  (omitted)
       _cons |  -27.11023   1.799003   -15.07   0.000    -30.64657   -23.57388
------------------------------------------------------------------------------
```

3.2 Estimating the Theoretical Gravity Model

Recall from above that the theory-consistent gravity model due to Anderson and Van Wincoop (2003) can be written as follows, omitting the sectoral superscripts k to focus on the case of aggregate trade:

$$\log X_{ij} = \log Y_i + \log Y_j - \log Y + (1-\sigma)\left[\log \tau_{ij} - \log \Pi_i - \log P_j\right] \qquad (31a)$$

$$\Pi_i = \sum_{j=1}^{C} \left\{\frac{\tau_{ij}}{P_j}\right\}^{1-\sigma} \frac{Y_j}{Y} \qquad (31b)$$

$$P_{\mathrm{j}} = \sum_{i=1}^{C} \left\{\frac{\tau_{ij}}{\Pi_i^k}\right\}^{1-\sigma} \frac{Y_i}{Y} \qquad (31c)$$

$$\log \tau_{ij}^k = b_1 \log distance_{ij} + b_2 contig + b_3 comlang\_off$$
$$+ b_4 colony + b_5 comcol \qquad (31d)$$

As noted above, this model has significant implications for the estimation technique adopted because it includes variables – the multilateral resistance terms – that are omitted from the intuitive model. Moreover, these variables are unobservable, because they do not correspond to any price indices collected by national statistical agencies. We therefore need an estimation approach that allows us to account for the effects of inward and outward multilateral resistance, even though these factors cannot be directly included in the model as data points. This section examines two strategies for doing so: fixed effects estimation; and an approximation technique due to Baier and Bergstrand (2009).

3.2.1 Fixed Effects Estimation

One approach to consistently estimating the theoretical gravity model is to use the panel data technique of fixed effects estimation. Grouping terms together for exporters and importers allows us to rewrite the gravity model from equation 31 as follows:[2]

$$\log X_{ij} = C + F_i + F_j + (1-\sigma)\left[\log \tau_{ij}\right] \tag{32a}$$

$$C = -\log Y \tag{32b}$$

$$F_i = \log Y_i - \log \Pi_i \tag{32c}$$

$$F_j = \log Y_j - \log P_j \tag{32d}$$

$$\log \tau_{ij} = b_1 \log distance_{ij} + b_2 contig_{ij} + b_3 comlang\_off_{ij} + b_4 colony_{ij} + b_5 comcol_{ij} \tag{32e}$$

The first term, C, is simply a regression constant. In terms of the theory, it is equal to world GDP, but for estimation purposes it can simply be captured as a coefficient multiplied by a constant term, since it is constant across all exporters and importers. The next term, F_i, is shorthand for a full set of exporter fixed effects. By fixed effects, we mean dummy variables equal to unity each time a particular exporter appears in the dataset. There is therefore one dummy variable for Australia as an exporter, another for Austria, another for Belgium, etc. We take the same approach on the importer side, specifying a full set of importer fixed effects F_j. In terms of the panel data literature, this approach can be seen as accounting for all sources of unobserved heterogeneity that are constant for a given exporter across all importers, and constant for a given importer

[2] In fact, the exporter and importer fixed effects model provides consistent estimates for any gravity model in which terms can be grouped together in this way. This class of models covers much of the field in applied international trade, including the Ricardian model of Eaton and Kortum (2002) and the heterogeneous firms model of Chaney (2008).

across all exporters. Theory provides a sound motivation for such an approach, as the GDP and multilateral resistance terms satisfy these criteria.

Estimation of fixed effects models is straightforward. Since the fixed effects are simply dummy variables for each importer and exporter, all that is necessary is it to create the dummies and then add them as explanatory variables to the model. Assuming its three key assumptions are satisfied, OLS remains a consistent, unbiased, and efficient estimator in this case. However, the introduction of fixed effects does introduce a major restriction on the model due to the third assumption: variables that vary only in the same dimension as the fixed effects cannot be included in the model, because they would be perfectly collinear with the fixed effects. For example, if we use fixed effects by importer, it is impossible to separately identify the impact of a variable like the importer's ETCR score, which is constant across all exporters for a given importer; it is subsumed into the fixed effects. It is only possible, therefore, to identify the effect of variables that vary bilaterally in fixed effects gravity models.

Two approaches are available in Stata for the estimation of gravity models with fixed effects by importer and by exporter. In both cases, it is first necessary to create variables that list exporters and importers according to numerical codes, rather than by letters as is common in gravity datasets. To do this, we use the *egen* command with the *group* option. The second stage of the process can be achieved either by applying the *i.variable* operator to automatically create dummies during the estimation process, or by using the *tabulate* command with the *generate* option to directly create dummies which must then be included manually in the estimation command.

Tables 6 and 7 present results from OLS estimation of a gravity model with exporter and importer fixed effects using these two approaches. For brevity, Stata's output is cut off after the first few exporter fixed effects. As can be seen from the table, the two approaches give exactly identical results in practice for the variables of interest. The only differences in the two sets of regression outputs come from the requirement that at least one dummy variable must be dropped in order to avoid perfect collinearity between the fixed effects and the constant: the first method automatically chooses a different dummy variable from the second method. There is thus a difference in the estimated fixed effects between the two methods, but this is of no consequence and simply represents scaling with respect to an omitted category. The key result is that regardless of which method is chosen, the estimated coefficients for the variables of interest – which all vary bilaterally – are identical.

Table 6: OLS estimates of a gravity model with fixed effects by importer and exporter (first method)

```
. egen exporters = group(exp)

. egen importers = group(imp)

. regress ln_trade ln_distance contig comlang_off colony comcol i.exporters
i.importers if sector=="SER", robust cluster(dist)

Linear regression                                      Number of obs =     4184
                                                       F(383,  2328) =        .
                                                       Prob > F      =        .
                                                       R-squared     =   0.7681
                                                       Root MSE      =   1.1333

                           (Std. Err. adjusted for 2329 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_distance |  -1.014767   .0469219   -21.63   0.000     -1.10678   -.9227543
      contig |    .235591    .202185     1.17   0.244    -.1608905    .6320725
 comlang_off |   .3982351   .0936922     4.25   0.000     .2145062    .5819639
      colony |   1.173628   .1159908    10.12   0.000     .9461722    1.401084
      comcol |   -.088625   .2584496    -0.34   0.732    -.5954404    .4181904
             |
   exporters |
          2  |  -.3386272     .530869   -0.64   0.524    -1.379652    .7023981
          3  |    2.01065   .6859219     2.93   0.003     .6655682    3.355731
          4  |   -.846119   .6116776    -1.38   0.167    -2.045609    .3533706
```

Table 7: OLS estimates of a gravity model with fixed effects by importer and exporter (second method)

```
. egen exporters = group(exp)

. egen importers = group(imp)

. quietly tabulate exporters, generate(exp_dum_)

. quietly tabulate importers, generate(imp_dum_)

. regress ln_trade ln_distance contig comlang_off colony comcol exp_dum_* imp_dum_*
if sector=="SER", robust cluster(dist)

Linear regression                                      Number of obs =     4184
                                                       F(383,  2328) =        .
                                                       Prob > F      =        .
                                                       R-squared     =   0.7681
                                                       Root MSE      =   1.1333

                           (Std. Err. adjusted for 2329 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_distance |  -1.014767   .0469219   -21.63   0.000     -1.10678   -.9227543
      contig |    .235591    .202185     1.17   0.244    -.1608905    .6320725
 comlang_off |   .3982351   .0936922     4.25   0.000     .2145062    .5819639
      colony |   1.173628   .1159908    10.12   0.000     .9461722    1.401084
      comcol |   -.088625   .2584496    -0.34   0.732    -.5954404    .4181904
   exp_dum_1 |  -.7798491   .4119459    -1.89   0.058    -1.587668      .02797
   exp_dum_2 |  -1.118476   .4488487    -2.49   0.013    -1.998661   -.2382913
   exp_dum_3 |   1.230801   .6464602     1.90   0.057     -.036897    2.498499
```

It is useful to compare results from the fixed effects gravity model with those from the intuitive model without fixed effects. The first notable feature is that, as expected, the model's explanatory power is much greater once the fixed effects are included: it increases from 54 per cent to 77 per cent. This change is unsurprising given that we have added a large number of additional variables to the model, but it underlines the important role played by factors such as multilateral resistance in explaining observed trade outcomes.

The second point to note is that a number of the coefficients are quite different under the two specifications. The distance elasticity, for example, is very close to -1 under fixed effects, which is the value typically observed in goods markets. The difference between the estimated elasticity from the intuitive model and the one from the theoretical model makes clear that the choice of estimation strategy, and the rationale for it, can make an economically significant difference to final results.

We can also use the fixed effects approach to estimate a gravity model augmented to include policy variables. Care is required, however, since the exporter and importer ETCR indicators are perfectly collinear with the corresponding fixed effects. One solution is to create a new variable equal to the product of the two scores, which will by definition vary bilaterally. Table 8 presents results from this approach, again using a significantly smaller sample due to lack of availability of

Table 8: OLS estimates of an augmented gravity model with fixed effects by importer and exporter

```
. gen etcr_both = etcr_exp*etcr_imp
(23562 missing values generated)

. regress ln_trade etcr_both ln_distance contig comlang_off colony comcol
i.exporters i.importers if sector=="SER", robust cluster(dist)
note: comcol omitted because of collinearity

Linear regression                                      Number of obs =     816
                                                       F( 63,    413) =   58.69
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8646
                                                       Root MSE      =  .93726

                                  (Std. Err. adjusted for 414 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   etcr_both |    -.30427   .0917592    -3.32   0.001    -.4846434   -.1238967
 ln_distance |  -.8979641   .1073258    -8.37   0.000    -1.108937   -.6869912
      contig |   .3251201   .2648705     1.23   0.220    -.1955424    .8457826
 comlang_off |   .2087727   .1919348     1.09   0.277    -.1685182    .5860636
      colony |   .4613652   .2723341     1.69   0.091    -.0739685     .996699
      comcol |          0  (omitted)
             |
   exporters |
         15  |  -.8825874   .4081986    -2.16   0.031    -1.684993   -.0801813
         18  |    .316488   .4292144     0.74   0.461    -.5272293    1.160205
         37  |   .4340733   .3321423     1.31   0.192    -.2188269    1.086974
```

the ETCR data for non-OECD countries. Again, we find that policy is a significant determinant of trade flows in services: increasing the product of two countries' ETCR scores by one point decreases trade by about 30 per cent, which is a very similar magnitude to the one found using the intuitive model. The effect is statistically significant at the 1 per cent level.

We have focused thus far on the simple case of aggregate trade, in which multilateral resistance can be accounted for by including exporter and importer fixed effects in the model. If we add sectors or time periods to the model, however, the situation becomes more complicated for the specification of fixed effects, as noted by Baldwin and Taglioni (2007). Consider a sectoral model, for example:

$$\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] \qquad (33a)$$

$$\Pi_i^k = \sum_{j=1}^{C} \left\{\frac{\tau_{ij}^k}{P_j^k}\right\}^{1-\sigma_k} \frac{E_j^k}{Y^k} \qquad (33b)$$

$$P_{\mathrm{j}}^k = \sum_{i=1}^{C} \left\{\frac{\tau_{ij}^k}{\Pi_i^k}\right\}^{1-\sigma_k} \frac{Y_i^k}{Y^k} \qquad (33c)$$

Collecting terms in this case produces a different arrangement of fixed effects from the aggregate trade model. Because trade costs potentially vary by sector, the multilateral resistance terms also vary in that dimension. They can therefore not be adequately captured by importer and exporter fixed effects. Instead, we need sector, exporter-sector, and importer-sector fixed effects, as follows:

$$\log X_{ij}^k = C^k + F_i^k + F_j^k + (1-\sigma_k)\left[\log \tau_{ij}^k\right] \qquad (34a)$$

$$C = -\log Y^k \qquad (34b)$$

$$F_i^k = \log Y_i^k - \log \Pi_i^k \qquad (34c)$$

$$F_j^k = \log Y_j^k - \log P_j^k \qquad (34d)$$

A second difficulty arises from the fact that the elasticity of substitution σ_k also varies across sectors. Since the reduced form parameters of the trade cost function are joint estimates of the elasticity of substitution and the elasticity of trade costs with respect to particular factors, it is important to take account of this variation in a model including multiple sectors. One option would be to interact the trade cost observables with estimates of the elasticity of substitution from Broda and Weinstein (2006). A simpler alternative would be to interact the trade cost observables with sector dummies. Although necessary to conform to theory, neither approach is regularly used in the applied literature.

Fixed effects estimation is a simple and feasible approach in aggregate gravity models. However, models including a large number of sectors quickly become unmanageable due to the number of parameters involved. There is no econometric limitation involved – the number of observations is always far greater than the number of parameters – but gravity models with large numbers of fixed effects and interaction terms can be slow to estimate, and may even prove impossible to estimate with some numerical methods such as Poisson and Heckman (see the next section). A more feasible alternative in such cases is to estimate the model separately for each sector in the dataset: a separate model for trade in business services versus trade in transport services, etc. With this approach, all that is needed for each model is a full set of exporter and importer fixed effects, as in the aggregate trade version of the model. The fact that each sector represents a separate estimation sample allows for multilateral resistance and the elasticity of substitution to vary accordingly. Indeed, it can often be useful from a research point of view to estimate separate sectoral models: knowledge of differences in the sensitivity of trade with respect to policy in particular sectors can be important in designing reform programmes, for example. This approach is therefore frequently used in the literature.

3.2.2 Estimation Without Fixed Effects

The fixed effects model provides a convenient way to consistently estimate the theoretical gravity model: unobservable multilateral resistance is accounted for by dummy variables. The method is simple to implement and is just an application of standard OLS. It has one important drawback, however: we need to drop from the model any variables that are collinear with the fixed effects. This restriction means that it is not possible to estimate a fixed effects model that also includes data that only vary by exporter (constant across all importers) or by importer (constant across all exporters). Unfortunately, many policy data – in fact, all policies that are applied on a most-favored nation basis – fall into this category, which means that the restriction poses a particular challenge for applied policy researchers.

One way of dealing with this problem is to take variables that vary by exporter or importer and transform them artificially into a variable that varies bilaterally. This was what we did with the ETCR scores above: by multiplying them together, the result is a variable that is unique to each country pair and therefore varies across importers for each exporter and across exporters for each importer. Such variables can be included in a fixed effects model without difficulty. However, the price of transforming variables in this way is that the model results become harder to interpret. In the last table, for example, we cannot distinguish the impact of changes in importer policies from that of exporter policies, which is potentially an important question. Although the overall policy message from the last regression was clear, such is not always the case with transformed variables: results can often carry perverse signs or unlikely magnitudes, which mean that transformation should be used cautiously in policy work.

The panel data econometrics literature provides an alternative to fixed effects estimation that still accounts for unobserved heterogeneity, but allows the inclusion of variables that would be collinear with the fixed effects. This alternative is the random effects model. Although it has been applied in gravity contexts – examples include Egger (2002) and Carrère (2006) – we will not discuss random effects estimation extensively here. There are two main reasons for not doing so. First, fixed effects estimation remains largely dominant in the literature because the random effects model is only consistent under restrictive assumptions as to the pattern of unobserved heterogeneity in the data. In the context of the theoretical gravity model, the random effects model requires us to assume that multilateral resistance is normally distributed, yet theory has nothing to say on that question. The fixed effects specification, by contrast, allows for unconstrained variation in multilateral resistance. Second, accounting for both inward and outward multilateral resistance requires specification of a two-dimensional random effects model – random effects by exporter and by importer – which is rarely treated in the literature. Although such models can be implemented straightforwardly in Stata using the *xtmixed* command, they have received scant consideration either in the econometrics literature or in the applied policy literature. The probable reason is that fixed effects modeling is generally preferred for gravity work because theoretical models do not say anything about the statistical distribution of trade costs or multilateral resistance.

A third, and determinant, consideration is that Baier and Bergstrand (2009) provide an alternative approach that fully accounts for arbitrary distributions of inward and outward multilateral resistance but without the inclusion of fixed effects. The Baier and Bergstrand (2009) methodology therefore makes it possible to consistently estimate a theoretical gravity model that also includes variables such as policy measures that vary by exporter or by importer, but not bilaterally. Their approach relies on a first order Taylor series approximation of the two nonlinear multilateral resistance terms. Concretely, Baier and Bergstrand (2009) show that the following model provides estimates almost indistinguishable from those obtained using fixed effects, but without the inclusion of dummy variables:

$$\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^{k^*}\right] \tag{35a}$$

$$\log \tau_{ij}^{k^*} = \log \tau_{ij}^k - \sum_{j=1}^{N} \theta_j^k \log \tau_{ij}^k - \sum_{i=1}^{N} \theta_i^k \log \tau_{ji}^k + \sum_{i=1}^{N}\sum_{j=1}^{N} \theta_i \theta_j \log \tau_{ij}^k \tag{35b}$$

$$\theta_j^k = \frac{Y_i^k}{Y^k} \tag{35c}$$

To deal with endogeneity concerns – see below – Baier and Bergstrand (2009) recommend estimating the model using simple averages rather than GDP-weights. For simplicity, we consider a Stata application of this approach using distance as the only trade cost variable. The calculations included here can easily be replicated for other variables, but they are omitted for brevity in this case. Table 9 presents results from a fixed effects model, and Table 10 presents results using the Baier and Bergstrand (2009) methodology with simple averages. Clearly, the two sets of results are very similar: the distance coefficient is only marginally different at the second decimal place, partly due to differences in the effective samples of the two regressions because of the absence of GDP data for a small number of countries. This finding shows that the Baier and Bergstrand (2009) approximation indeed performs well when it comes to capturing the effects of multilateral resistance in the data without including fixed effects.

Table 9: OLS estimates of a simple gravity model with fixed effects by importer and exporter

```
. regress ln_trade ln_distance i.exporters i.importers if sector=="SER", robust
cluster(dist)

Linear regression                                      Number of obs =    4184
                                                       F(379,  2328) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.7483
                                                       Root MSE      =  1.1801

                                (Std. Err. adjusted for 2329 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_distance |  -1.128312   .0462662   -24.39   0.000    -1.219039   -1.037585
             |
   exporters |
          2  |  -.6754425    .634101    -1.07   0.287    -1.918904    .5680191
          3  |   1.919711    .846599     2.27   0.023     .2595445    3.579878
          4  |  -.3117572   .5766871    -0.54   0.589    -1.442631    .8191167
```

Table 10: OLS estimates of a simple gravity model estimated using the Baier and Bergstrand (2009) methodology

```
. egen temp1 = mean(ln_distance), by(exp sector)
(303 missing values generated)

. egen temp2 = mean(ln_distance), by(imp sector)
(303 missing values generated)

. egen temp3 = sum(ln_distance), by(sector)

. gen ln_distance_star = ln_distance - temp1 - temp2 + (1/218^2)*temp3
(606 missing values generated)

. regress ln_trade ln_distance_star ln_gdp_exp ln_gdp_imp if sector == "SER", robust
cluster(dist)

Linear regression                                      Number of obs =    3884
                                                       F(  3,  2151) =  971.38
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5019
                                                       Root MSE      =  1.5947
```

Table 10: *(continued)*

(Std. Err. adjusted for 2152 clusters in dist)

ln_trade	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_distance_star	-1.142375	.050619	-22.57	0.000	-1.241642	-1.043108
ln_gdp_exp	.572658	.0131492	43.55	0.000	.5468715	.5984445
ln_gdp_imp	.5883749	.0140075	42.00	0.000	.5609052	.6158445
_cons	-35.01455	.7036024	-49.76	0.000	-36.39436	-33.63474

3.3 Dealing with Endogeneity

Regardless of whether we are estimating the intuitive gravity model or its theoretical counterpart, we need to pay particular attention to the problem of endogeneity, particularly when policy variables are included in the model. The reason is that policies are often determined to some extent by the level of a country's integration in international markets: more open economies have an incentive to implement more liberal policies, for example, which creates a circular causal chain between policies and trade. From an econometric point of view, endogeneity of an explanatory variable violates the first OLS assumption by creating a correlation between that variable and the error term. To see this, we can write down two equations that summarize the problem. The first is our gravity model:

$$\log X_{ij}^k = C + F_i^k + F_j^k + (1-\sigma_k)\left[\log \tau_{ij}^k\right] + e_{ij}^k \tag{36}$$

The second equation says that trade costs (particularly those driven by policy) are endogenous to trade flows:

$$\log \tau_{ij}^k = D + G_i^k + G_j^k + b \log X_{ij}^k + w_{ij}^k \tag{37}$$

By substitution:

$$\log X_{ij}^k = C + F_i^k + F_j^k + (1-\sigma_k)\left[D + G_i^k + G_j^k + b \log X_{ij}^k + w_{ij}^k\right] + e_{ij}^k \tag{38}$$

The first OLS assumption will only hold if w_{ij}^k and e_{ij}^k are uncorrelated, which is often unlikely in a practical context. As a result, researchers need to be extremely cautious when interpreting the results of gravity models with policy variables: the estimated parameters could be severely biased due to endogeneity, if they are left uncorrected.

Thankfully, basic econometrics provides us with a simple technique to deal with such endogeneity problems. If we can find an instrumental variable – a piece of data that is correlated with the potentially endogenous variable but not with trade through any other mechanism – then we can use it to purge the problematic variable of its endogenous variation. Various techniques are available for instrumental variables estimation, the simplest of which is two stage least squares (TSLS). As the name suggests, it consists in running OLS twice. The first regression uses the potentially endogenous variable as the dependent variable, and includes all the exogenous variables from the model as independent variables, along with at least one additional instrument. The second regression uses the estimated values of the dependent variable from the first stage regression in place of the problematic variable in the gravity model itself. We can think of the estimated values from the first stage as the part of the problematic variable that varies due to exogenous influences (the instrument and exogenous variables), which solves the endogeneity problem.[3]

For the TSLS estimator to work properly and provide results superior to OLS, three conditions must be satisfied. The first is that there must be at least as many instruments as potentially endogenous variables, and preferably one extra. Having the same number of instruments as potentially endogenous variables is a necessary condition for model identification, but including at least one additional instrument makes it possible to perform an additional diagnostic test that is an important indicator of instrument validity. The second condition is that the instrumental variable or variables must be strongly correlated with the potentially endogenous explanatory variable. To test whether this is in fact the case, we perform an F-test of the null hypothesis that the coefficients on the instruments are jointly equal to zero in each of the first stage regressions. First stage F-tests should be systematically reported whenever TSLS is used. The third condition is that the instruments must be validly excludable from the second stage regression, in the sense that they do not influence the dependent variable other than through the potentially endogenous variable. In an over-identified model, we can test whether this condition holds using the Hansen J-statistic. The null hypothesis for the test is that the residuals from both stages of the regression are uncorrelated, which is equivalent to assuming that the exclusion condition holds. A high value of the test statistic (low prob. value) indicates that the instruments may not be validly excludable, and the TSLS strategy needs to be rethought. Like the first stage F-tests, Hansen's J should be routinely reported when it is available.

Although it is possible to run the TSLS estimator manually in Stata, researchers should generally avoid doing so. One reason is that the standard errors from the second stage regression need to be corrected in order to avoid downward bias. It is also preferable on a practical level to use a built in TSLS estimator as it automatically includes the right set of variables in the first stage regression, i.e. all exogenous variables from the main model plus the instrument.

[3] Although correct parameter estimates can be obtained by running OLS twice manually, the estimated standard errors from the second stage will be biased downwards as they do not correct for the first stage estimation procedure. Researchers should always use Stata's built in instrumental variables estimation commands rather than estimating the models manually.

Stata's built in TSLS estimator is the *ivregress* command with the *tsls* option. However, it is generally preferable to use the user-written *ivreg2* command, which contains a host of additional test statistics and diagnostic information that is important in assessing the performance of the TSLS estimator. To install *ivreg2*, simply type *findit ivreg2* and follow the prompts. The format for *ivreg2* is similar to *regress*, but with the addition of some specific information on the endogenous variables and instruments in parentheses:

ivreg2 dependent_variable exogenous_variables (endogenous_variables = instruments), options

If no additional options are specified, *ivreg2* uses TSLS as the estimator. In addition to the standard *robust* and *cluster* options, it is also important to include the *first* option. This option presents the first stage regression results, which always need to be reported when TSLS is used. Another useful option is *endog(endogenous variable)*, which provides a test of the null hypothesis that the listed variables are in fact exogenous to the model. If the null hypothesis is not rejected and all other tests for the validity of the TSLS estimator are satisfied, that is an indication that endogeneity may not be a serious problem in the data. *Ivreg2* automatically presents other test statistics, such as Hansen's J, if appropriate.

As an example of how the TSLS estimator can be applied to gravity models, we will revert to the intuitive model augmented to include policy variables, namely the exporter and importer ETCR scores. We are using the intuitive model for expositional clarity only. In applied work, it would be important to include fixed effects in addition to the variables discussed here, and TSLS works as normal in the presence of fixed effects. However, the large number of additional parameters and the need to transform both the policy variables and instruments to be bilaterally varying makes it problematic to present such an approach as an example. We therefore use the simpler model for this purpose.

The first step in applying the TSLS estimator is to identify at least two instruments for the policy variables (exporter and importer ETCR scores), which are potentially endogenous. Identification of appropriate instruments is often extremely difficult due to the twin requirements of strength and excludability discussed above. As an example, we use the absolute value of a country's latitude as an instrument for its ETCR score. The rationale is that countries that are further away from the equator tend to be more developed than those close to the equator, and this is reflected in a more liberal policy stance. Latitude could also be a proxy for the level of institutional and governance development, which is also correlated with more liberal policies.

Is latitude likely to be a valid choice of instrument? We will need to examine the first stage F-tests before deciding whether the correlation with the potentially endogenous variables is strong enough. We can say with certainty, however, that latitude is genuinely exogenous to the model. Indeed, researchers often use geographical or historical features as instruments precisely because they must be exogenous to current variables such as trade flows. The final criterion is excludability. Because the model is just identified, we will be unable to test that condition directly using Hansen's J. We therefore need to rely on economic logic to make the argument that latitude does not affect trade except through the policy measures captured in the ETCR scores. Clearly, this part of the instrument validity argument is potentially problematic, for at least two reasons. One is that institutional quality as proxied by latitude might be directly correlated with trade as a source of trade costs in its own right. A second problem is that distance from the equator is likely to be correlated with distance from major trading partners, which provides a third possible way of influencing trade. It is important to stress, therefore, that the instrumental variables strategy used here is presented as an example only. In published work, it would be necessary to go further down the path of identifying a more strongly excludable instrument. It would also be highly preferable to over-identify the model by including at least one extra instrument. Nonetheless, the basic approach outlined here demonstrates the basic logic of TSLS estimation, and is sufficient for present purposes.

Table 11 presents estimation results using TSLS. The first part of the Stata output (Table 11a) shows the first stage regression results for the first potentially endogenous variable, i.e. the exporter's ETCR score. We see that the appropriate instrument, namely the exporter's latitude, is indeed strongly correlated with the policy variable: the t-test rejects the null hypothesis that the coefficient is zero at the 1 per cent level, as does the first stage F-test reported at the bottom of the table. The difference between the two tests is of course that the F-test uses both instruments, whereas the t-test focuses on one only. Based on these results, we conclude that our instruments are indeed strongly correlated with the potentially endogenous variables, as required. Moreover, the direction of the correlation is as expected: countries that are further away from the equator tend to be more developed, have more liberal trade-related policies, and thus lower ETCR scores (negative correlation).

Table 11b presents the same output for the second potentially endogenous variable, i.e. the importer's ETCR score. Results are nearly identical in every respect. We therefore draw similar conclusions: latitude is a strong and appropriate instrument for the importing country's ETCR score.

Second stage results appear in Table 11c. We find that even after correcting for the potential endogeneity of the policy variables, they are still negatively and statistically significantly associated with trade flows. The magnitudes of the coefficients are different from those in the OLS model, though not by very much. This is a preliminary indication that any bias induced by endogeneity may not be severe in these data. This impression is reinforced by the endogeneity test (at the bottom of the table), which does not reject the null hypothesis that the two policy variables are in fact exogenous to the model. Subject to the validity of the instruments – see above – we therefore conclude that endogeneity is not a major problem with the policy variables in this dataset, and that once it is corrected for, the original insight still stands.

Table 11a: TSLS estimates of an augmented gravity model

```
. gen ln_lat_exp = ln(abs(lat_exp))
(292 missing values generated)

. gen ln_lat_imp = ln(abs(lat_imp))
(292 missing values generated)

. ivreg2 ln_trade (etcr_exp etcr_imp = ln_lat_exp ln_lat_imp) ln_gdp_exp ln_gdp_imp
ln_distance contig comlang_off colony comcol, robust cluster(dist) first endog
(etcr_exp etcr_
> imp)
Warning - collinearities detected
Vars dropped:       comcol

First-stage regressions
-----------------------

First-stage regression of etcr_exp:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on dist

Number of clusters (dist) = 415                          Number of obs =     3590
                                                         F(  8,    414) =    57.61
                                                         Prob > F      =   0.0000
Total (centered) SS     =  1596.948251                   Centered R2   =   0.3547
Total (uncentered) SS   =  19720.29201                   Uncentered R2 =   0.9477
Residual SS             =  1030.498577                   Root MSE      =    .5364

------------------------------------------------------------------------------
             |               Robust
    etcr_exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ln_gdp_exp |  -.1685608   .0101994   -16.53   0.000    -.1886099   -.1485118
  ln_gdp_imp |    .011711    .015874     0.74   0.461    -.0194926    .0429146
 ln_distance |  -.1557105   .0248945    -6.25   0.000    -.2046458   -.1067751
      contig |   -.119582   .1152261    -1.04   0.300    -.3460832    .1069192
 comlang_off |  -.1896983   .1055387    -1.80   0.073    -.3971569    .0177602
      colony |   .0315528   .1496084     0.21   0.833    -.2625339    .3256395
  ln_lat_exp |  -1.924733   .1016979   -18.93   0.000    -2.124641   -1.724824
  ln_lat_imp |  -.1585988   .1198288    -1.32   0.186    -.3941476      .07695
       _cons |    15.6712   1.054823    14.86   0.000     13.59772    17.74468
------------------------------------------------------------------------------
Included instruments: ln_gdp_exp ln_gdp_imp ln_distance contig comlang_off
                      colony ln_lat_exp ln_lat_imp
------------------------------------------------------------------------------
Partial R-squared of excluded instruments:   0.3205
Test of excluded instruments:
  F(  2,   414) =   179.52
  Prob > F      =   0.0000

First-stage regression of etcr_imp:
```

Table 11b: TSLS estimates of an augmented gravity model

```
OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on dist

Number of clusters (dist) = 415                         Number of obs =     3590
                                                        F(  8,   414) =    55.76
                                                        Prob > F      =   0.0000
Total (centered) SS     =  1564.952922                  Centered R2   =   0.3501
Total (uncentered) SS   =  19562.95282                  Uncentered R2 =   0.9480
Residual SS             =  1017.044609                  Root MSE      =    .5329

------------------------------------------------------------------------------
             |               Robust
    etcr_imp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ln_gdp_exp |   .0063751   .0155364     0.41   0.682     -.024165    .0369151
  ln_gdp_imp |  -.1794399   .0104756   -17.13   0.000    -.2000318   -.1588479
 ln_distance |  -.1274362   .0242096    -5.26   0.000    -.1750253   -.0798471
      contig |  -.0668288   .1140222    -0.59   0.558    -.2909635     .157306
 comlang_off |  -.1994695   .1102494    -1.81   0.071    -.4161878    .0172488
      colony |   .0241459   .1541118     0.16   0.876    -.2787933    .3270851
  ln_lat_exp |  -.1223865    .113011    -1.08   0.279    -.3445335    .0997604
  ln_lat_imp |  -1.934644   .1094315   -17.68   0.000    -2.149755   -1.719533
       _cons |   15.76784   1.017276    15.50   0.000     13.76817    17.76751
------------------------------------------------------------------------------
Included instruments: ln_gdp_exp ln_gdp_imp ln_distance contig comlang_off
                      colony ln_lat_exp ln_lat_imp
------------------------------------------------------------------------------
Partial R-squared of excluded instruments:   0.3114
Test of excluded instruments:
  F(  2,   414) =   156.35
  Prob > F      =   0.0000
```

Table 11c: TSLS estimates of an augmented gravity model

```
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on dist

Number of clusters (dist) = 415                        Number of obs =      3590
                                                       F(  8,    414) =    112.12
                                                       Prob > F      =    0.0000
Total (centered) SS     =  26167.87607                 Centered R2   =    0.4340
Total (uncentered) SS   =  61089.90172                 Uncentered R2 =    0.7575
Residual SS             =  14812.32073                 Root MSE      =     2.031

------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    etcr_exp |  -.5422142   .1498024    -3.62   0.000    -.8358215   -.2486069
    etcr_imp |  -.4674795   .1077936    -4.34   0.000    -.6787511   -.2562078
  ln_gdp_exp |   .6185911   .0329753    18.76   0.000     .5539608    .6832214
  ln_gdp_imp |   .6987833   .0337433    20.71   0.000     .6326476    .7649191
 ln_distance |  -1.149847   .0651766   -17.64   0.000     -1.27759   -1.022103
      contig |  -.5352804     .28467    -1.88   0.060    -1.093223    .0226625
 comlang_off |   1.128319   .2443539     4.62   0.000     .6493942    1.607244
      colony |  -.0253281   .2938231    -0.09   0.931    -.6012108    .5505546
       _cons |   -20.4278   1.845306   -11.07   0.000    -24.04454   -16.81107
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):            377.037
                                                   Chi-sq(1) P-val =    0.0000
------------------------------------------------------------------------------
Weak identification test (Kleibergen-Paap rk Wald F statistic):       1046.102
Stock-Yogo weak ID test critical values: 10% maximal IV size              7.03
                                         15% maximal IV size              4.58
                                         20% maximal IV size              3.95
                                         25% maximal IV size              3.63
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         0.000
                                                 (equation exactly identified)
-endog- option:
Endogeneity test of endogenous regressors:                               0.495
                                                   Chi-sq(2) P-val =    0.7809
Regressors tested:    etcr_exp etcr_imp
------------------------------------------------------------------------------
```

4

Alternative Gravity Model Estimators

4 Alternative Gravity Model Estimators

The previous section primarily used OLS as the estimation methodology for a variety of gravity models, both intuitive and theoretical. Just as the basic model has been subject to increasing scrutiny from a theoretical point of view, so too has OLS as the baseline estimator been subject to criticism from an econometric point of view. This section describes two alternative estimators from the literature – Poisson and Heckman – and discusses their application to the gravity model, as well as their respective advantages and disadvantages. The bottom line for applied policy researchers is that both estimators are now very commonly used in the literature, and it is therefore important to ensure that results obtained using OLS are robust to their application.

4.1 The Poisson Pseudo-Maximum Likelihood Estimator

Consider the nonlinear form of the Anderson and Van Wincoop gravity model with a multiplicative error term:

$$X_{ij}^{k} = \frac{Y_i^k E_j^k}{Y^k}\left(\frac{\tau_{ij}^k}{\Pi_i^k P_j^k}\right)^{(1-\sigma_k)} e_{ij}^k \tag{39}$$

Taking logarithms gives the standard gravity model in linearized form, but makes clear that the error term is in logarithms too:

$$\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + \log e_{ij}^k \tag{40}$$

The mean of $\log e_{ij}^k$ depends on higher moments of e_{ij}^k, thus including its variance. If e_{ij}^k is heteroskedastic, which is highly probable in practice, then the expected value of the error term depends on one or more of the explanatory variables because it includes the variance term. This violates the first assumption of OLS and suggests that the estimator may be biased and inconsistent. It is important to note that this kind of heteroskedasticity cannot be dealt with by simply applying a robust covariance matrix estimator, since it affects the parameter estimates in addition to the standard errors. The presence of heteroskedasticity under the assumption of a multiplicative error term in the original nonlinear gravity model specification requires adoption of a completely different estimation methodology.

Santos Silva and Tenreyro (2006) present a simple way of dealing with this problem. They show that under weak assumptions – essentially just that the gravity model contains the correct set of explanatory variables – the Poisson pseudo-maximum likelihood estimator provides consistent estimates of the original nonlinear model. It is exactly equivalent to running a type of nonlinear least squares on the original equation. Since we are dealing with a pseudo-maximum likelihood estimator, it is not necessary that the data be in fact distributed as Poisson. So although Poisson is more commonly used as an estimator for count data models, it is appropriate to apply it far more generally to nonlinear models such as gravity.

The Poisson estimator has a number of additional desirable properties for applied policy researchers using gravity models. First, it is consistent in the presence of fixed effects, which can be entered as dummy variables as in simple OLS. This is an unusual property of nonlinear maximum likelihood estimators, many of which have poorly understood properties in the presence of fixed effects. The point is a particularly important one for gravity modeling because most theory-consistent models require the inclusion of fixed effects by exporter and by importer.

Second, the Poisson estimator naturally includes observations for which the observed trade value is zero. Such observations are dropped from the OLS model because the logarithm of zero is undefined. However, they are relatively common in the trade matrix, since not all countries trade all products with all partners (see e.g., Haveman and Hummels, 2004). Although the issue has mainly arisen to date in the context of goods trade, it is also relevant for services trade (see further below). Dropping zero observations in the way that OLS does potentially leads to sample selection bias, which has become an important issue in recent empirical work (see further below). Thus the ability of Poisson to include zero observations naturally and without any additions to the basic model is highly desirable.

Third, interpretation of the coefficients from the Poisson model is straightforward, and follows exactly the same pattern as under OLS. Although the dependent variable for the Poisson regression is specified as exports in levels rather than in logarithms, the coefficients of any independent variables entered in logarithms can still be interpreted as simple elasticities. The coefficients of independent variables entered in levels are interpreted as semi-elasticities, as under OLS.

Stata contains a built in *poisson* command that can easily be applied to the gravity model, but it suffers from a number of numerical issues that result in sometimes unstable or unreliable results. A better option for applied researchers is to use the *ppml* command developed by Santos Silva and Tenreyro (2011b). The command can be installed by typing *findit ppml* and following the prompts. The *ppml* command automatically uses the *robust* option for estimation, so it is not necessary to specify it. Clustering can be corrected for in the usual way, i.e. by specifying the *cluster(dist)* option. If the command experiences estimation problems, it is sometimes possible to work around them by expressing the dependent variable as trade in thousands or millions of dollars, rather than in dollars: large values of the dependent variable are more difficult to treat numerically, and dividing through by a constant does not make any difference to the final result, as the estimator is scale-invariant.

Table 12 presents results for a fixed effects gravity model estimated using the *ppml* command. The first point to note is that, as expected, the number of observations is greater using Poisson than using OLS: 6580 compared with 3884. This difference shows that there is a large number of zero observations in the dataset, which is typical for gravity data. Those observations were dropped from the OLS estimates because the dependent variable was in logarithms, but they can be included naturally by Poisson.

Table 12: Poisson estimates of a fixed effects gravity model

```
. ppml trade ln_distance contig comlang_off colony comcol exp_dum_* imp_dum_* if
sector=="SER", cluster(dist)
note: checking the existence of the estimates
note: starting ppml estimation
note: exp_dum_218 omitted because of collinearity
note: imp_dum_218 omitted because of collinearity
note: trade has noninteger values

Iteration 1:   deviance =  919576.3
Iteration 2:   deviance =  571271.1
Iteration 3:   deviance =    505370
Iteration 4:   deviance =  494184.4
Iteration 5:   deviance =  492564.3
Iteration 6:   deviance =    492333
Iteration 7:   deviance =  492304.4
Iteration 8:   deviance =  492302.2
Iteration 9:   deviance =  492302.1
Iteration 10:  deviance =  492302.1
Iteration 11:  deviance =  492302.1

Number of parameters: 407
Number of observations: 6580
Number of observations dropped: 148
Pseudo log-likelihood: -256224.74
R-squared: .86676832
                                (Std. Err. adjusted for 3357 clusters in dist)
------------------------------------------------------------------------------
             |               Semirobust
       trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_distance |    -.55767   .0483891   -11.52   0.000    -.6525108   -.4628292
      contig |   .2205841   .1670074     1.32   0.187    -.1067443    .5479126
 comlang_off |   .4592715   .1174326     3.91   0.000      .229108    .6894351
      colony |   .1420645   .1153311     1.23   0.218    -.0839803    .3681094
      comcol |  -.5527961   .3745067    -1.48   0.140    -1.286816    .1812235
   exp_dum_1 |  -.2121395   .8042537    -0.26   0.792    -1.788448    1.364169
   exp_dum_2 |  -1.537334   .6055117    -2.54   0.011    -2.724115   -.3505525
   exp_dum_3 |    2.25762   .5087708     4.44   0.000     1.260448    3.254793
```

It is also notable that the Poisson model fits the data much better than does the original OLS model. R2 for the former is around 87 per cent, compared with 77 per cent for OLS. Given that the set of explanatory variables is the same in both cases, this difference suggests that the change in estimator is important in order to pick up significant features of the data, most likely heteroskedasticity of the type outlined above.

Finally, the coefficient estimates are significantly different under Poisson compared with OLS. In particular, the distance coefficient is smaller in absolute value. This result is typical of Poisson gravity regressions, and largely reflects the impact of heteroskedasticity on the original OLS estimates (Santos Silva and Tenreyro, 2006).

The Poisson estimator is becoming steadily more popular in the literature, but it is not free from divergent opinions. Applied researchers need to be aware of some of the issues that have been highlighted in relation to the Poisson estimator, and of the additional properties that have been demonstrated for it as a result.

On the one hand, some researchers have used alternative count data models in place of Poisson, such as the negative binomial model, on the assumption that trade data are likely to exhibit over-dispersion (variance greater than the mean). However, this approach is erroneous for two reasons. First, Poisson is consistent as a pseudo-maximum likelihood estimator regardless of how the data are in fact distributed. The only improvement that could come from allowing for over-dispersion would be in terms of efficiency. For the efficiency gain to be real, moreover, the exact nature of the over-dispersion would need to be known, which it usually is not. Second, the negative binomial estimator has an undesirable property in a trade context: it is not scale invariant. Thus, results from a model with trade in dollars as the dependent variable will be different from those obtained with trade in millions of dollars as the dependent variable. This feature of the negative binomial model is not problematic in its usual count data setting, but becomes worrying in the gravity modeling context. Applied researchers should therefore avoid the negative binomial model in practice.

A second argument that has been made is that it other estimators may be superior to Poisson because they allow for a greater proportion of zeros in the observed trade matrix. However, the response to this argument is identical to the previous one: Poisson is consistent regardless of how the data are in fact distributed, assuming only that the zero and non-zero observations are produced by the same data generating process (see further below). In any case, recent simulation evidence (Santos Silva and Tenreyro, 2011a) shows that Poisson performs strongly even in datasets with large numbers of zeros.

Taking all of these points together, there is a strong argument for using Poisson as the workhorse gravity model estimator. From an applied policy research point of view, the desirable properties of Poisson suggest that estimates of policy impacts should generally be based on Poisson results rather than OLS.[4] In any case, Poisson results should always be presented for comparative purposes or as a robustness check.

[4] The choice between OLS and Poisson is, of course, an empirical one. Santos Silva and Tenreyro (2006) present a test for determining whether the OLS estimator is appropriate, and another for determining whether Poisson or another pseudo-maximum likelihood estimator is likely to be efficient. However, a detailed presentation of these tests is outside the scope of the current user guide.

4.2 The Heckman Sample Selection Estimator

As noted above, zero trade flows are relatively common in the bilateral trade matrix. As the level of product disaggregation becomes greater, so do zeros become more frequent. Even using aggregate trade data, Helpman et al. (2008) report that around half of the bilateral trade matrix is filled with zeros. Dropping these observations – as OLS automatically does because the logarithm of zero is undefined – immediately gives rise to concerns about sample selection bias. The sample from which the regression function is estimated is not drawn randomly from the population (all trade flows), but only consists of those trade flows which are strictly positive. One way of thinking of this problem is that the probability of being selected for the estimation sample is an omitted variable in the standard gravity model. To the extent that that variable is correlated with any of the variables included in the model – which it certainly is, since the probability of trading no doubt depends on trade costs – then a classic case of omitted variable bias arises. This is a violation of the first OLS assumption, which can lead to biased and inconsistent parameter estimates.

One way of dealing with this problem is to use the sample selection correction introduced by Heckman (1979). Helpman et al. (2008) developed a model of international trade that yields a gravity equation with a Heckman correction combined with an additional correction for firm heterogeneity. We explore only the first part of their model here (the Heckman correction), not the second.

To apply the Heckman sample selection model to the gravity model, we first need to split it into two parts: an outcome equation, which describes the relationship between trade flows and a set of explanatory variables, and a selection equation, which describes the relationship between the probability of positive trade and a set of explanatory variables. The outcome equation takes the form of the standard gravity model, but makes clear that it only applies to those observations within the estimation sample:

$$\left.\begin{array}{l}\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + e_{ij}^k \; if \; p_{ij}^k > 0 \\ \log X_{ij}^k = missing \; if \; p_{ij}^k \leq 0\end{array}\right\} \quad (41)$$

The variable p_{ij}^k is a latent (unobserved) variable that can be interpreted as the probability that a particular data point is included in the estimation sample. The selection equation relates the latent variable to a set of observed explanatory variables. That set must include all variables in the outcome equation, and preferably at least one additional variable that affects the probability that two countries engage in trade, but not the volume of such trade once it takes place.[5] One possible candidate is the cost of market entry of the exporter and the importer from the World

[5] Strictly speaking, the Heckman model is just-identified when the two sets of explanatory variables are the same. However, identification is only achieved due to the fact that the inverse Mill's ratio is a nonlinear function of the explanatory variables. Results therefore tend to be more stable when an additional over-identifying variable is included in the selection model. Applied researchers should try to specify an over-identified model whenever possible.

Bank's Doing Business dataset, as used by Helpman et al. (2008) in robustness checks. Using this variable, the selection equation takes the following form, where p_{ij}^k is the (unobserved) probability of selection, and d_{ij}^k is an (observed) dummy variable equal to unity for those observations that are in the sample, and zero for those that are not.

$$p_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + b \log entry_{ij} + w_{ij}^k \quad (42)$$

$$\left.\begin{array}{l} d_{ij}^k = 1 \; if \; p_{ij}^k > 0 \\ d_{ij}^k = 0 \; if \; p_{ij}^k \leq 0 \end{array}\right\} \quad (43)$$

where w_{ij}^k is a standard error term.

Another way of looking at the sample selection problem is that it creates bias if the error terms in the selection and outcome equations are correlated. In the trade context, we have good reason to believe that such correlation will be significant, in light of the tendency of firms to self-select into export status (Helpman et al., 2008).

Intuitively, the solution proposed by Heckman (1979) amounts to a two-step procedure. The first step is to estimate the probability that a particular observation is included in the gravity model sample, using a probit estimator. We therefore estimate:

$$Prob\left(d_{ij}^k = 1\right) = \Phi\left(\log Y_i^k + \log E_j^k - \log Y^k \right.$$

$$\left. + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + b \log entry_{ij}\right) \quad (44)$$

The probit estimates are then used to calculate the inverse Mill's ratio $\left(\frac{\phi}{\Phi}\right)$, which corresponds to the probability of selection variable omitted from the original equation.[6] Inclusion of the additional variable solves the omitted variable bias and produces estimates that are consistent in the presence of non-random sample selection.

There are a number of technical drawbacks to actually performing the Heckman (1979) correction as a two-step procedure, however. Most researchers therefore use a maximum likelihood procedure in which the selection and outcome equations are estimated simultaneously. That approach is implemented by default in Stata's *heckman* command. The format for *heckman* is similar to that for *regress*, but the option *select(variables)* must always be specified. This option tells Stata which variables to include in the selection equation. As noted above, that list must always include the full set of variables from the original gravity model, and preferably one additional variable, such as entry costs, that effects the probability that two countries engage in trade, but not the volume of trade conditional on the existence of a trading link.

[6] In technical terms, the inverse Mill's ratio is the ratio of the probability density function to the cumulative distribution function.

Table 13 presents results for a Heckman sample selection model using our services data. The first part of the output is the outcome equation, i.e. the usual gravity model. Coefficients are quite close to their OLS counterparts, except for the common colonizer dummy, which has the expected positive sign in the Heckman results but is statistically insignificant.

Table 13: Heckman estimates of a fixed effects gravity model

```
. heckman ln_trade ln_distance contig comlang_off colony comcol i.exporters
i.importers if sector == "SER", select(ln_distance contig comlang_off colony comcol
ent_cost_both i.ex
> porters i.importers) robust cluster(dist)

Iteration 0:   log pseudolikelihood = -6436.0627
Iteration 1:   log pseudolikelihood = -6433.5323
Iteration 2:   log pseudolikelihood = -6433.5284
Iteration 3:   log pseudolikelihood = -6433.5284

Heckman selection model                         Number of obs      =      5164
(regression model with sample selection)        Censored obs       =      1681
                                                Uncensored obs     =      3483

                                                Wald chi2(322)     =         .
Log pseudolikelihood = -6433.528                Prob > chi2        =         .

                             (Std. Err. adjusted for 2617 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_trade     |
 ln_distance |   -1.05512   .0461723   -22.85   0.000    -1.145617   -.9646245
      contig |   .2037951   .2015601     1.01   0.312    -.1912554    .5988456
 comlang_off |    .461497   .0964098     4.79   0.000     .2725372    .6504568
      colony |   1.115957    .115507     9.66   0.000     .8895671    1.342346
      comcol |   .2982764   .2700312     1.10   0.269     -.230975    .8275277
             |
   exporters |
           3 |   2.565928   .6870591     3.73   0.000     1.219317    3.912539
           5 |  -1.370435   .8193477    -1.67   0.094    -2.976327    .2354564
           8 |    3.77945   .4426346     8.54   0.000     2.911903    4.646998


-------------+----------------------------------------------------------------
select       |
 ln_distance |   -.6646448   .1862166    -3.57   0.000    -1.029623   -.2996671
      contig |   .9822229   .4741379     2.07   0.038     .0529297    1.911516
 comlang_off |   .4065678   .1439841     2.82   0.005     .1243641    .6887715
      colony |   .8266087   .2456122     3.37   0.001     .3452177       1.308
      comcol |   .1321516   .2518237     0.52   0.600    -.3614138    .6257171
ent_cost_both |  -.2595793   .2237018    -1.16   0.246    -.6980268    .1788682


-------------+----------------------------------------------------------------
     /athrho |  -.1166728   .0709321    -1.64   0.100    -.2556972    .0223515
    /lnsigma |   .0552157   .0166455     3.32   0.001     .0225912    .0878402
-------------+----------------------------------------------------------------
         rho |  -.1161463   .0699752                     -.2502665    .0223478
       sigma |   1.056769   .0175904                      1.022848    1.091814
      lambda |  -.1227397   .0744213                     -.2686028    .0231234
------------------------------------------------------------------------------
Wald test of indep. eqns. (rho = 0): chi2(1) =     2.71   Prob > chi2 = 0.1000
```

The second part of Table 13 presents results for the selection equation, i.e. the probit model of export participation. In line with intuition, distance has a negative and 1 per cent statistically significant impact on the probability that two countries engage in trade. The historical and cultural variables all have the expected positive signs and are at least 5 per cent statistically significant, except for the common colonizer dummy, which is statistically insignificant. We can therefore see that geography as well as cultural and historical ties do not just influence the volume of trade between countries (outcome equation), but also the probability that two countries engage in trade at all (selection equation). Finally, the entry cost variable – which we have transformed to vary bilaterally by taking the product of the exporter and importer values – has the expected negative sign, but its coefficient is statistically insignificant. In this case, further research would be required with alternative over-identifying variables to try and deal with this problem. To produce stable and robust results, researchers should be careful to choose an over-identifying variable that both has a coefficient that accords with intuition, and one which is statistically significant at conventional levels.

The final part of the Stata output provides information on the relationship between the outcome and selection equations. As noted above, sample selection only creates bias if the error terms of the two equations are correlated. That information is contained in Stata's estimate of the parameter rho: an estimate that is large in absolute value (up to a maximum of one) and statistically significant suggests that sample selection is a major problem in a given dataset. In fact, the final line of Stata's output is a test of the null hypothesis that rho is equal to zero, i.e. that the two error terms are uncorrelated. If the null hypothesis is rejected, we conclude that sample selection is a serious issue. In this case, the test statistic is marginal at the 10 per cent level. It would be dangerous to conclude, however, that sample selection is not an issue: use of a better over-identifying variable may well give different results. Indeed, evidence from goods markets suggests that sample selection can indeed create significant bias (Helpman et al., 2008).

Like the Poisson estimator, the Heckman model provides a natural way of including zero trade observations in the dataset. As yet, the literature does not provide any decisive guidance on which model should be preferred in applied work. Each has its own set of advantages and disadvantages. For example, Poisson deals well with heteroskedasticity, but Heckman does not. Similarly, fixed effects Poisson models are well understood and have desirable statistical properties, but fixed effects probit models suffer from a technical issue – the incidental parameters problem – that introduces bias and inconsistency into the estimates, but the empirical extent of that issue is still unclear. On the other hand, Heckman allows for separate data generating processes for the zero and non-zero observations, whereas Poisson assumes that all observations are drawn from the same distribution. For the moment, then, applied researchers generally need to present both Poisson and Heckman results in the interests of showing that their results are robust to the use of different, but commonly used, estimators.

5

Conclusion: Using the Gravity Model for Policy Research

5 Conclusion: Using the Gravity Model for Policy Research

As noted at the outset, various versions of the gravity model have been widely used in the applied international trade literature for over half a century. The model represents the standard starting point for much empirical work in international trade, and for that reason is of particular interest to applied policy researchers. However, as the previous sections have noted, the gravity literature has undergone a series of major changes in the last decade or so. To produce policy research that is credible and robust, it is necessary to take full account of those changes when undertaking research using the gravity model. Increasingly, research that does not use the latest models and techniques does not represent a sound basis for drawing policy conclusions.

The first point for applied researchers to take away from this user guide is that the gravity model is no longer just an intuitive way of summarizing the relationship among trade, economic size, and distance. A variety of theoretical gravity models now exist, which provide firm micro-foundations for gravity-like models. As demonstrated by the "gravity with gravitas model", the inclusion of theory can make a major difference to the way the dataset is set up, the way in which the model is estimated, and most importantly, to the results and policy conclusions that flow from the model. It is therefore important that research based on the gravity model make explicit reference to theory, and incorporate in so far as possible the insights that flow from it. Policy conclusions are only as robust as the model behind them, and it is increasingly necessary to use a theory-consistent gravity model to convince readers that model results are meaningful. As a starting point, all gravity model research should now include appropriate dimensions of fixed effects, or otherwise correct for the multilateral resistance terms introduced by Anderson and Van Wincoop (2003), for example using the Baier and Bergstrand (2009) methodology.

A second point of particular importance to policy researchers, but which is often overlooked, relates to the inclusion of policy variables in gravity models. There is a long tradition of augmenting gravity models in that way, and there is an increasingly large body of literature that uses policy variables, including behind-the-border barriers. However, the possible endogeneity of these measures is always a serious issue in the gravity context. Since endogeneity can introduce serious bias into the model's results – and thus affect policy conclusions – it is important that researchers attempt to correct for it whenever possible. The simplest way to do so is using the TSLS estimator, with at least as many exogenous and excludable instruments as potentially endogenous variables. Although not technically difficult to implement, the TSLS estimator is challenging for researchers because of the need to identify appropriate instruments: they must be strong, exogenous, and excludable. If one of these conditions is not met, the TSLS estimator is no longer valid, and results can even be worse than with OLS. It is therefore important to pay attention to the standard diagnostic statistics, and to report them systematically when the TSLS estimator is used.

Another way in which econometrics is important in the applied gravity modeling context relates to the recent literature on the appropriate estimator to use to estimate gravity models. The literature in this area remains particularly unsettled, with two major contributions focusing on the Poisson estimator as a way of overcoming heteroskedasticity, and the Heckman sample selection estimator as a way of modeling zero trade flows. The bottom line for applied researchers is that it is important to ensure that results are robust to estimation using different techniques. Much of the empirical literature now presents results using Poisson and/or Heckman at least as a robustness check, if not as a first line approach. It is therefore important to ensure that policy conclusions are robust to the estimation of the model using these techniques, as well as others that may be developed in the literature subsequently.

More fundamental than all of these points, however, is the need for applied researchers to focus on questions where gravity modeling has a comparative advantage. In particular, the gravity model describes the behavior of trade flows, but not economic welfare as such. For applications that focus on economic welfare, it would be more appropriate to use other methodologies, such as computable general equilibrium modeling, rather than gravity. The same applies to reallocations of labor and capital across sectors as a result of trade liberalization: gravity is very poorly placed to answer such questions, and alternative methodologies, such as computable general equilibrium modeling, need to be considered. Gravity's comparative advantage lies in the use of data to assess the sensitivity of trade to particular trade cost factors, including policies. To the extent that policy data are available, they can be combined with the gravity model to provide useful information on the likely response of trade flows to reforms. Indeed, in an extension of the approaches presented here, gravity modeling can also be used to perform counterfactual evaluations of the behavior of trade flows following reforms. However, counterfactuals need to be performed very carefully: see Baier and Bergstrand (2009) for a simple way of performing them while taking proper account of the impact of multilateral resistance. Taking account of multilateral resistance is important because it allows counterfactual simulations to properly capture third-country effects such as trade creation and trade diversion. Counterfactual simulations using the intuitive or fixed effects gravity model only measure pure impact effects, and do not consider the general equilibrium implications of policy changes, which is a very significant disadvantage.

If these points are kept in mind, the gravity model can be a useful tool for applied trade policy researchers. As the richness of applications over the last half-century demonstrates, there is enormous scope for adapting the model to changing circumstances and policy priorities. It continues to provide valuable insights in a policy context, and appears likely to remain the workhorse of the applied international trade literature for some time to come.

References

Anderson, J. 1979. "A Theoretical Foundation for the Gravity Model." *American Economic Review*, 69(1): 106-116.

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

Baier, S., and J. Bergstrand. 2009. "Bonus Vetus OLS: A Simple Method for Approximating International Trade Cost Effects using the Gravity Equation." *Journal of International Economics*, 77(1): 77-85.

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

Broda, C., and D. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics*, 121(2): 541-585.

Carrère, C. 2006. "Revisiting the Effects of Regional Trade Agreements on Trade Flows with Proper Specification of the Gravity Model." *European Economic Review*, 50(2): 223-247.

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

Deardorff, A. 1995. "Determinants of Bilateral Trade: Does Gravity Work in a Neo-Classical World?" in J. Frankel (ed.) *The Regionalization of the World Economy*. Chicago: University of Chicago Press.

De Benedictis, L., and D. Taglioni. 2011. "The Gravity Model in International Trade" in L. De Benedictis and L. Salvatici (eds.) *The Trade Impact of European Union Preferential Policies: An Analysis through Gravity Models*. Berlin: Springer.

Disdier, A.-C., and K. Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics*, 90(1): 37-48.

Eaton, J., and S. Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741-1779.

Egger, P. 2002. "An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials." *The World Economy*, 25(2): 297-312.

Francois, J., O. Pindyuk, and J. Woerz. 2009. "Trends in International Trade and FDI in Services: A Global Database of Services Trade." Discussion Paper No. 20090802, IIDE.

Haveman, J., and D. Hummels. 2004. “Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization.” *Canadian Journal of Economics*, 37(1): 199-218.

Heckman, J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica*, 47(1): 153-161.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. “Estimating Trade Flows: Trading Partners and Trading Volumes.” *Quarterly Journal of Economics*, 103(2): 441-487.

Kimura, F., and H.-H. Lee. 2006. “The Gravity Equation in International Trade in Services.” *Review of World Economics*, 142(1): 92-121.

Krugman, P. 1979. “Increasing Returns, Monopolistic Competition, and International Trade.” *Journal of International Economics*, 9(4): 469-479.

Leamer, E., and J. Levinsohn. 1995. “International Trade Theory: The Evidence” in G. Grossman and K. Rogoff (eds.) *Handbook of International Economics*. Amsterdam: Elsevier.

Moulton, B. 1990. “An Illustration of the Pitfall in Estimating the Effects of Aggregate Variables on Micro Units.” *Review of Economics and Statistics*, 72(2): 334-338.

Santos Silva, J., and S. Tenreyro. 2006. “The Log of Gravity.” *Review of Economics and Statistics*, 88(4): 641-658.

Santos Silva, J., and S. Tenreyro. 2011a. “Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator.” *Economics Letters*, 112(2): 220-222.

Santos Silva, J., and S. Tenreyro. 2011b. “Poisson: Some Convergence Issues.” *Stata Journal*, 11(2): 207-212.

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

ARTNeT

Collaborating with ARTNeT

ARTNeT, an open network of research and academic institutions and think-tanks in developing countries in the Asia-Pacific region, aims to increase the amount of high quality, topical and applied research in the region by harnessing existant research capacity and developing new capacities. ARTNeT undertakes regional team research projects, enhances research dissemination mechanisms, increases interactions between policymakers and researchers, and organizes specific capacity building activities catering to researchers, research institutions and think-tanks from low-income developing countries.

Governments, development agencies and other regional or international organizations are encouraged to partner with ARTNeT and explore the potential for collaboration. Research and academic institutions interested in taking part in ARTNeT activities are invited to contact the ARTNeT Secretariat or visit the ARTNeT website at www.artnetontrade.org where information on forthcoming activities is available.

More information on ARTNeT membership guidelines may be obtained via:
http://www.unescap.org/tid/artnet/how-to-join.asp

ARTNeT Secretariat
United Nations
Economic and Social Commission for
Asia and the Pacific
Trade and Investment Division
United Nations Building
Rajadamnern Nok Avenue
Bangkok 10200, Thailand
Tel.: +66 2 288 1902
Fax: +66 2 288 1027, 288 3066
E-mails: artnetontrade@un.org
escap-tid@un.org

www.artnetontrade.org

Follow ARTNeT on

Facebook   Twitter   LinkedIn Group

ISBN 978-974-680-346-5

United Nations publication
Printed in Bangkok
June 2013